Crowding out disclosure: Amplification and stress test design[*]

Ansgar Walther and Daniel Quigley[†]

October 2015 – Job market paper

Abstract

This paper introduces public information in a model of verifiable disclosure to assess the equilibrium impact of public signals such as ratings and stress tests. We show that public information crowds out incentives to disclose evidence, and that this effect is amplified by a 'reverse unraveling' mechanism. Consequently, more informative public signals can leave agents worse informed in equilibrium when this indirect effect dominates the direct effect of better signals. We explore the implications of these effects for welfare and policy in a model of financial crises. Banks' incentives to remain opaque in crises respond to the precision of stress tests. Stress test design is therefore subject to the Lucas critique, and we characterize optimal stress tests given these constraints. Our model has further implications for corporate finance and certification in consumer markets.

Please download the current version: http://www.economics.ox.ac.uk/Academic/ansgar-walther

1 Introduction

Asymmetric information can disrupt markets (Akerlof, 1970), but is often mitigated by two factors: First, informed parties, such as potential borrowers in a credit market or sellers of a good, can disclose verifiable evidence of their quality. Second, uninformed agents, such as potential lenders or buyers, have access to public information about quality. In this paper, we analyze the interaction

---

[†]Department of Economics and Nuffield College, University of Oxford.
Email: ansgar.walther@economics.ox.ac.uk; daniel.quigley@economics.ox.ac.uk.

between private disclosures and public information and study how this interaction affects market efficiency. We consider a model in which an informed Sender can communicate verifiable evidence to an uninformed Receiver. Evidence is costly, and the Receiver observes public information in addition to the Sender's message. We show that public information can crowd out private incentives to disclose evidence, especially for Senders whose quality is high. This mechanism is subject to strong amplification effects. When public information improves, this strongly reduces incentives to disclose evidence. We show that the latter effect can dominate, in which case better public information exacerbates informational asymmetries in equilibrium.

Our results are particularly relevant for financial crises and macroprudential policy. During the crisis of 2007-9, for instance, regulators performed stress tests to elicit the quality of banks' assets and made their results public, in an attempt to reduce the asymmetry of information between banks and investors. At the same time, banks voluntarily disclosed more detailed evidence (Sowerbutts and Zimmerman, 2013). A growing literature studies the optimal way to publish stress test results, but mostly assumes that banks will not disclose evidence voluntarily.[1] Thus, its policy implications are potentially subject to the Lucas (1978) critique: If stress tests affect banks' incentives to disclose, then changing their design can lead to endogenous responses in information, with further repercussions for financial stability and welfare. By applying our framework to a model of financial crises based on Morris and Shin (2000), we can characterize optimal stress tests in a more general environment where information is endogenous.

Beyond banking, our analysis sheds light on disclosure incentives in other markets. In corporate finance, for instance, firms publish verifiable accounting information while anticipating the release of public analyst reports and credit ratings. Also, producers can decide whether to acquire and display quality certificates, while their consumers have public access to third-party ratings and reviews. As we discuss below, the mechanisms we emphasize help to explain recent empirical findings on the correlation between product quality and disclosures, and yield new testable predictions for corporate disclosure policy.

Positive results.   The following example illustrates our argument. It is a special case of the Sender-Receiver model we develop in Section 2. Consider a population of banks, where the value of each bank's assets is a random variable θ, drawn independently across banks. Investors wish to roll over their loans to banks worth more than a threshold c, and to run on banks worth less than c.[2]

---

[1]Bouvard et al. study the problem of a regulator who chooses between revealing the portfolios of all banks and revealing nothing. Goldstein and Leitner (2015) allow the regulator to choose flexible information structures in the spirit of Bayesian Persuasion (Kamenica and Gentzkow, 2011), and Ely (2015) applies a model of dynamic Bayesian Persuasion to stress tests. These papers do not address voluntary disclosures by informed banks.

[2]More precisely, one can think of a coordination game between depositors in the Diamond and Dybvig (1983) model. According to the standard global games refinement (Morris and Shin, 2000), investors run on the bank whenever they expect asset values to fall below a threshold c. We explore this micro-foundation in Section 4.

Banks would like to convince investors to roll over. Communication works as follows: First, banks privately observe their quality, and decide whether to verifiably disclose it. Disclosure comes at a cost, but this cost is smaller than the benefit to the bank of avoiding a run. Second, a noisy public signal s of each bank's quality is observed by investors. Third, investors decide whether to run on their bank.

Suppose that we are in a financial crisis: The expected value of assets is low with $E[\theta] < c$, and without any further information, there would be a run on all banks. If public information is very noisy, then information unravels, as predicted by the classic literature (Grossman and Hart, 1980; Grossman, 1981; Milgrom, 1981; Jovanovic, 1982): Banks who are worth more than the threshold c choose to avoid a run by disclosing their quality. Banks worth less than c remain quiet, but face a run because investors interpret their silence as bad news. As a result, equilibrium outcomes are as if investors had perfect information.

If public information is sufficiently precise, then incentives change dramatically. Public signals above a certain threshold $s^\star$ reveal high quality and entice investors to roll over, even when their bank does not disclose anything. Consequently, the very best banks prefer to stay quiet: They expect that the public signal of their quality will likely exceed the hurdle rate $s^\star$ even without disclosure, so the marginal benefit of disclosure is small. Moreover, this reaction is amplified. When the best banks stay quiet, silence itself becomes better news, and yet more high-quality banks prefer to stay quiet. This feedback loop, which we dub 'reverse unraveling', amplifies opacity. There is much less disclosure in equilibrium than there would be without public information.

In Section 2 and 3, we show that 'reverse unraveling' arises in the general Sender-Receiver model. We also show that more informative public information can leave the Receiver less informed in equilibrium (in the sense of Blackwell's (1953) criterion).[3] Thus, more public information can increase uncertainty among uninformed agents.

These results are derived under relatively mild conditions on preferences and communication possibilities. The salient assumption is that verifiable disclosures are costly. Empirical studies consistently find that disclosures are costly, due to the direct costs of producing verifiable information (Lewis, 2011) and the indirect 'proprietary' costs of revealing sensitive information to competitors (Hayes and Lundholm, 1996; Harris, 1998; Berger and Hann, 2003).[4]

Normative results and stress test design. In our example, suppose that a policy-maker can conduct stress tests of banks' assets and publish their results. By publishing more or less detailed results,

---

[3]Blackwell's (1953) criterion calls signal s more informative than signal t if any Bayesian decision-maker would prefer observing s to observing t before taking her decision.

[4]For our positive analysis, it is not important whether disclosure costs are direct or proprietary, but the two may have different welfare implications, as we discuss in Section 5. A more complete literature review on disclosure costs is provided by Leuz and Wysocki (2008).

she can control the precision of public information to maximize welfare. Assume, for now, that bank managers have the right incentives: The costs of runs and of disclosure are truly social costs. Then, maximizing welfare is equivalent to minimizing a weighted sum of disclosure costs, the probability of bank runs, and the administrative costs of the test itself.

A useful benchmark is to think of banks' voluntary disclosures as fixed and unresponsive to policy, in line with the existing literature (Bouvard et al., 2015; Goldstein and Leitner, 2015). The policy-maker would like to protect weak banks (with values below c) from facing runs. This is achieved by introducing noise, i.e. by giving weak and strong banks the same stress test results some of the time. Under this scheme, strong banks insure weak banks against runs.[5] However, there are limits to insurance: If there is too much noise, investors learn nothing from stress tests, and proceed to run on all banks if $E[\theta] < c$. Thus, a minimal degree of informativeness is required in bad times.

We analyze how optimal policy changes when banks' voluntary disclosures are endogenous, which introduces further constraints on stress test design. We show that optimally, stress tests should be made more precise than in the benchmark in order to exploit the crowding out mechanism. To understand this result, consider a marginal increase in the precision of test results. This has a direct effect on investor behavior, which is present in the benchmark, and the additional indirect effect that more precise information crowds out disclosure by strong banks. Other things equal, the indirect effect is socially beneficial: When strong banks stay quiet, then silence is better news, which in turn enhances the insurance effect on weak banks. Consequently, it is worth making stress tests marginally more precise than in the benchmark to exploit the indirect effect on welfare.

In Section 4, we generalize this argument in a micro-founded model of financial crises. More-over, we examine the sensitivity of our conclusion to the assumption that bank managers have the right incentives. In economies with contracting frictions or reputational concerns, the managers of a failing bank might have excessive incentives to avoid liquidation, so that their perceived cost of liquidation overstates the social cost. In this case, the insurance benefit of stress tests loses impor-tance. Moreover, the perceived cost of disclosure might be due to the fear of revealing investment portfolios to competitors, the social costs of which are again less obvious. In that case, it becomes attractive to crowd in disclosures by making stress test results noisier.

Empirical implications. Our positive results relate to the common empirical finding that quality disclosures are not monotonic: The best and the worst agents tend to stay quiet, while mediocre ones disclose evidence (Feltovich et al., 2002; Luca and Smith, 2015; Bederson et al., 2015).[6]

---

[5]Although we think of stress test results as verifiable, this insurance mechanism qualitatively similar to the result that imprecise 'cheap talk' can be an optimal policy choice in Stein (1989).

[6]Feltovich et al. (2002) provide experimental evidence of this non-monotonicity. Luca and Smith (2015) show that mid-ranked business schools are most likely to disclose their rankings, while Bederson et al. (2015) find that restaurants

The 'countersignaling' hypothesis (Feltovich et al., 2002) is consistent with non-monotonicity: High quality types do not pay for costly signals (such as disclosure) if their quality is likely to be revealed by public signals. In the light of our analysis, the empirical relevance of this hypothesis is strengthened substantially. 'Reverse unraveling' amplifies non-disclosure at the top, and as long as the very best agents are enticed to stay quiet, we do not require very precise public information to generate substantial amounts of opacity.

In the case of the financial system, historical analyses suggest that asymmetric information is pervasive in financial crises. Banking panics, including the panic of 2008, are usually described as situations where (i) there is bad news about the banking system as a whole, and (ii) there is asymmetric information because investors do not know which individual banks are most exposed (Mishkin, 1990; Gorton, 2008). Our model provides a testable explanation of the fact that asymmetric information prevails in crises, even though disclosures in times of distress are feasible (Sowerbutts and Zimmerman, 2013), and public and regulatory data on individual banks are considered by investors (Jordan et al., 1999). The predictions which set our explanation apart from alternative stories are that (i) disclosures which resolve asymmetric information will be non-monotonic in the quality of banks' assets, and (ii) more precise public information, such as stress tests, will reduce voluntary disclosures, especially by high-quality banks.

Moreover, our framework can be applied to corporate finance, where it also yields new testable predictions. In particular, we predict that more precise or more frequent analyst reports and credit ratings lead to less thorough voluntary disclosures by corporations, and that this effect is concentrated on corporations with high underlying quality. We hope to test these predictions for banking and corporate finance formally in future research.

Related literature. Our 'reverse unraveling' argument describes a new mechanism which generates substantial amounts of non-disclosure. This is complementary to the literature on potential failures of the classic 'unraveling' result. In particular, by emphasizing the interaction between disclosure costs and public information, we add to the predictions of theories based on disclosure costs (Grossman and Hart, 1980; Verrecchia, 1983; Wagenhofer, 1990) or restrictions on communication possibilities (Okuno-Fujiwara et al., 1990; Hagenbach et al., 2014). Moreover, we complement theories based on uncertainty about the informed party's information (Dye, 1985; Shin, 1994, 2003; Acharya et al., 2011) by proposing a mechanism that generates non-monotonic disclosure strategies.[7]

The idea that more public information can lead to less information in equilibrium is related to the macroeconomic literature on dispersed information. Amador and Weill (2010) show that public

---

with top hygiene ratings did not post these ratings online. Moreover, Edelman (2011) finds that websites displaying trust certificates are less likely to be trustworthy on average.

[7]Milgrom (2008) and Dranove and Jin (2010) conduct more exhaustive literature reviews on disclosure.

information can reduce the emphasis which agents place on their dispersed private information, thus inhibiting the informational efficiency of the price system and exacerbating uncertainty (see also: Vives, 1997; Morris and Shin, 2005; Kohlhas, 2015). Our results show that neither this problem nor its relevance to welfare are confined to the price system, since public information can further reduce informativeness by crowding out the strategic disclosures of informed agents.[8]

In our financial application, we show that 'reverse unraveling' generates opacity in the banking sector. Existing theories of opacity in banking (Dang et al., 2015) attribute it to banks' desire to issue information-insensitive securities in good times. By contrast, we offer an explanation for why opacity prevails in bad times. Thus, we provide a potential foundation of macroeconomic models with financial frictions, which rely on asymmetric information to generate persistent downturns (e.g. Mankiw, 1986; Boissay et al., 2015; Heider et al., 2015).

The policy-maker in our normative analysis solves a problem which is akin to Bayesian persuasion (Kamenica and Gentzkow, 2011): Before she knows which banks are of high and low quality, she commits to a decision which influences the distribution of signals about quality.[9] We add the possibility of further disclosures once an informed agent has learned the state of the world. The policy maker needs to consider the interim incentives of informed agents when choosing signals ex ante. In this context, our policy problem can be thought of as a simple example of ex ante Bayesian Persuasion subject to interim incentive constraints, which complements the analysis of interim persuasion in Perez-Richet (2014).

Outline. In Section 2, we establish our main results in a general Sender-Receiver model. In Section 3, we further characterize the equilibrium implications of crowding out effects in a model where the Sender takes a binary action. In Sections 4 and 5 we apply our framework to a model of financial crises and study normative implications and the optimal design of stress tests. Section 6 concludes.

2   A Sender-Receiver model with public information

In this Section, we develop a model of the interaction between public information and private incentives to disclose evidence. The reader mainly interested in the finance application and stress test design can skip to Section 4, which is intended to be self-contained.

---

[8] The impact of public information in our model is also distinct from the effects discussed by Morris and Shin (2002) and Angeletos and Pavan (2007), where public information affects the outcome of a coordination game between private agents. However, a related effect is driving the multiplier induced by 'reverse unraveling', since there are strategic complementarities in informed agents' disclosure decisions.

[9] Unlike Kamenica and Gentzkow (2011), we consider only a restricted version where the policy-maker chooses a noise parameter, rather than the entire information structure.

Players and payoffs. We consider a game of communication between an informed *Sender* (he) and a less informed *Receiver* (she). The Receiver needs to decide on an *action* $a \in A$. Payoffs depend on this action and on the realization of a random variable $\theta \in \Theta$ which we call the Sender's *type*. The Sender's payoff is $v(a, \theta)$ and the Receiver's payoff is $u(a, \theta)$. Both v and u are continuous.

We assume that the Sender's payoff $v(a, \theta)$ is strictly increasing in a so that he always prefers high actions.[10] The Receiver's payoff $u(a, \theta)$ is log-supermodular in a and θ, so that she optimally takes higher actions when she is more optimistic about θ.[11] Let $a^\star(\theta) = \arg\max u(a, \theta)$ be the Receiver's preferred action under full information, and assume for simplicity that this action is unique. Moreover, we assume that the Receiver's preferences are non-trivial: $a^\star(\theta)$ is not the same for all θ.

In this Section, we assume that the sets A and Θ are finite subsets of $\mathbb{R}$.[12] For any finite set $X \subset \mathbb{R}$, we let $\underline{x}$ and $\bar{x}$ denote the smallest and largest elements of X. Thus, $\underline{a}$ is the lowest action, $\bar{\theta}$ is the highest type and so forth. Without loss of generality, we assume that $a^\star(\underline{\theta}) = \underline{a}$ and $a^\star(\bar{\theta}) = \bar{a}$.

Information. The Sender privately observes his type θ, which is drawn from a probability distribution with density $f(\theta)$ and support Θ. The Receiver observes only a noisy *public signal* s of θ, drawn from a conditional distribution with density $g(s|\theta)$ and support $S(\theta)$. We write $S = \cup_{\theta \in \Theta} S(\theta)$ for the set of all possible signals.

To model the possibility of communication, we assume that the Sender can send the Receiver a message m. The set of messages that type θ can send is finite and denoted $M(\theta)$.

We define two types of messages, cheap talk and verifiable disclosure. *Cheap talk* messages $m \in \cap_\theta M(\theta)$ can be sent independently of the Sender's type. *Verifiable disclosure* is a message which is not cheap talk, and therefore offer hard information about the type. A special case of disclosure is *full disclosure,* i.e. a message which can only be sent by a single type θ, and therefore certifies this type for certain. Verifiable disclosure is costly in that it reduces the Sender's utility by $\delta(\theta)$. The Sender's overall utility is thus $v(a, \theta) - \delta(\theta)$ if m is verifiable disclosure, and $v(a, \theta)$ if it is cheap talk. We say that a Sender who sends a cheap talk message *stays quiet.*

We assume further that cheap talk messages exists, $\cap_\theta M(\theta) \neq \emptyset$, and that each type has at least one full disclosure message which certifies θ for certain.[13] A common special case is the message

---

[10]Seidmann and Winter (1997) analyze a Sender-Receiver game where the Sender's preference is not monotonic in actions.

[11]More precisely, the Receiver's optimal action increases whenever her beliefs about θ become more optimistic in the sense of Milgrom's (1981) Monotone Likelihood Ratio Property. Athey (2002) characterizes the general relationship between log-supermodularity and comparative statics under uncertainty.

[12]Finiteness is assumed to avoid difficulties with the definition and existence of equilibria in extensive form games with infinite types and actions (see Myerson and Reny (2015)). Assuming subsets of $\mathbb{R}$ is convenient because it generates a natural ordering, but without loss of generality. Our arguments apply as long as $A \times \Theta$ is any finite lattice.

[13]These assumptions clarify the exposition, but could be weakened. We expect that the existence of certifying messages can be relaxed as long as the set of costly messages admits an *evidence base* in the sense of Hagenbach et al.

space introduced by Grossman and Hart (1980), who allow messages of the form 'my type belongs to X', where X can be any subset of Θ that contains the Sender's true type. In that case, $X = \Theta$ is cheap talk, all other messages are verifiable disclosure, and $X = \{\theta\}$ certifies θ.

Timing. The timing of the game is as follows: First, the Sender learns his type θ, and then sends the Receiver a message $m \in M(\theta)$. Second, the Receiver observes the message and the public signal, and then chooses her action. Figure 1 illustrates.

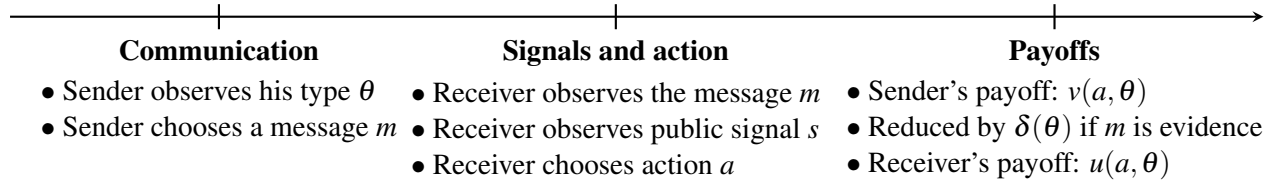| Communication	Signals and action	Payoffs
• Sender observes his type θ	• Receiver observes the message m	• Sender's payoff: $v(a, \theta)$
• Sender chooses a message m	• Receiver observes public signal s	• Reduced by $\delta(\theta)$ if m is evidence
	• Receiver chooses action a	• Receiver's payoff: $u(a, \theta)$

Figure 1: **Timeline of the Sender-Receiver model**

The key friction is that the Sender cannot reveal further verifiable information after the public signal is realized and before the Receiver takes her action. This assumption is motivated by natural frictions in communication. For example, when public news is revealed to financial markets, market participants (Receivers) react instantly, and firms (Senders) do not have time to prepare, audit and publish new financial reports in the meantime. Moreover, if the information in question in complex, it may not be possible for Receivers to process it before it is necessary to decide on an action.

Equilibrium definition. We consider *Perfect Bayesian Equilibria* (Fudenberg and Tirole, 1991). Perfect Bayesian Equilibrium requires (i) that the Receiver chooses her action optimally given her beliefs about the type θ after observing message m and signal s, (ii) that the Sender chooses his message optimally given the Receiver's strategy, and (iii) that the Receiver's beliefs be consistent with Bayes' rule on the equilibrium path. Off the equilibrium path, we restrict beliefs only by requiring that after observing a signal s, investors attach zero probability to types with $s \notin S(\theta)$ who are not physically able to send this signal.[14]

To be precise, a *strategy profile* $\sigma = (\sigma_S, \sigma_R)$ specifies the probability distribution over the Sender's messages $\sigma_S(\theta) \in \Delta M(\theta)$ for each type θ, and a probability distribution $\sigma_R(m, s) \in \Delta A$

(2014).

[14]Our notion of Perfect Bayesian Equilibrium is not necessarily equivalent to Sequential Equilibrium (Kreps and Wilson, 1982) in this setting. There is an interim move by Nature in revealing s and the Receiver has more than two types in general, so the equivalence result of Fudenberg and Tirole (1991) does not directly apply. However, all of our main results can be stated equivalently in terms of Sequential Equilibria.

over the Receiver's action, for each message m and each signal s.[15] A *belief* specifies a probability distribution $\mu(\theta|m,s) \in \Delta\Theta$ for the Receiver, given a message and a public signal.

Definition 1. A *Perfect Bayesian Equilibrium* is a pair (σ, μ) such that

1. The Sender's message is optimal with $\sigma_S(.|\theta) \in \arg\max_{\sigma \in \Delta M(\theta)} E[v(a,\theta)|\theta]$,

2. The Receiver's action is optimal with $\sigma_R(.|m,s) \in \arg\max_{\sigma \in \Delta A} E_\mu[u(a,\theta)|m,s]$,

3. Beliefs are calculated using Bayes' rule for all events (m,s) which occur with positive probability under the strategy profile σ; For zero probability events, $\mu(\theta|m,s) = 0$ if $s \notin S(\theta)$,

where the expectation operator $E[.|\theta]$ considers the probability distribution over $a \in A$ induced by σ, and $E_\mu[.|m,s]$ considers the joint distribution over $(a,\theta) \in A \times \Theta$ induced by σ and μ.

2.1 Reverse unraveling

We begin by establishing general results which formalize the arguments that (i) sufficiently precise public information crowds out incentives to disclose evidence, (ii) this effect is amplified by a 'reverse unraveling' mechanism. The following two concepts are central to our results. First, define as the set of signals s for which type θ is the *worst case,*

$$\underline{S}(\theta) = S(\theta)\backslash\{\cup_{\theta'<\theta}S(\theta')\}. \tag{1}$$

Intuitively, $s \in \underline{S}(\theta)$ if θ is the lowest type who sends signal s with positive probability.[16] Second, define the *maximal punishment,*

$$P(\theta) = \sum_{\theta'<\theta} Pr[s \in \underline{S}(\theta')|\theta] \times \left[v(a^\star(\theta),\theta) - v(a^\star(\theta'),\theta)\right]. \tag{2}$$

$P(\theta)$ captures the largest loss in expected utility that a Sender with type θ can experience by staying quiet instead of fully disclosing his type. It is based on a set of beliefs for which the Receiver 'assumes the worst'. For each realization of the public signal s, these beliefs attach probability one to the lowest possible type who could have sent that signal. If θ' is the lowest type who can send a signal s (that is, $s \in \underline{S}(\theta')$), then a Receiver assuming the worst chooses $a^\star(\theta')$. The loss in utility for the Sender, relative to disclosing his true type θ, is $v(a^\star(\theta),\theta) - v(a^\star(\theta'),\theta)$. The expression for $P(\theta)$ takes the expected value of this loss.[17]

---

[15]ΔX denotes the set of probability distributions over a set X.

[16]Note that the sets $\{\underline{S}(\theta)\}_{\theta\in\Theta}$ form a partition of the set of all possible signals: $\underline{S}(\theta)$ is disjoint from $\underline{S}(\theta')$, and $\cup\underline{S}(\theta) = S$.

[17]When taking the expected value in (2), we only need to sum over types $\theta' \leq \theta$. For $\theta' = \theta$, the loss in utility is zero. For types $\theta' > \theta$, we have $P[s \in \underline{S}(\theta')|\theta] = 0$ by definition.

The maximal punishment $P(\theta)$ is closely related to the precision of public signals. Intuitively, precise signals tend to reveal the true quality to the Receiver, which limits the extent to which the Sender can suffer from non-disclosure. To understand why, suppose that the public signal is $s = \theta + \sigma \varepsilon$, where $\varepsilon \in [-1, 1]$ is random error and σ is a noise parameter. Thinking in terms of continuous signals and types for simplicity, the signal for which type θ' is the worst case is $\underline{S}(\theta') = \{\theta' + \sigma\}$. The probability that the worst case is θ' or lower, given the true type θ, is $Pr[\varepsilon \leq 1 - \frac{\theta - \theta'}{\sigma}]$, which is increasing in σ. Thus, noisier signals lead to more severe worst case beliefs, which increases the maximal punishment for staying quiet. Conversely, more precise signals decrease the punishment.

We now show that the maximal punishment $P(\theta)$ and the disclosure cost $\delta(\theta)$ are sufficient statistics for the equilibrium behavior of high-quality types. This generalizes the notion that high-quality types will stay quiet when signals are precise enough, which we presented in the Introduction.

Proposition 1. *Let $\theta_P = \max\{\theta : P(\theta) \geq \delta(\theta)\}$ be the highest type for which the maximal punishment exceeds the cost of disclosure. If $\theta_P < \bar{\theta}$, then types $\theta > \theta_P$ stay quiet with probability one in any equilibrium.*

In other words: If there is a group of consecutive types at the top for whom the maximal punishment is strictly less than the cost of disclosure, then all of these types stay quiet in equilibrium. This follows by iterated deletion of strictly dominated strategies. Suppose that there are n types at the top ($\theta > \theta_P$) for whom we have $P(\theta) < \delta(\theta)$. This includes the best type $\bar{\theta}$ and $(n-1)$ types below him. First, consider the best type $\bar{\theta}$. If this type decides to make a verifiable disclosure, then he can do no better than to fully disclose that he is the best, since doing so induces the Receiver to take the highest possible action $\bar{a}$. The most he can lose by staying quiet, therefore, is the expected loss he would suffer if the Receiver switched from believing that he was the best to assuming the worst. This loss is exactly the maximal punishment $P(\bar{\theta})$. However, by assumption, this loss does not justify the cost of disclosure. In this case, the best type has a dominant strategy to stay quiet.

Second, having deleted disclosure from the best type's possible strategies, consider the second-best type. Again, if he decides to make a verifiable disclosure, then he can do no better than full disclosure: Indeed, no disclosure can ever convince the Receiver that he might be facing best type, because everybody knows that the best type's strategy is to remain quiet. Therefore, the best possible disclosure the Sender can make is to certify that he is the second-best. The above argument goes through as before, and the second-best type prefers to stay quiet as well. The formal proof in the Appendix extends this argument to all types $\theta > \theta_P$ by induction.[18]

---

[18]Iterated deletion is needed in this proof only to ensure that the second-best type, say, cannot make a verifiable disclosure which allows him to pool with $\bar{\theta}$. If such pooling were possible, then the the expected loss from staying quiet could be more than the maximal punishment $P(\theta)$. If actions are continuous and strictly increasing in the Receiver's beliefs, then a simpler argument is sufficient. For example, consider the Buyer-Seller model of Grossman

Under what circumstances should we expect a group of types at the top who prefer to stay quiet? The connection between the maximal punishment $P(\theta)$ and signal precision implies that such a group is more likely to exist when signals are precise. Moreover, the following simple condition is grossly sufficient to ensure that $\theta_P < \bar{\theta}$:

$$\sum_{\{\theta':a^\star(\theta)\neq\bar{a}\}} Pr[s \in S(\theta')|\bar{\theta}] < \frac{\delta(\bar{\theta})}{v(\bar{a},\bar{\theta}) - v(\underline{a},\bar{\theta})}.$$

This inequality ensures that the best type $\bar{\theta}$ prefers to stay quiet. The left-hand side is the probability that the best type receives a public signal which could have come from a type that does not obtain the highest action $\bar{a}$ under full information. This probability falls when signals become more precise. The right-hand side is the cost of disclosure $\delta(\bar{\theta})$, divided by the largest possible benefit of disclosure. This ratio rises when disclosures become more costly, or when the best type becomes less concerned about which action he obtains.[19]

So far, we have characterized the strategies of high quality types $\theta > \theta_P$, who have an incentive to stay quiet even when the Receiver assumes the worst. However, this implies that it is no longer rational for the Receiver to assume the worst when she faces a quiet Sender, since she must realize in equilibrium that the highest quality types are among the pool of quiet Senders. We show that further types $\theta \leq \theta_P$ are tempted to stay quiet once we correct the Receiver's beliefs accordingly. This captures the 'reverse unraveling' effect we described in the Introduction.

We need two further definitions to state this result. First, we define the *density* of types as

$$\zeta = \left(\max_{\theta,\theta'} \|z(\theta) - z(\theta')\|\right)^{-1}, \tag{3}$$

where $z(\theta) = (u(.|\theta), v(.|\theta), \delta(\theta), g(.|\theta))$ is a vector collecting type θ's attributes, and $\|.\|$ is the Euclidean norm.[20] The density measures the 'closeness' of types, and loosely speaking, a high density represents a type space which is almost continuous.[21]

Second, we say that types above θ_P are *minimally persuasive* if for there exists a signal $s \in$

(1981), where the Buyer's continuous 'action' is the market price, which is simply the expected value of the Seller's good. Any pooling – among those types making verifiable disclosures – would create incentives for better sellers to marginally increase the market price by switching to full disclosure. Thus, our argument would apply without iterated deletion.

[19]The latter effect is common in models with type-dependent outside options such as Akerlof (1970). Suppose that the seller of a good, for example, has an outside option to consume the good himself, and that the utility he derives from consuming it himself is increasing in his quality. Then, high-quality sellers care less about market offers, implying that $u(\bar{a},\bar{\theta}) - u(\underline{a},\bar{\theta})$ is low.

[20]The length of $z(\theta)$ is $2(\#A) + (\#S) + 1$, where $\#X$ denotes the cardinality of any set X.

[21]For example, we can continually fill in the type space by taking all pairs of neighboring types (θ,θ') and creating an 'average type' $\theta'' = (\theta + \theta')/2$ with attributes $z(\theta'') = (z(\theta) + z(\theta'))/2$ in between. As we repeat this exercise, the type space will approach continuity and the density $\zeta \to \infty$.

$S(\theta_P)$ such that

$$\arg\max_{a \in A} E\left[u(a, \theta) \mid s, m \in M^c; \sigma_S'\right] > \arg\max_{a \in A} E\left[u(a, \theta) \mid s, \theta = \theta'\right] \tag{4}$$

where σ_S' is any alternative disclosure strategy for the sender, in which types $\theta > \theta_P$ stay quiet with probability one. To interpret this, consider a Receiver who observes a quiet Sender. If she does assume the worst, she takes the action on the right-hand side of (4). But if she is forced to acknowledge that high quality types $\theta > \theta_P$ also stay quiet, she takes the action on the left-hand side of (4). If types above θ_P are minimally persuasive, acknowledging that these types stay quiet leads the Receiver to change her mind with positive probability. We now derive a condition under which reverse unraveling occurs.

Proposition 2. *Suppose that $\theta_p < \bar{\theta}$, and that types above θ_P are minimally persuasive. If the state space is sufficiently dense, then there is a non-empty set of types $\theta \le \theta_P$ who stay quiet with probability one in any equilibrium.*

The logic of the proof is as follows. Suppose that types above θ_P are minimally persuasive. For types close to θ_P, there is a positive probability that the Receiver will change her mind once she acknowledges that types above θ_P stay quiet (which is the case in equilibrium, according to Proposition 2). For these types, the maximal punishment from staying quiet is now falls below $P(\theta)$: The worst case scenario improves because the Receiver might change her mind. Finally, note that types 'just above' θ_P satisfy $\delta(\theta) > P(\theta)$ by definition. If the state space is sufficiently dense, then types 'just below' have $\delta(\theta) \simeq P(\theta)$, and when the maximal punishment falls below $P(\theta)$, such types will strictly prefer to stay quiet.

This argument captures reverse unraveling because types below θ_P don't have a strong fundamental incentive to stay quiet, but do so nonetheless. They are swayed by the opportunity to pool with types who are better than themselves. When the state space is dense, this pooling effect is meaningful because types close to θ_P are likely to send similar signals to those above them.

2.2 Informativeness

Our final result in this Section shows that more public information can lead to less information for the Receiver in equilibrium. Given the crowding out effects we have described, the basic idea is simple: Better public information constitutes an obvious increase in informativeness for the Receiver, but an indirect decrease if it crowds out disclosures by the Sender. We show that the indirect effect can dominate in a strong sense, namely in Blackwell's (1953) informativeness order.

A signal x is more informative than another signal y in the sense of Blackwell (1953) if any utility maximizing decision-maker, whose utility depends on an underlying state θ, would prefer

having access to x over having access to y. Blackwell's theorem shows that this notion of informativeness is equivalent to y being a 'garbled' version of x, in that there exists a garbling function $\gamma(y|x)$ such that $\sum_y \gamma(y|x) = 1$ and

$$Pr[y|\theta] = \sum_x \gamma(y|x)Pr[x|\theta].$$

In the context of our model, a *Blackwell-improvement in public information s* occurs if the Receiver instead observes a signal $\hat{s}$, with conditional distribution $\hat{g}(\hat{s}|\theta)$, which is more informative than s. This requires that $g(s|\theta) = \sum_{\hat{s}} \gamma_s(s|\hat{s})\hat{g}(\hat{s}|\theta)$ for some function γ_s such that $\sum_s \gamma_s(s|\hat{s}) = 1$.

The information available to the Receiver in equilibrium is the pair of random variables (s, m). Recall that $\sigma_S(m|\theta)$ is the probability that a Sender of type θ sends message m. Since the Sender chooses m before he observes s, the message is conditionally independent of the signal. Hence, the joint distribution of (s, m) is

$$Pr[s, m|\theta] = \sigma_S(m|\theta)g(s|\theta) \equiv \tau(m, s|\theta).$$

A *Blackwell-deterioration in the Receiver's information* occurs if the Receiver instead observes an information structure $\{\hat{s}, \hat{m}\}$, with conditional distribution $\hat{\tau}(\hat{s}, \hat{m}|\theta)$, which is less informative than $\{s, m\}$. This requires that $\hat{\tau}(\hat{s}, \hat{m}|\theta) = \sum_{s,m} \gamma_{s,m}(\hat{s}, \hat{m}|s, m)\tau(s, m|\theta)$ for some function $\gamma_{s,m}$ which satisfies the adding-up constraint $\sum_{\hat{s}, \hat{m}} \gamma_{s,m}(\hat{s}, \hat{m}|s, m) = 1$.

Since there are potentially multiple equilibria in this model, it is difficult to perform equilibrium comparative statics with respect to informativeness in general. For the sake of clarity, we focus on parametric regions where a transparent equilibrium exists. A transparent equilibrium is one in which every type, except for types who receive the worst action under full information, fully discloses his information. This arises naturally, for example, when disclosure costs $\delta(\theta)$ are sufficiently small. In the next Section, we show that such equilibria also arise under weaker conditions when the Receiver takes a binary action.

Proposition 3. *If a transparent equilibrium exists, then there exists an alternative public signal $\hat{s}$ which induces (i) a Blackwell-improvement in public information relative to $\hat{s}$ and (ii) a Blackwell-deterioration in the Receiver's information in any equilibrium, relative to the most informative equilibrium of the initial game in which s is the public signal.*

This Proposition relies on the crowding out effect we characterized in Proposition 1. Starting from a full disclosure equilibrium, there is always a Blackwell-improvement in public information which gives the best type $\bar{\theta}$ a dominant strategy to stay quiet. To achieve this, we make signals sent by high quality types more precise, until the disclosure costs for the best type exceed the maximal punishment, $\delta(\bar{\theta}) < P(\bar{\theta})$. In any equilibrium after this change, the best type will stay quiet in

equilibrium. Indeed, if this type is minimally persuasive, then other types – at the top but below $\bar{\theta}$ – will stay quiet too. On balance, the Receiver now observes strictly noisier information than before about types that stay quiet, and no more information about types that continue to disclose. As a result, there is a Blackwell-deterioration in the Receiver's information.

This logic does not rely on the initial equilibrium exhibiting full disclosure, but without imposing more structure on preferences, we cannot characterize equilibrium play tightly enough to perform more general comparative statics. In the next Section, we re-state a more comprehensive version of Proposition 3 for models in which the Receiver takes a binary action.

3 Binary actions

In this Section, we study a version of the Sender-Receiver model in which the Receiver takes a binary action $a \in \{0, 1\}$. One example of binary actions is the bank run scenario discussed in the Introduction. Other examples with binary actions include legal judgments (guilty or not guilty), corporate investment decisions (invest in a project or not), employment decisions (hire an applicant or not) or college admissions. Binary actions allow us to study the 'reverse unraveling' mechanism and its implications in more detail. We can fully characterize equilibrium disclosure strategies and extend our analysis to the case where public signals have full support. Moreover, we further develop the informativeness result of Proposition 3, and characterize comparative statics with respect to signal precision in terms of a 'multiplier' which captures reverse unraveling.

We assume that the Sender's type θ is a continuous random variable with density $f(\theta)$ and support $\Theta = [\underline{\theta}, \overline{\theta}] \subset \mathbb{R}$. The public signal is $s = \theta + \sigma \varepsilon$, where ε is a continuous random variable with density $h(\varepsilon)$, cumulative distribution $H(\varepsilon)$ and support $[\underline{\varepsilon}, \overline{\varepsilon}] \subset \mathbb{R}$. The parameter $\sigma > 0$ captures the amount of noise in public information. The density $h(\varepsilon)$ is log-concave, so that a high s is good news about θ in the sense of Milgrom's (1981) Monotone Likelihood Ratio Property.

All densities are continuously differentiable in θ and s, the bounds of Θ and $S(\theta)$ may be infinite. $S(\theta) = [\theta + \sigma\underline{\varepsilon}, \theta + \sigma\overline{\varepsilon}] \equiv [\underline{s}(\theta), \overline{s}(\theta)]$ denotes the set of type θ's possible signals, and $S = [\underline{s}(\underline{\theta}), \overline{s}(\overline{\theta})]$ is the set of all possible signals.

The Sender's preferences are described by the costs of disclosure $\delta(\theta)$ and the function $\beta(\theta) = v(1, \theta) - v(0, \theta) > 0$, which captures type θ's marginal benefit from obtaining the high action instead of the low one. We assume that $\delta(\theta) < \beta(\theta)$.[22] The Receiver's preferences are summarized by the function $\gamma(\theta) = u(1, \theta) - u(0, \theta)$, which captures her marginal utility from taking the high action if the Sender is of type θ. Note that $\gamma(\theta)$ is strictly increasing in θ because we have assumed that the Receiver's utility is log-supermodular. We let $c \in (\underline{\theta}, \overline{\theta})$ denote the type for which she is

---

[22]If disclosure costs outweigh the benefit of obtaining the high action, then the Sender never wishes to disclose. By imposing $\delta(\theta) < \beta(\theta)$, we focus on the more interesting case where non-disclosure is driven by strategic forces.

indifferent, defined by $\gamma(c) = 0$.

For simplicity, we assume that the Sender sends a binary message $m \in \{\emptyset, \theta\}$ which corresponds to either staying quiet or fully disclosing his type.[23] Moreover, we impose the following regularity condition on the Sender's preferences.

Assumption. *For all values of the public signal s_0, the function*

$$J(\theta) = H\left(\frac{s_0 - \theta}{\sigma}\right) - \frac{\delta(\theta)}{\beta(\theta)}$$

crosses zero at most once on the range $\theta \in [\underline{\theta}, \bar{\theta}]$. If it crosses once, then it crosses from above.

The function $J(\theta)$ in our regularity condition compares two terms. The first term is the probability of receiving a public signal s in the left tail, given that the true state is θ. This is strictly decreasing in θ, since high types are likely to receive good news. The second term measures the ratio of the cost of disclosure to the benefit of obtaining the high action. This is not necessarily monotonic in θ, but it is always less than one. The single crossing property we require holds under reasonable conditions. First, when disclosure costs are proportional to marginal benefits, then $\delta(\theta)/\beta(\theta)$ is a constant, and the function $J(\theta)$ is strictly decreasing. This arises naturally when disclosure costs are interpreted as the proprietary cost of revealing a firm's business model to competitors.[24] Second, when disclosure costs are fixed, the condition holds if public signals are precise enough. For example, if the noise σ in the public signal is small enough, then $G(s_0|\theta)$ is close to one for types $\theta < s_0$ and close to zero for types $\theta > s_0$. Since the relative cost term satisfies $0 < \delta/\beta < 1$, the difference between this probability can only have one crossing with zero.

3.1 Reverse unraveling

We begin by showing that disclosure strategies have a cutoff property: The best and the worst banks stay quiet in equilibrium, while *marginal types* disclose. This generalizes the Intuition of our bank run example in the Introduction. To economize on notation, we suppose that when the players are indifferent, the Sender discloses and the Receiver takes the high action. This restriction has no meaningful impact on equilibrium outcomes since it affects only events with measure zero.

Lemma 1. *In any equilibrium, there is a cutoff θ_1 such that the Sender discloses his type if and only if $\theta \in [c, \theta_1]$.*

---

[23] Since types who pay the cost of disclosure tend to fully reveal their quality, and the cheap talk messages of quiet types do not convey information, this is without loss of generality in a wide class of models.

[24] For example, consider an industry where the Sender is a monopolist with constant marginal costs, and revealing θ publicly triggers entry by a competitor with probability a. If the monopoly profit is $\beta(\theta)$ and firms engage in Bertrand competition upon entry, then the expected loss from revealing θ is $a\beta(\theta)$.

We establish this Lemma in two steps. First, we show that regardless of the Sender's disclosure strategy, the Receiver's best response is to take the high action if the public signal lies above a threshold $s \geq s^\star$. Intuitively, the Receiver takes more optimistic actions after good news. Formally, this is a consequence of the Monotone Likelihood Ratio Property.

Second, we show that as a best response, the best and the worst types of Sender will stay quiet. Bad types with $\theta < c$ have a dominant strategy to stay quiet: Disclosure is costly and leads to the low action for sure, while staying quiet allows them to free-ride with good types who stay quiet. Among good types with $\theta > c$, the best have the strongest incentive to stay quiet, since they are confident that they will receive a public signal $s \geq s^\star$. Hence, the only types that will disclose are mediocre types in an interval $[c, \theta_1]$, who are anxious that they will not receive a good enough public signal and thus prefer to guarantee themselves a favorable action by means of full disclosure. The single crossing assumption on $J(\theta)$ establishes the cutoff property by guaranteeing that the best types have the strongest incentive to remain quiet.

Lemma 1 allows for *transparent equilibria* in which all Senders with $\theta \geq c$ disclose, so that $\theta_1 = \overline{\theta}$. In transparent equilibria, outcomes are as if the Sender had perfect information. Moreover, Lemma 1 allows for opaque equilibria without any disclosure, which are denoted by $\theta_1 = c$.

In the general model, Proposition 1 and 2 established a 'reverse unraveling' mechanism: When signals are precise enough, the Sender stays quiet in equilibrium if he is one of the best types, and this entices further types to stay quiet because silence becomes better news in the Receiver's eyes. Here, we can characterize this mechanism more directly. For this purpose, we define the *overlap* of type θ as

$$L(\theta) = H\left(\frac{\overline{s}(c) - \theta}{\sigma}\right).$$

The overlap $L(\theta)$ measures the probability that type θ sends a signal which could have come from a type below the threshold c. This is closely related to the concept of maximal punishment in Section 2: $L(\theta)$ measures the probability that type θ will obtain the low action if he stays quiet, given that the Receiver interprets the public signal in the most negative way. The overlap also inherits the relationship to signal precision. If signals are precise, then $L(\theta)$ is low for good types, because it becomes less likely that good and bad types send the same signals. Figure 2 further illustrates the concept of overlap and its relationship with signal precision.
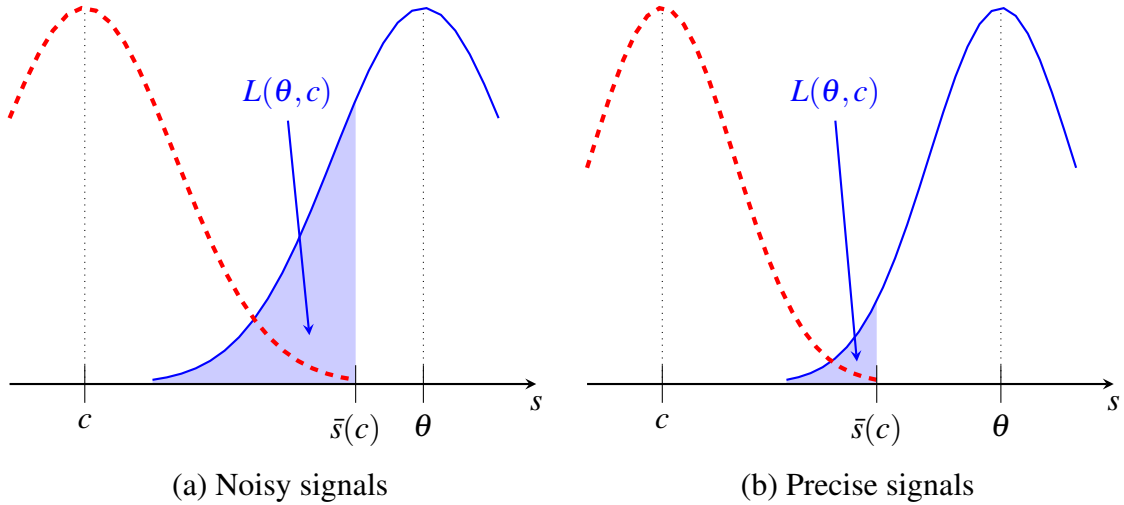
Figure 2: **Overlap**. The two panels show the overlap $L(\theta)$ for different signal precisions. The dashed (red) line is the density of signals sent by type c, for whom the Receiver is indifferent between high and low actions. This type sends signals up to $\bar{s}(c) = c + \sigma\bar{\varepsilon}$. The solid (blue) line is the density of signals sent by an other type $\theta > c$. The overlap $L(\theta)$ is the probability in the tail for which type θ's signal could have been drawn from type c's distribution. In panel (a), signals are noisy (large σ) and the overlap is large. In panel (b), signals are more precise and the overlap is small.

As a result of Lemma 1, the task of finding equilibria reduces to finding a cutoff θ_1 which is consistent with optimality. We show that there is a relationship between overlap and θ_1 which captures the idea of reverse unraveling in its strongest form.

Proposition 4. *Suppose that the support of ε is bounded. Then a full information equilibrium with $\theta_1 = \bar{\theta}$ exists if and only if*

$$L(\bar{\theta}) \geq \frac{\delta(\bar{\theta})}{\beta(\bar{\theta})}, \tag{5}$$

If (5) is violated, then there is a type $\tilde{\theta} < \bar{\theta}$ such that $\theta_1 \leq \tilde{\theta}$ in any equilibrium.

The first part of the Proposition shows that the best type $\bar{\theta}$ stays quiet in equilibrium if their overlap is low, i.e. when signals are precise enough so that (5) is violated. The second part shows that, whenever (5) is violated, the highest disclosing type is bounded strictly away from $\bar{\theta}$. Reverse unraveling occurs in its strongest form occurs at the threshold where (5) holds with equality. At this point, a fully transparent equilibrium exists, but even an infinitesimal improvement in signal precision (or reduction in overlap) leads to a discrete downward jump in equilibrium disclosures. This is because a small improvement at the threshold entices the best type to stay quiet, which in turn makes silence better news and gives further types near the top an incentive to stay quiet.

Proposition 4 is based on a fixed point equation which determines equilibrium. Let the function $BR(\theta_1)$ denote the Sender's *best response*, i.e. the highest type above c who is willing to disclose given that the Receiver expects him to use the cutoff strategy 'disclose if $\theta \in [c, \theta_1]$'.[25] Rational expectations in equilibrium require that

$$\theta_1 = BR(\theta_1). \tag{6}$$

The best response $BR(\theta_1)$ is increasing in θ_1 due to strategic complementarities between different types of the Sender: If more good types above the threshold c are expected to disclose, then quietness becomes a worse signal. Indeed, the marginal Receiver, who observes the critical signal $s^\star$ and believes that the value of taking the high action $\gamma(\theta)$ is zero on average, considers it bad news when types above c with $\gamma(\theta) > 0$ are removed from the quiet region. Hence, it becomes tougher for a quiet Sender to obtain the high action, and more good types prefer to disclose in response. Figure 3 further illustrates this effect.
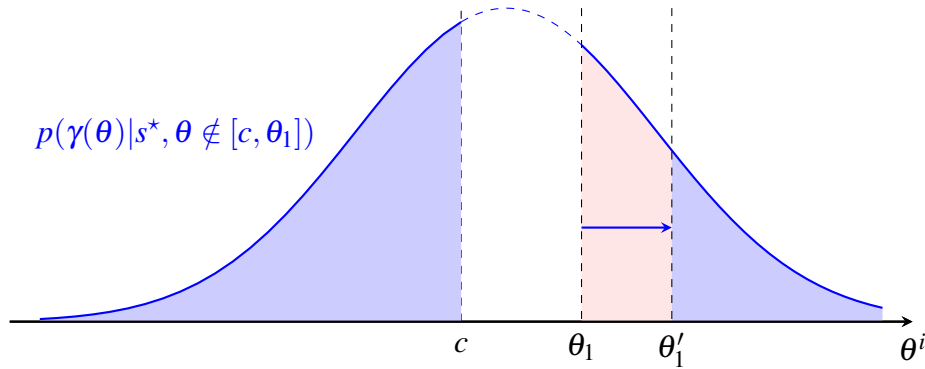
Figure 3: **The marginal Receiver and strategic complementarities**. The solid (blue) line is the posterior density of the Receiver's marginal benefit $\gamma(\theta)$ when she (i) observes the critical signal $s^\star$, (ii) observes that the Sender stays quiet, and (iii) believes that the Sender's strategy is to stay quiet if and only if $\theta \notin [c, \theta_1]$. The mean of this distribution is c. When θ_1 increases to θ_1', this removes above-average types (the light shaded area) from the Receiver's posterior. Therefore, she becomes less optimistic.

Figure 4 illustrates the fixed point problem (6). Equilibrium cutoffs θ_1 lie at the intersections of the best response $BR(\theta_1)$ and the 45-degree line. In panel (a), there is a full information equilibrium, since $BR(\overline{\theta}) = \overline{\theta}$, while in panel (b) there is not.

---

[25]The formal proof considers the possibility of beliefs off the equilibrium path, in which case $BR(\theta_1)$ may be a correspondence.

Figure 4: **Best responses and equilibrium.** The two panels show equilibria for different levels of public signal precision and overlap $L(\overline{\theta})$. In panel (a) signals are noisy, overlap is high and condition (5) holds. Thus there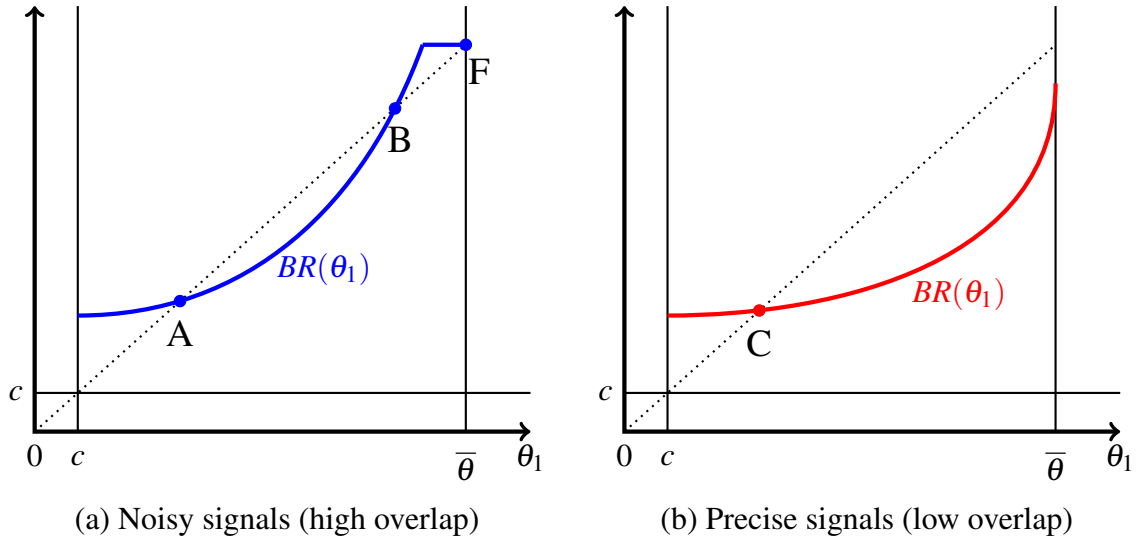 is an equilibrium with full information at point F. In panel (b), signals are precise and (5) fails. Thus there is only an equilibrium with partial disclosure at point C. In panel (a), the equilibrium at B is unstable while the equilibria at A and F are stable.

The proof of Proposition 4 is completed in two steps. First, we show that the best response touches the 45-degree line at $\overline{\theta}$, creating a full information equilibrium, if and only if (5) holds. Second, we show that whenever (5) does not hold, the best response is very steep near $\overline{\theta}$, so that we can bound the any equilibrium – such as point A in panel (b) – away from the best type $\overline{\theta}$.

3.2 Signal distributions with full support

We have emphasized cases where the overlap between different types' signals is limited. If any type can send any signal, then (5) is guaranteed to hold and there is always a transparent equilibrium. Intuitively, the transparent equilibrium in this case is sustained by skeptical beliefs: Whenever she faces a quiet Sender, the Receiver assumes that the Sender's type must be below the threshold c, and accordingly takes the low action. These beliefs are reasonable with full support because public signals cannot prove beyond reasonable doubt that a quiet Sender has a type above θ. We now argue that reverse unraveling is important even with full support. Thus we focus on the case where the support of the noise in the public signal is unbounded with $\underline{\varepsilon} = -\infty$ and $\overline{\varepsilon} = +\infty$.

We analyze the stability properties of equilibria. For an intuitive notion of stability, we return to panel (a) of Figure (6). Equilibrium cutoffs θ_1 where the best response crosses the 45-degree line from below are not stable. For example, consider the middle equilibrium at point B. If the

Receiver started expecting slightly less disclosure by good types, this would be self-fulfilling, since the Sender's best response would indeed be to disclose less. By a process of iterated best responses, the system would converge back to the stable equilibrium at A.

Accordingly, we say that a solution to 6 is a *stable equilibrium* if the best response curve crosses the 45-degree line from above at θ_1. This definition of stability is heuristic and is based on Schelling's (1978) 'diagrammatics of critical mass'. In Appendix B, we formalize our definition of stability in terms of perturbations and best response dynamics.

A corollary of Proposition 4 is that full information equilibria are always stable if the noise ε has bounded support (except in the knife-edge case where (5) holds with equality). With unbounded support, this changes and we obtain the following condition, which for the sake of expositional clarity we present for the case where $\delta(\theta) \propto \beta(\theta)$.[26]

Proposition 5. *Full transparency is an unstable equilibrium for all disclosure costs $\delta(\theta) > 0$ if and only if for any $K \in \mathbb{R}$, $\exists \tilde{\theta}'$ such that $\forall \theta' \geq \tilde{\theta}'$:*

$$-\frac{Pr(\theta \geq \theta'|s = \theta' + K).E[\gamma(\theta)|\theta \geq \theta', s = \theta' + K]}{Pr(\theta \leq 0 \mid s = \theta' + K).E[\gamma(\theta)|\theta \leq 0, s = \theta' + K]} > 1 \tag{7}$$

We establish Proposition 5 by considering a small deviation from a transparent equilibrium. Instead of expecting every type $\theta \geq c$ to disclose, the Receiver mistakenly expects an small portion of high quality types $[\theta_1, \overline{\theta}]$ to stay quiet. This implies that very high signals have the potential to convince the Receiver to take the high action, even if the Sender stays quiet. If signals are precise enough in the sense of condition (7), then the Receiver's mistake becomes self-fulfilling, since types $\theta \geq \theta_1$ are confident to receive a high public signal and prefer to stay quiet given the new set of beliefs. As a result, the small deviation is followed by the familiar reverse unraveling mechanism: When types above θ_1 stay quiet, then yet more types stay quiet because silence has become better news, and so forth until convergence.

More concretely, the condition (7) states that the critical signal required of quiet Senders after the deviation is not unreasonably large, even as we consider the limit of arbitrarily small deviations and let θ_1 converge to $\overline{\theta}$. If the critical signal s^* does not grow as fast as θ_1, type θ_1 believes that the probability of R playing 0 when he is quiet is a very unlikely tail event since $H\left(\frac{s^* - \theta_1}{\sigma}\right)$ becomes small. Proposition 5 identifies necessary and sufficient conditions under which the marginal signal fails to 'keep up' with θ_1. To see why condition (7) is important, suppose that θ_1 were arbitrarily increased and consider the effect on the Receiver's incentive to take the high action. whenever the public signal satisfies $s \geq \theta_1 + K$ for some $K \in \mathbb{R}$. Under these circumstances, the Receiver will

---

[26]A formal statement for, and proof of, the result when $\delta(\theta)$, $\beta(\theta)$ are arbitrary is available on request.

take the high action if

$$E\left[\gamma(\theta)\mid\theta\notin[c,\theta_1],s\right] = Pr[\theta<c|\theta\notin[c,\theta_1],s]E\left[\gamma(\theta)\mid\theta<c,s\right]$$
$$+Pr[\theta>\theta_1|\theta\notin[c,\theta_1],s]E\left[\gamma(\theta)\mid\theta>\theta_1,s\right]>0.$$

Equivalently, the Receiver will take the high action if her posterior-weighted expectation that the state is from the 'good' interval, $[\theta_1,\overline{\theta})$ exceeds the weighted expectation that it is from the 'bad' interval, $[\underline{\theta},c)$. In other words, under condition (7), R will strictly prefer to keep investing if the signal which satisfies this grows at the same rate with θ_1. Therefore, she needs only a lower s to substitute against the loss of further 'good' types as θ_1 increases.

When θ and ε are jointly Normally distributed with $\theta\sim N\left(\mu_\theta,\sigma_\theta^2\right)$ and $\varepsilon\sim N\left(0,\sigma^2\right)$, condition (7) has a particularly natural interpretation. In this case, (7) holds if and only if the signal-to-noise ratio is greater than one, $\sigma_\theta^2>\sigma^2$. Intuitively, when the signal-to-noise ratio is greater than 1, the Receiver puts a lot of weight on her signals, s. In such circumstances, observing a high signal more than offsets the Receiver's concern that the 'quiet' signal gets worse as θ_1 increases. Since the Receiver does not require large increases in signal to compensate for higher θ_1, then for sufficiently high θ_1, the cost of staying quiet becomes small and reverse unraveling is bound to occur.

3.3 Informativeness

We reconsider the effect of an improvement in the public signal on the information available to investors in equilibrium. Improvements and deteriorations of information are defined in terms of Blackwell's (1953) criterion: Signals x are more informative than signals y if any utility maximizing decision maker whose payoff depends on θ would prefer having access to x over having access to y. (See Section 2.1 for a formal definition.)

Proposition 6. *All equilibria can be ranked according to Blackwell's criterion. Moreover, there is a feasible Blackwell-improvement in public information which induces a Blackwell-deterioration in investors' information in the most informative equilibrium.*

In the binary actions model, the Sender discloses his type if he is mediocre with $\theta\in[c,\theta_1]$. If there are several equilibria, then ones with higher cutoffs θ_1 are strictly more informative: The Receiver obtains more information about mediocre types, and the same information (i.e. the public signal s) about the best and the worst types. Thus, we can rank equilibria according to Blackwell's criterion.

We establish the second part of the proposition by considering the cutoff type θ_1 in the most informative equilibrium. We construct an improvement in the precision of the public signal which focuses on an interval of types $[\theta_1-x,\theta_1]$ just below the cutoff, and entices these types to stay

quiet. For example, this can be done by introducing an additional public signal which reveals these types for certain with some probability. As a result, the Receiver becomes less informed about a group of types around θ_1 who used to engage in full disclosure, and no better informed about any other types. Hence, her information has strictly deteriorated.

3.4 Comparative statics and multipliers

To further characterize the impact of signal precision on equilibrium disclosures, we consider some local comparative statics. For the sake of clarity, we make the dependence of equilibrium strategies on signal precision more explicit: Let $s^\star(\theta_1; \sigma)$ be the critical value of the signal below which the Receiver takes the low action if she expects disclosure by types $\theta \in [c, \theta_1]$, and $BR(\theta_1; \sigma)$ the highest type of Sender who wishes to disclose given these beliefs.

Proposition 7. *The marginal effect of increasing the noise parameter σ on the disclosure cutoff in an interior equilibrium with $\theta_1 \in (c, \bar{\theta})$ is*

$$\frac{d\theta_1}{d\sigma} = \frac{1}{1 - \frac{\partial BR(\theta_1; \sigma)}{\partial \theta_1}} \times \frac{\partial BR(\theta_1; \sigma)}{\partial \sigma}. \tag{8}$$

In a stable equilibrium, this effect has same sign as

$$\frac{\theta_1 - s^\star(\theta_1, \sigma)}{\sigma} + \frac{\partial s^\star(\theta_1, \sigma)}{\partial \sigma}. \tag{9}$$

The first part of the Proposition uses the implicit function theorem to characterize the impact of noise on disclosure cutoffs. Equation (8) captures a multiplier effect which is driven by reverse unraveling. First, the direct effect of σ on disclosure incentives is that the marginal type θ_1 for whom staying quiet is a best response changes by $\Delta\theta_1^{(0)} = \partial BR/\partial \sigma$. Second, silence becomes worse news since $\Delta\theta_1^{(0)}$ fewer good types disclose, and the number of types for whom staying quiet is a best response changes further by $\Delta\theta_1^{(1)} = \Delta\theta_1^{(0)} \times \partial BR/\partial \theta_1$. Third, silence becomes yet worse news due to the second-round effect and the lack of disclosure by $\Delta\theta_1^{(1)}$ fewer good types, which yields a further change of $\Delta\theta_1^{(2)} = \Delta\theta_1^{(1)} \times \partial BR/\partial \theta_1$. This process continues, and the total effect is

$$\sum_{n=0}^{\infty} \Delta\theta_1^{(0)} = \sum_{n=0}^{\infty} \left(\frac{\partial BR(\theta_1; \sigma)}{\partial \theta_1} \right)^n \times \left(\frac{\partial BR(\theta_1; \sigma)}{\partial \sigma} \right).$$

In a stable equilibrium, $\partial BR/\partial \theta_1 < 1$, so that the sum converges to the expression in (8) and the effect of noise on disclosure has the sign of $\partial BR/\partial \sigma$, which in turn has the same sign as the expression in (9).

Equation (9) shows the sum of two effects. First, if the cutoff type θ_1 lies above the critical

signal $s^\star$, then the event of not obtaining the high action ($s < s^\star$) is a tail event from his perspective. Increased noise makes this tail event more likely, leading to increased incentives to disclose. Second, more noise may make the marginal Receiver more pessimistic, leading to an increase in the critical public signal $s^\star$, which would also increase incentives to disclose.

3.5 Crowding out and crowding in

It turns out that the sign of both effects in (9) is ambiguous in general, meaning that an increase in signal noise σ can either increase or decrease the equilibrium disclosure cutoff θ_1. Put differently, more precise signals can either crowd out or crowd in disclosure.

To see this point clearly, we specialize our focus to what we call the *normal model* in which we assume that θ and s are jointly normally distributed, that the Receiver's utility from taking the high action is simply $\gamma(\theta) = \theta$ and that disclosure costs $\delta(\theta)$ are proportional to the Sender's benefit from obtaining the high action $\beta(\theta)$.[27] We now characterize the response of the disclosure cutoff θ_1 to a change in the noise parameter σ.

Lemma 2. *Consider any stable interior equilibrium of the normal model. More precise public information crowds out private disclosures ($\frac{d\theta_1}{d\sigma} > 0$) if*

   1. $\mu < 0$; and

   2. $\frac{\delta(\theta)}{\beta(\theta)} \leq \frac{1}{2}$

Conversely, starting from any equilibrium θ_1 it is possible to increase μ and decrease $\frac{\delta(\theta)}{\beta(\theta)}$ sufficiently that (i) θ_1 remains an equilibrium and (ii) $\frac{d\theta_1}{d\sigma} < 0$. In other words, crowding in occurs.

Loosely speaking, crowding out (where more precise public information reduces equilibrium disclosure) is more likely to happen in bad times (where the common prior mean of θ, μ, is low) while crowding in is more likely to be a feature of healthy markets (high μ). In other words, crowding out prevails when the Receiver is sufficiently pessimistic *ex ante* and disclosure costs are not too large. To see why, note that the marginal Receiver's posterior expectations given the public signal are a weighted average of the prior μ and her signal s^*

$$E[\theta|s^*] = \alpha\mu + (1 - \alpha)s^*$$

where the weight $\alpha = \frac{\sigma^2}{\sigma_\theta^2 + \sigma^2}$ depends only on the signal-to-noise ratio. As public signals get more informative, the Receiver therefore increases the weight on the signal at the expense of her prior belief. But if the prior mean is low, then rebalancing the weights away from the prior increases

---

[27]In Appendix C, we flesh out the properties of equilibrium in this case fully.

expectations, and the marginal Receiver must become more willing to take the high action if the Sender stays quiet. Moreover, when $\delta(\theta) \leq \frac{1}{2}\beta(\theta)$, type θ_1 believes that the risk of drawing a signal below s^* is a tail event (otherwise, he would strictly prefer to disclose given that the cost is low). Thus, an increase in signal precision additionally makes type θ_1 more confident that signal s will exceed the hurdle rate s^*. As a result, the cost of staying quiet unambinguously reduces for the Sender. We return to this effect in our discussion of financial crises in Section 4.

Conversely, Lemma 2 also tells us that whenever μ is 'sufficiently high' relative to the relative cost of disclosure $\delta(\theta)/\beta(\theta)$, then more precise public signals ($d\sigma < 0$) *crowd in* disclosure by the Sender ($d\theta > 0$). Again, it is easiest to grasp the result by considering the marginal Receiver's beliefs when μ is high. When the prior expectation is sufficiently high, the marginal Reciever must be relatively pessimistic: that is, it must be that $s^*(\theta_1) < \mu$. Intuitively, since the marginal Receiver's posterior beliefs are anchored at $E[\theta|s^*, \theta \notin [c, \theta_1]]$, the critical signal s^* must change to offset the impact of a high prior mean on the Reciever's beliefs - otherwise, she would strictly prefer to take the high action. But when $s^*(\theta_1) < \mu$, increasing the precision of public information causes her to redistribute weight onto her private signal and become *more pessimistic* in the event that the Sender stays quiet. In other words, the Receiver's best response is to marginally increase the hurdle rate s^*. Anticipating this, the Sender becomes more reluctant to stay quiet, and as a result equilibrium disclosures actually increase with public signals.

Strictly, this heuristic argument does not take into account that in such situations, s^* is likely to fall below θ_1. In this case, more precise public information induces an offsetting benefit to staying quiet by reducing the probability of the tail event $s < s^*$. The Proof of Lemma 2 deals with this complication by simultaneously reducing the cost of disclosure to keep θ_1 unchanged. Using this thought experiment, we are able to establish that the former effect can dominate the latter, such that for sufficiently high μ and low δ, crowding in can occur.

4 Financial crises and stress tests

This Section applies our general framework to a model of financial crises based on Morris and Shin (2000). We employ a version of the general Sender-Receiver game in Section 2, where *banks* (Senders) decide whether to disclose evidence of their asset values (their type), and where *investors* (Receivers) decide whether or not to withdraw their investments from banks. As in Diamond and Dybvig (1983), illiquidity introduces a potential coordination problem among investors: Withdrawing investments is socially wasteful, but if others are withdrawing, then any individual investor's may prefer to also withdraw her investment.

By developing an explicit application, we can examine welfare and policy implications. The welfare implications of improving public access to information are particularly salient in financial

settings because (i) the data driven nature of financial markets makes information policy (such as the design of stress tests, or regulation of ratings agencies) feasible; (ii) inherent imperfections in financial contracting introduce a second-best rationale for meaningful information policy which does not simply mandate full transparency. For instance, it is well known that investors in banks face a coordination problem in their funding decisions, such that improved public information can actually increase the tendency for runs against solvent but illiquid banks. As discussed in the Introduction, we can also address a current policy question – whether stress test results should be made public – in a way that is robust to the Lucas critique.

Players, payoffs and technology. There is a continuum of banks indexed by $i \in [0,1]$. Each bank interacts with a continuum of its investors. Everybody is risk-neutral, and there are three dates $t \in \{0,1,2\}$. At date 0, each investor is endowed with one unit of cash. Investors lend their cash endowment to their bank. The bank invests this cash in a long-term project, which yield a stochastic gross return r^i at date 2. Projects are illiquid: If a proportion l^i of the long-term investment is withdrawn at time 1, then the return on the remaining projects is reduced to $r^i - 2cl^i$ (when they eventually mature at time 2). The parameter $c > 0$ captures the degree of illiquidity, and multiplying by 2 helps to save notation later on.

At date 1, each investor in bank i decides whether to *roll over* her investment or to *withdraw*.[28] Investors who withdraw are entitled to an immediate payment of one unit of cash. Investors who roll over are residual claimants on the bank's assets at time 2.[29] Thus, an investor who withdraws receives a certain payoff of 1, while an investor who rolls over receives a stochastic payoff $r^i - 2cl^i$.

We have followed the approach of Morris and Shin (2000), who make the analysis of investor's incentives particularly tractable by imposing two assumptions. First, bank's asset structure and its contract with investors is taken as given. Second, the liquidation technology is linear in the return r^i and the proportion of withdrawals l^i. Both assumptions can be relaxed using the techniques of Goldstein and Pauzner (2005).

Banks wish to maximize the aggregate utility of their investors. The coordination problem among investors introduces a conflict of interest, because individual investors will not necessarily choose actions that maximize aggregate utility. We allow for further conflicts of interest between investors and managers in Section 5.

---

[28]Relative to the Sender-Receiver model, each Sender (bank) faces a continuum of investors instead of one Receiver. This does not change the analysis, since the equilibrium of the roll-over subgame among investors can be interpreted as the action of one 'market receiver'.

[29]We make the latter assumption in order to follow convention (see, for example, Diamond and Dybvig (1983), Morris and Shin (2000) and Goldstein and Pauzner (2005)), but similar results would obtain if investors who roll over had a debt claim.

Returns and information. The return on assets at date 2 is $r^i = 1 + \theta^i + \eta^i$. The bank privately observes the first component θ^i, which is a continuous random variable with density $f(\theta^i)$ and support $\Theta = [0, \overline{\theta}] \subset \mathbb{R}$. θ^i is independently and identically distributed across banks.[30] We call θ^i the bank's *type*.

Investors do not observe θ^i directly, but rather observe a noisy public signal s^i. As in Section 3, the signal is $s^i = \theta^i + \sigma \varepsilon^i$, where the disturbance ε^i has density $h(\varepsilon^i)$, cumulative distribution $H(\varepsilon^i)$ and support $[\underline{\varepsilon}, \overline{\varepsilon}] \subset \mathbb{R}$, whose bounds may be infinite. $S(\theta) = [\theta + \sigma\underline{\varepsilon}, \theta + \sigma\overline{\varepsilon}] \equiv [\underline{s}(\theta), \overline{s}(\theta)]$ denotes the set of type θ's possible signals. We assume that high s is good news about θ in the sense of Milgrom's (1981) Monotone Likelihood Ratio Property.[31]

To model communication, we assume that the bank can send its investors a message $m^i \in \{\emptyset, \theta\}$. The message $m^i = \emptyset$ corresponds to *staying quiet* (or cheap talk) and $m^i = \theta^i$ corresponds to *disclosure* (or certification of θ^i). Disclosure is costly in that it reduces investors' utility by $\delta(\theta^i) < \theta^i$. As discussed above, this reduction captures the direct accounting cost of disclosure and the proprietary cost of revealing the bank's portfolio of investments to competitors. For simplicity, we assume that the cost of disclosure $\delta(\theta^i)$ does not depend on the number of investors withdrawing. Our results extend easily to cases where it does.

Neither the bank nor investors observe the second return component η^i. We introduce it to perturb common knowledge among bank i's investors, so that we can use a standard global games argument to find a unique equilibrium of the coordination game. We assume that $\eta^i \sim N(0, \sigma_\eta^2)$, and that each of bank i's investors $j \in [0,1]$ receives a noisy private signal $t_j^i = \eta^i + \zeta_j^i$, where $\zeta_j^i \sim N(0, \sigma_\zeta^2)$. The variables $(\zeta_j^i)_{j \in [0,1]}$ and η^i are all independent of each other and independent of θ^i.[32] In most of the following analysis, we focus on a limiting case where both η^i and t_j^i collapse to zero.

Stress test design. We consider the problem of a regulator who can create public information by publishing stress test results. By choosing the accuracy of those results, she can influence the accuracy of overall public information, which amounts to choosing the noise parameter from a set $\sigma \in [\underline{\sigma}, \overline{\sigma}]$. There is a direct resource cost of making stress tests precise, denoted $\chi(\overline{\sigma} - \sigma)$, where χ is a strictly convex increasing function.

---

[30]By assuming independence, we are eliminating the possibility of *contagion* across banks. In particular, disclosures about θ^i will contain no information about θ^j. Our goal in this paper is to characterize externalities across the banking system which arise even in the absence of contagion. We leave the effects of contagion and its interaction with the externalities we emphasize for future work.

[31]See Section 3 for a formal definition.

[32]The specification with two return components is adapted from Bouvard et al. (2015). It is useful because it separates the perturbation of η^i from the incentives to disclose θ^i.

Timing. The timing of the game at date 1 is as follows: First, the regulator chooses the stress test design σ. Second, each bank learns its type θ^i, and then sends its investors a message $m^i \in \{0, \emptyset\}$. Third, each investor observes her bank's message, the public signal s^i and the private signal t^i_j, and then chooses whether to withdraw or roll over. Figure 5 depicts the timeline.

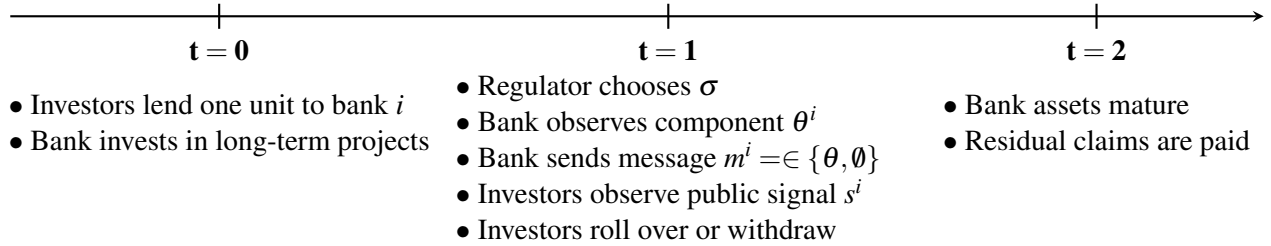| $t = 0$	$t = 1$	$t = 2$
• Investors lend one unit to bank i • Bank invests in long-term projects	• Regulator chooses σ • Bank observes component θ^i • Bank sends message $m^i = \in \{\theta, \emptyset\}$ • Investors observe public signal s^i • Investors roll over or withdraw	• Bank assets mature • Residual claims are paid

Figure 5: **Timeline of the financial crises model**

As we emphasized in Section 2, banks cannot produce new evidence in between the realization of public signals and investors' roll-over decisions. An obvious interpretation of this assumption is a time lag. Investors who hold short-term claims in banks can withdraw their investment very quickly – often overnight – when public information arrives. Moreover, the spontaneity of the market's response to public information is likely to be particularly acute in financial settings, where coordination problems among investors and a lack of common knowledge in how participants will respond to news events makes it extremely difficult to predict the timing of a bank run. It therefore seems reasonable that banks cannot produce verifiable reports about the quality of their assets at such short notice. Alternatively, one could think of situations in which the signal s^i is not observed directly by the bank itself. The analysis of such scenarios is also captured by our model.

Equilibrium definition. We consider symmetric *Perfect Bayesian Equilibria* (Fudenberg and Tirole, 1991). Perfect Bayesian Equilibrium requires (i) that investors chooses make optimal roll-over decisions, given their beliefs about the type θ^i after observing message m^i and signal s^i, (ii) that banks choose their messages optimally given their investors' strategy, and (iii) that investors' beliefs be consistent with Bayes' rule on the equilibrium path. Off the equilibrium path, we restrict beliefs only by requiring that after observing a signal s^i, investors attach zero probability to types with $s^i \notin S(\theta^i)$ who could not have sent this signal. The formal definition of equilibrium is as in Section 2.

From now on, we omit the i superscript on bank-specific variables and strategies whenever the context is clear.

4.1 The coordination game among investors: Liquidity versus solvency

At time 1, all investors in a given bank observe the same public signal s and message from the bank m. They commonly share the belief that the expected value of the first return component is $E_\mu[\theta|m,s]$. Investor $j \in [0,1]$ then uses her private signal t_j to update her beliefs about the second return component η, and makes her withdrawal decision based on that assessment.

Morris and Shin (2000) show that when investors' signals are sufficiently noisy relative to the prior distribution of η (that is, as long as the signal-to-noise ratio $\sigma_\eta^2/\sigma_\zeta^2$ is small), there is a unique equilibrium in which investor j withdraws if and only if $E[\eta|t_j] < \eta^\star$, where $\eta^\star$ is a critical value. The uniqueness is preserved in a limiting case approaching certainty, where $\sigma_\eta^2 \to 0$ and $\sigma_\zeta^2 \to 0$, as long as the signal-to-noise ratio is held constant.

In the unique limiting equilibrium, all investors in bank i withdraw if and only if their beliefs about the principal return component θ are sufficiently pessimistic. In particular, a bank run occurs if and only if

$$E_\mu[\theta|m,s] < c. \tag{10}$$

This condition is intuitive: Investors run whenever they have pessimistic beliefs about fundamentals, and they are more likely when bank assets are highly illiquid, i.e. when c is high. In the Online Appendix, we show how these conditions are derived.

Banks with $\theta < c$ are *illiquid* in the sense that they would face a run under full information. Banks $\theta \geq c$ are *liquid* and could avoid a run by disclosing their true type. The coordination problem among investors introduces a distinction between liquidity and solvency. Since we have assumed that $\theta \geq 0$, all banks are solvent in the sense that their assets have positive Net Present Value. We return to the impact of insolvent banks in Section 5.

4.2 Disclosures in equilibrium

In the limiting equilibrium investor behavior is binary: All investors run on their bank if (10) holds and all of them roll over otherwise. Thus we can use the tools we developed for the model with binary actions in Section 3 to characterize equilibrium. As before, we impose a regularity condition.

Assumption. *For all values of the public signal s_0, the function*

$$J(\theta) = H\left(\frac{s_0 - \theta}{\sigma}\right) - \frac{\delta(\theta)}{\theta}$$

crosses zero at most once on the range $\theta \in [\underline{\theta}, \bar{\theta}]$. If it crosses once, then it crosses from above.

As we discussed in more detail in Section 3, our regularity condition is satisfied under natural conditions, in particular if the cost $\delta(\theta)$ is proportional to the benefit of avoiding a run θ, or if pub-

lic information is precise enough. The proportional cost case is particularly relevant for financial intermediaries, because disclosing lending portfolios publicly leads to increased competition. The 'proprietary cost' of disclosure, due to the reduction in interest margins as a result of competition, is then likely to be proportional to total profits.[33]

The argument of Lemma 1 which generalizes the intuition presented in the Introduction applies here, and disclosure strategies in equilibrium are as depicted in Figure 6. In equilibrium, there is a cutoff θ_1 such that banks disclose if and only if $\theta \in [c, \theta_1]$. Banks worth less than c are illiquid in the sense that they would face a run under full disclosure. Thus they have a dominant strategy to remain quiet, and indeed free-ride on the reputation of better banks who stay quiet also. Banks worth more than θ_1 are liquid and could avoid a run by fully disclosing their value. However, they are confident that the public signal will convince investors of their quality, and thus disclosure is not worth the cost. Mediocre banks with $\theta \in [c, \theta_1]$ are also liquid, but they are anxious that their public signal will be insufficient to avoid a run. Thus, they prefer to pay the cost of disclosure.
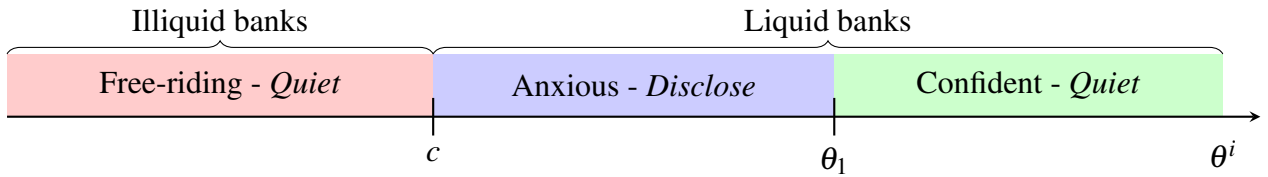
Figure 6: **Bank disclosure strategies in equilibrium**.

We find equilibria as before. Let $BR(\theta_1; \sigma)$ denote the highest quality bank who wishes to disclose, given that investors expect banks to play the cutoff strategy 'disclose if $\theta \in [c, \theta_1]$' and that the regulator has chosen the stress test design σ. In the case where the best response is an interior type, $BR(\theta_1; \sigma) \in (c, \overline{\theta})$, is can be neatly defined with $s^*(\theta_1; \sigma)$ as the solution to the system

$$\frac{\delta(\theta)}{\theta} = H\left(\frac{s^*(\theta_1; \sigma) - BR(\theta_1; \sigma)}{\sigma}\right) \tag{11}$$

$$E[\theta \mid s^*(\theta_1; \sigma), \theta \notin [c, \theta_1]] = c \tag{12}$$

To understand system (11)-(12), suppose that banks are playing an arbitrary disclosure strategy 'disclose if and only if $\theta \in [c, \theta_1]$' and consider the investor's best response to such a strategy. On observing no private disclosure by the bank and receiving a public signal s, we know that the investor withdraws if her posterior expectation of the bank's net return falls short of c and rolls over

---

[33]Moreover, even though in our particular model the benefit of avoiding a bank run is exactly the asset quality θ, this need not be the case. In particular, it is possible for this benefit to be decreasing in asset quality θ because high quality banks have less severe liquidity problems or easier access to liquidity assistance from central banks. Under those conditions, the regularity condition holds generally, even when disclosure costs are fixed and public information is noisy.

otherwise. Condition (12) therefore defines the lowest signal $s^*(\theta_1; \sigma)$ at which the investor will be willing to roll over her funds to the bank. Knowing this, a bank of type $\theta > c$ will choose to disclose if it believes the probability of investors receiving a signal $s < s^*(\theta_1; \sigma)$ is high enough to outweigh the cost of disclosure. Given the hurdle s^*, (12) therefore defines the highest type of bank $BR(\theta_1; \sigma)$ who is just willing to disclose when it expects other banks to disclose in $[c, \theta_1]$.

Rational expkectations in equilibrium yield the fixed point equation $BR(\theta_1; \sigma) = \theta_1$. In general, there can be multiple equilibria because disclosure decisions are strategic complements: More disclosure by high-quality types makes silence worse news, and increases incentives to disclose (see Figures 3 and 4 above for an illustration).

5 Stress test design

We now characterize welfare as a function of the stress test design σ, and emphasize the effect of changing σ on the equilibrium disclosure cutoff θ_1. In cases where multiple cutoffs constitute an equilibrium, this comparative statics exercise depends on which equilibrium is played. We focus on the local effects on one particular equilibrium cutoff θ_1, and assume that this equilibrium exists for all feasible policy choices $\sigma \in [\underline{\sigma}, \overline{\sigma}]$. This makes for a particularly simple exposition, and similar effects obtain when we allow equilibria to be selected according to the realization of a 'sunspot'.[34]

5.1 Aligned incentives and solvent banks

As a benchmark, we continue to assume that bank managers have the right incentives and that all banks are solvent with $\theta \geq 0$. For a bank with asset values θ, $\delta(\theta)$ captures the true social cost of making disclosures, and θ captures the true social benefit of avoiding a bank run. In this case, welfare in equilibrium is

$$W(\sigma, \theta_1) = \int_{\theta} (1 + \theta) dF(\theta) - \int_{\theta \in [c, \theta_1]} \delta(\theta) dF(\theta) - \int_{\theta \notin [c, \theta_1]} H\left(\frac{s^\star(\theta_1; \sigma) - \theta}{\sigma}\right) \theta dF(\theta) - \chi(\bar{\sigma} - \sigma),$$

Recall that $s^\star(\theta_1; \sigma)$ denotes the critical value of the public signal such that investors run on a non-disclosing bank if $s < s^\star(\theta_1; \sigma)$. The critical value depends on θ_1 and σ because banks' disclosure strategies and the amount of noise both influence investors' signal extraction problem on observing a bank who stays quiet.

The first term in the welfare function represents the expected Net Present Value of banks' assets in the first-best scenario, where nobody needs to make disclosures and nobody faces bank runs. The

---

[34]Moreover, for many reasonable parameter values and ranges of σ, equilibria are indeed unique: For instance, the analysis of (approximately) normal distributions in AppendixC demonstrates that equilibria are unique whenever the signal-to-noise ratio of public information exceeds $3/2$.

second term is the social cost of disclosures in equilibrium by banks in the interval $\theta \in [c, \theta_1]$. The third term is the expected cost of bank runs: A bank who does not disclose (since its assets' net return is 'mediocre', satisfying $\theta \notin [c, \theta_1]$) faces a bank run whenever it draws a public signal below the threshold $s^\star$. This occurs with probability $\Pr[s < s^\star | \theta] = H\left(\frac{s^\star - \theta}{\sigma}\right)$, and in that event it loses the continuation value θ.

A naive regulator takes the banks' policy θ_1 as given and maximizes welfare by considering the following effects: First, an increase in the noise σ directly affects the probability $\Pr[s < s^\star | \theta]$. Second, this increase also changes investors' signal extraction problem, and thus affects the critical signal $s^\star(\theta_1; \sigma)$. Third, increasing σ raises the resource cost $\chi(\bar{\sigma} - \sigma)$.

Our goal is to examine to what extent such a policy is vulnerable to the Lucas critique, i.e. in what manner the policy-maker would change her choice of σ if she realized the impact on banks' disclosure policies. A useful approach is to posit that the naive regulator has guessed θ_1 correctly, and that she is at an interior optimal policy $\sigma^\star$.

Proposition 8. *At the naive regulator's optimal choice $\sigma^\star$, the marginal effect of further increasing σ on welfare is*

$$\frac{dW(\sigma^\star, \theta_1)}{d\sigma} = \left\{\frac{d\theta_1}{d\sigma}\right\} \times \left\{\frac{\partial s^\star(\theta_1, \sigma)}{\partial \theta_1}\right\} \times \left\{-\frac{c}{\sigma} \int\limits_{\theta \notin [c, \theta_1]} h\left(\frac{s^\star(\theta_1, \sigma) - \theta}{\sigma}\right) dF(\theta)\right\} \qquad (13)$$

and this effect has the opposite sign to $d\theta_1/d\sigma$.

In this Proposition we decompose the effect of increasing the noise parameter σ beyond the naive optimum, and use this decomposition to sign the overall effect on welfare. At her optimal choice $\sigma^\star$, the naive regulator exactly trades off the three effects we outlined above. The remaining effects on welfare, which are not internalized by the naive regulator, are caused by changes in the disclosure strategy θ_1. On one hand, some banks with asset values near the cutoff θ_1 either disclose more (if $\partial \theta_1 / \partial \sigma > 0$) or stop disclosing (if $\partial \theta_1 / \partial \sigma < 0$). However, this direct change such a bank's utility only has a second-order effect on welfare. To see this, note that banks near the cutoff are almost indifferent between disclosing and not disclosing. Because incentives are aligned and their cost-benefit trade-off reflects social welfare, society is also almost indifferent. On the other hand, the change in θ_1 has further indirect effects on the signal extraction problem of investors.

We decompose the latter into three terms in (13), which illustrates that the indirect effect has first-order importance for welfare . First, the mass of high-quality banks that disclose changes by $\partial \theta_1 / \partial \sigma$. Second, the critical signal required to avoid a bank run changes by $\partial s^\star / \partial \theta_1$ in response. This partial derivative is always positive: When more high quality banks disclose ($\uparrow \theta_1$), then silence becomes worse news, and a higher signal $s^\star$ is required to avoid a run on a quiet bank.

Finally, the change in the critical signal has an impact on welfare, because it changes the probability with which banks who stay quiet in equilibrium experience a run. In Equation (13), we characterize this final effect as $(-c)$ times a scaling factor, which is always negative. Intuitively, the effect of a small increase in $s^\star$ on welfare is that the marginal bank, who receives exactly the critical signal $s^\star$, is no longer able to avoid a run. But on the margin (for public signal realizations close to s^*) this bank is worth exactly c in expectation – this follows from the definition of the critical signal. Since $c > 0$, the fact that the marginal bank loses funding implies a welfare loss, and indeed this loss in welfare is scaled by the *ex ante* probability that investors observe a signal close to s^*.

The second part of Proposition 8 uses this decomposition to sign the overall welfare effect of increasing σ. If noisier information leads to more disclosure, then $\partial\theta_1/\partial\sigma > 0$ and the effect on welfare is negative. Put differently, if more precise information crowds out disclosure, then $\partial\theta_1/\partial\sigma > 0$ and it pays to make stress tests *more precise* once we account for the endogenous response of disclosure strategies.

At the margin, crowding out affects liquid banks around θ_1 who are indifferent between disclosing and staying quiet. It turns out that, from a welfare perspective, this is a good thing: By staying quiet, liquid banks strengthen the insurance provided to illiquid banks who also stay quiet, because silence has become better news. Thus, welfare increases unambiguously.

5.2 *Managerial incentive problems*

We now relax the assumption that bank managers have the right incentives, but maintain the assumption that there are no insolvent banks. In particular, managers have a contract with investors which implies that the benefit of avoiding a run is $B(\theta)$ and the cost of disclosure is $D(\theta)$, while the true social costs and benefits are $\delta(\theta)$ and θ respectively, as before.

As before, we consider the effect of increasing σ beyond the naive regulator's optimal choice. The effects we described in Proposition 8 are unchanged: Crowding out disclosures by liquid banks is still welfare-improving because it enhances the insurance provided to illiquid ones. Moreover, there is an additional effect which depends on managerial incentives.

Proposition 9. *At the naive regulator's optimal choice $\sigma^\star$, the marginal effect of further increasing σ on welfare is the sum of the expression in Equation ((13)) and a term which has the same sign as*

$$\frac{\partial\theta_1}{\partial\sigma} \times \left[\frac{D(\theta_1)}{B(\theta_1)} - \frac{\delta(\theta_1)}{\theta_1} \right] \tag{14}$$

If increasing σ increases disclosures, then $\partial\theta/\partial\sigma > 0$. Proposition 9 shows that in this case, the additional welfare effect has the sign of the difference between the private relative cost of disclosure and the social relative cost. Intuitively, when mangers underestimate the social cost

of disclosure, then they privately decide to disclose too little at the margin, and any policy that increases disclosures in equilibrium further improves welfare.

This specification can capture a variety of situations. First, managers may not internalize the entire benefit of avoiding a run when they have limited liability, so that they would overstate the relative cost of disclosure and disclose too little. In a financial crisis, where more precise public information tends to crowd out disclosure, this means that optimal stress tests ought to be made less precise in order to encourage more disclosure. Second, managers may *overstate* the benefit of avoiding a run if they wish to preserve their reputation or to take advantage of long-term compensation arrangements. Finally, managers may overstate the cost of disclosure if this is mainly the proprietary cost of revealing sensitive information to competitors, since the profits lost from increased competition constitute only a welfare-neutral transfer from a social perspective. In this case, stress tests ought to be more precise in order to reduce disclosures which are made purely to ensure the survival of managers or preserve rents.

5.3   Insolvent banks and resolution policy

In this final Subsection, we allow the bank's Net Present Value θ to be drawn from an interval $[\underline{\theta}, \overline{\theta}] \subset \mathbb{R}$, where $\underline{\theta} < 0$. There are now insolvent banks with $\theta < 0$ for whom the welfare-maximizing policy is to liquidate all assets at date 1. If the incentives of managers and investors are aligned, then managers who find out that their bank is insolvent will voluntarily liquidate assets. Assuming that this liquidation is observed by everybody, welfare is the same as in Subsection 5.1, since insolvent banks effectively leave the market.

We obtain more interesting results by introducing insolvent banks in the model of managerial incentive problems from Subsection 5.2. In particular, managers have incentives which imply that the benefit of avoiding a run to a manager is $B(\theta)$ and the cost of disclosure is $D(\theta)$. We assume that $B(\theta) > D(\theta) \geq 0$ for all θ, so that even managers of insolvent banks prefer to avoid a run.

Equilibrium disclosure strategies are as before: Insolvent banks join the pool of illiquid banks who stay quiet, and free-ride on the reputation of liquid banks. Among liquid banks, the best ones are confident and stay quiet, while mediocre ones with $\theta \in [c, \theta_1]$ are anxious and disclose.

Perhaps surprisingly, the basic welfare analysis is also unchanged. Crowding out disclosure has a positive effect on welfare, as demonstrated in Proposition 8, since it strengthens the insurance provided by liquid banks who stay quiet to illiquid banks. This remains true despite the fact that liquid banks now also insure their insolvent peers. To see why that is the case, recall that the insurance effect works through the impact of disclosure strategies on the critical public signal $s^\star$ below which investors run on their bank. In particular, less disclosure by liquid banks decreases the critical signal, which insures 'marginal banks' who receive signals close to $s^\star$ against a run. However, the critical signal is defined such that investors who observe $s^\star$ consider the bank to be

worth exactly c. Thus, 'marginal banks' are worth approximately $c > 0$ from an *ex ante* perspective. Insuring them always yields an *average* welfare improvement which is proportional to c, even though the increase in insurance also benefits insolvent banks in some states of the world.

Although the cost-benefit trade-off regarding the precision of stress tests is not affected by the presence of insolvent banks, there is value in introducing any policy which serves to remove insolvent banks from the market. For example, one could allow regulators to scrutinize banks' assets at date1 and force banks with $\theta < 0$ into resolution. Such policies are implemented in practice, for example, through the FDIC's 'prompt corrective action' approach, which involves the immediate resolution of banks whose capitalization falls short of certain predetermined standards.[35]

In our model, resolution policy can further improve welfare for three reasons: First, closing down an insolvent bank which would only stay open to serve managers directly improves the welfare of that bank's investors. Second, when banks with $\theta < 0$ are removed from the model, it is easy to show that the critical signal $s^\star$ unambiguously decreases. Investors are reassured by the removal of the worst banks, and this increases welfare by strengthening the insurance provided to illiquid but solvent banks. Finally, since the 'quality' of the quiet signal exogenously improves, more 'good' banks are marginally incentivized to stay quiet. If managers' incentives are correctly aligned, this further improves welfare (as discussed above). Of course, we make the caveat that this third effect can be socially costly when managerial incentive problems are present.

6 Conclusion

We have studied the interaction between private disclosures and public information. In a general Sender-Receiver model, we have shown that it is common for precise public information to crowd out incentives to disclose, and that this effect is amplified by 'reverse unraveling'. Moreover, more public information can lead to less information overall in equilibrium, in the strong sense of Blackwell's criterion.

These findings have repercussions for policy and for empirical research. On the policy side, we established that the design of stress tests in financial crises should consider additional, and amplified, trade-offs when it takes into account the response of banks' private disclosures. In particular, stress tests should be more precise so as to crowd out disclosures in financial crises, although this recommendation is mitigated when bank managers have skewed incentives which lead them to understate the benefit of avoiding bank runs, or overstate the social cost of making disclosures. On the

---

[35] Another approach to resolution policy, which would have similar effects, is the recent proposal to 'bail-in' the debt of banks whose solvency is in danger, which involves writing down banks' debt obligations to improve solvency. Resolution policy can introduce further incentive problems for the regulator, who may shy away from aggressive resolution policies in order to avoid revealing bad news about aggregate states to investors. Walther and White (2015) analyze this forbearance problem and show that it can be resolved using commitment devices such as contingent capital instruments or CoCos.

empirical side, our findings help to rationalize recent findings on disclosure policies in consumer markets which are non-monotonic in underlying product quality. Moreover, our model yields novel testable predictions. In future work, we hope to use these predictions to empirically explore the determinants of opacity in financial crises, and the determinants of corporate disclosure policies.

References

Acharya, Viral V., Peter DeMarzo, and Ilan Kremer, "Endogenous Information Flows and the Clustering of Announcements," *American Economic Review*, 2011, *101* (7), 2955–2979.

Akerlof, George E., "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, 1970, *84* (3), 488–500.

Amador, Manuel and Pierre-Olivier Weill, "Learning from Prices: Public Communication and Welfare," *Journal of Political Economy*, 2010, *118* (5), 866–907.

Angeletos, George-Marios and Alessandro Pavan, "Efficient Use of Information and Social Value of Information," *Econometrica*, 2007, *75* (4), 1103–1142.

Athey, Susan, "Monotone Comparative Statics under Uncertainty," *Quarterly Journal of Economics*, 2002, *117* (1), 187–223.

Bederson, Benjamin B., Ginger Zhe Jin, Phillip Leslie, Alexander J. Quinn, and Ben Zou, "Incomplete Disclosure: Evidence of Signaling and Countersignaling," 2015. Mimeo.

Berger, Philip G. and Rebecca Hann, "The Impact of SFAS No. 131 on Information and Monitoring," *Journal of Accounting Research*, 2003, *41* (2), 263–223.

Blackwell, David, "Equivalent Comparison of Experiments," *Annals of Mathematical Statistics*, 1953, *24* (2), 265–272.

Boissay, Frédéric, Fabrice Collard, and Frank Smets, "Booms and Banking Crises," *Journal of Political Economy (forthcoming)*, 2015.

Bouvard, Matthieu, Pierre Chaigneau, and Adolfo de Motta, "Transparency in the Financial System: Rollover Risk and Crises," *Journal of Finance*, 2015, *70* (4), 1805–1837.

Dang, Tri Vi, Gary Gorton, Bengt Holmstrom, and Guillermo Ordonez, "Banks as Secret Keepers," 2015. Mimeo.

Diamond, Douglas W and Philip H Dybvig, "Bank Runs, Deposit Insurance, and Liquidity," *Journal of Political Economy*, 1983, *91* (3), 401–19.

Dranove, David and Ginger Zhe Jin, "Quality Disclosure and Certification: Theory and Practice," *Journal of Economic Literature*, 2010, *48* (4), 935–963.

Dye, Ronald A., "Disclosure of Nonproprietary Information," *Journal of Accounting Research*, 1985, *23* (1), 123–145.

Edelman, Benjamin, "Adverse Selection in Online Trust Certifications and Search Results," *Electronic Commerce Research and Applications*, 2011, *10* (1), 17–25.

Ely, Jeffrey C., "Beeps," 2015. Mimeo.

Feltovich, Nick, Richmond Harbaugh, and Ted To, "Too Cool for School? Signalling and Countersignalling," *RAND Journal of Economics*, 2002, *33* (4), 630–649.

Fudenberg, Drew and Jean Tirole, "Perfect Bayesian Equilibrium and Sequential Equilibrium," *Journal of Economic Theory*, 1991, *53* (2), 236–250.

Goldstein, Itay and Ady Pauzner, "Demand–Deposit Contracts and the Probability of Bank Runs," *Journal of Finance*, 2005, *60* (3), 1293–1327.

_ **and Yaron Leitner**, "Stress Tests and Information Disclosure," 2015. Mimeo.

Gorton, Gary, "The Panic of 2007," NBER Working Paper 14358, National Bureau of Economic Research 2008.

Grossman, Sanford J, "The Informational Role of Warranties and Private Disclosure About Product Quality," *Journal of Law and Economics*, 1981, *24* (3), 461–483.

_ **and Oliver D Hart**, "Disclosure Laws and Takeover Bids," *Journal of Finance*, 1980, *35* (2), 323–334.

Hagenbach, Jeanne, Frederic Koessler, and Eduardo Perez-Richet, "Certifiable Pre-Play Communication: Full Disclosure," *Econometrica*, 2014, *83* (3), 1093–1131.

Harris, Mary S., "The Association between Competition and Managers' Business Segment Reporting Decisions," *Journal of Accounting Research*, 1998, *36* (1), 111–128.

Hayes, Rachel M. and Russell Lundholm, "Segment Reporting to the Capital Market in the Presence of a Competitor," *Journal of Accounting Research*, 1996, *34* (2), 261–279.

Heider, Florian, Marie Hoerova, and Cornelia Holthausen, "Liquidity Hoarding and Interbank Market Rates: The Role of Counterparty Risk," *Journal of Financial Economics (forthcoming)*, 2015.

Jordan, John S., Joe Peek, and Eric S. Rosengreen, "The Impact of Greater Bank Disclosure Amidst a Banking Crisis," Working Paper, Federal Reserve Bank of Boston 1999.

Jovanovic, Boyan, "Truthful Disclosure of Information," *The Bell Journal of Economics*, 1982, *13* (1), 36–44.

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

Kohlhas, Alexandre N., "An Informational Theory of Stabilization Policy," 2015. Mimeo.

Kreps, David M. and Robert Wilson, "Sequential Equilibria," *Econometrica*, 1982, *50* (4), 863–894.

Leuz, Christian and Peter Wysocki, "Economic Consequences of Financial Reporting and Disclosure Regulation: A Review and Suggestions for Future Research," 2008. Mimeo.

Lewis, Gregory, "Asymmetric Information, Adverse Selection and Online Disclosure: The Case of eBay Motors," *American Economic Review*, 2011, *101* (4), 1535–1546.

Luca, Michael and Jonathan Smith, "Strategic Disclosure: The Case of Business School Rankings," *Journal of Economic Behavior and Organization*, 2015, *112*, 17–25.

Lucas, Robert E., "Econometric Policy Evaluation: A Critique," in Karl Brunner and Alan Meltzer, eds., *Carnegie-Rochester Conference Series on Public Policy*, Vol. 1, Elsevier, 1978, pp. 19–46.

Mankiw, Gregory, "The Allocation of Credit and Financial Collapse," *Quarterly Journal of Economics*, 1986, *101* (3), 455–470.

Milgrom, Paul, "What the Seller Won't Tell You: Persuasion and Disclosure in Markets," *Journal of Economic Perspectives*, 2008, *22* (2), 115–132.

Milgrom, Paul R, "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 1981, *12* (2), 380–391.

Mishkin, Frederic S., "Asymmetric Information and Financial Crises: A historical perspective," NBER Working Paper 3400, National Bureau of Economic Research 1990.

Morris, Stephen and Hyun Song Shin, "Rethinking Multiple Equilibria in Macroeconomic Modeling," *NBER Macroeconomics Annual*, 2000, *15*, 139–182.

_ **and** _ , "Social Value of Public Information," *American Economic Review*, 2002, *92* (5), 1521–1534.

_ **and** _ , "Central Bank Transparency and the Signal Value of Prices," *Brookings Papers on Economic Activity*, 2005, *2*, 1–66.

Myerson, Roger B. and Philip J. Reny, "Sequential Equilibria of Multi-Stage Games with Infinite Sets of Types and Actions," 2015. Mimeo, University of Chicago.

Okuno-Fujiwara, Masahiro, Andrew Postlewaite, and Kotaro Suzumura, "Strategic Information Revelation," *Review of Economic Studies*, 1990, *57* (1), 25–47.

Perez-Richet, Eduardo, "Interim Bayesian Persuasion: First steps," *American Economic Review*, 2014, *104* (5), 469–74.

Schelling, Thomas, *Micromotives and Macrobehavior*, WW Norton & Company, 1978.

Seidmann, Daniel J and Eyal Winter, "Strategic Information Transmission With Verifiable Messages," *Econometrica*, 1997, *65* (1), 163–169.

Shin, Hyun Song, "News Management and the Value of Firms," *RAND Journal of Economics*, 1994, *25* (1), 58–71.

_ , "Disclosures and Asset Returns," *Econometrica*, 2003, *71* (1), 105–133.

Sowerbutts, Rhiannon and Peter Zimmerman, "Banks' Disclosure and Financial Stability," Quarterly Bulletin 2013 Q4, Bank of England 2013.

Stein, Jeremy C., "Cheap Talk and the Fed: A Theory of Imprecise Policy Announcements," *American Economic Review*, 1989, *79* (1), 32–42.

Verrecchia, Robert E., "Discretionary Disclosure," *Journal of Accounting and Economics*, 1983, *5*, 179–194.

Vives, Xavier, "Learning From Others: A Welfare Analysis," *Games and Economic Behavior*, 1997, *20* (2), 177–200.

Wagenhofer, Alfred, "Voluntary Disclosure With a Strategic Opponent," *Journal of Accounting and Economics*, 1990, *12* (4), 341–363.

Walther, Ansgar and Lucy White, "Rules Versus Discretion in Bank Resolution," 2015. Mimeo.

A   Proofs

Throughout this Appendix, we refer to the Sender as S and to the Receiver as R. With a slight notational abuse, we refer to $\sigma_S^*(m, \theta)$ as the probability that type θ send message m under σ_S.

Proof of Proposition 1

Proposition. *Let $\theta_P = \max\{\theta : P(\theta) \geq \delta(\theta)\}$ be the highest type for which the maximal punishment exceeds the cost of disclosure. If $\theta_P < \overline{\theta}$, then types $\theta > \theta_P$ stay quiet with probability one in any equilibrium.*

Proof. We first establish that in any equilibrium, the Sender's optimal disclosure strategy σ_S^* satisfies $\sigma_S^*(\overline{\theta}) \in \Delta(M^c)$. We then use an iterative argument to show that all types $\theta > \theta_P$ also prefer to play some $m \in M^c$.

Consider type $\overline{\theta}$'s expected payoff from playing some verifiable message $m \in M(\overline{\theta}) \setminus M^c$ and let $a_{\sigma_S^*}^*(m, s)$ denote R's equilibrium action on observing message m and signal s (where S uses strategy σ_S^*). For any such message, his expected payoff $\sum_{s \in S} g(s \mid \theta) . v\left(a_{\sigma_S^*}^*(m, s), \theta\right)$ is bounded above by

$$\sum_{s \in S} g(s \mid \overline{\theta}) . v\left(a_{\sigma_S^*}^*(m, s), \overline{\theta}\right) - \delta(\overline{\theta}) \leq v\left(a^*(\overline{\theta}), \overline{\theta}\right) - \delta(\overline{\theta}) \tag{15}$$

since log-monotonicity of R's payoff and $\overline{\theta} \geq \theta$, $\forall \theta \in \Theta$, jointly imply that for any message m and signal s we have[36] $a_{\sigma_S^*}^*(m, s) \leq a^*(\overline{\theta})$ and furthermore since v is strictly increasing in a. In particular, for any message $m \in M(\overline{\theta}) \setminus \{\cup_{\theta \neq \theta'} M(\theta)\}$ the upper bound (15) is tight since such messages verifiably identify the Sender's type as $\overline{\theta}$.

Now, consider type $\overline{\theta}$'s expected payoff from sending a message $m' \in M^c$. This is bounded below by

$$\sum_{s \in S} g(s \mid \overline{\theta}) . v\left(a_{\sigma_S^*}^*(m', s), \overline{\theta}\right) \geq \sum_{\theta' \in \Theta} \Pr\left(s \in \underline{S}(\theta') \mid \overline{\theta}\right) . v\left(a^*(\theta'), \overline{\theta}\right) \tag{16}$$

where again, log-monotonicity of R's payoff and $\theta \geq \theta'$ when $s \in \underline{S}(\theta)$ imply that $a_{\sigma_S^*}^*(m', s) \geq a^*(\theta')$. Since v is increasing, (16) then follows immediately. Further, this lower bound is tight for a strategy profile in which only type $\underline{\theta}$ plays messages in M^c. It can be supported by the most pessimal beliefs that $\theta = \theta'$ for any signal $s \in \underline{S}(\theta')$.[37]

---

[36] See Athey (2002).

[37] Recall that we do not allow beliefs following signal s that are inconsistent with types for whom such a signal can be observed.

Thus, comparing (15)-(16) type $\overline{\theta}$ will prefer to play some $m \in M^c$ *regardless* of any other type's disclosure strategy if

$$\sum_{\theta' \in \Theta} \Pr\left(s \in \underline{S}\left(\theta'\right) \mid \overline{\theta}\right) . v\left(a^*\left(\theta'\right), \overline{\theta}\right) > v\left(a^*\left(\overline{\theta}\right), \overline{\theta}\right) - \delta\left(\overline{\theta}\right)$$

or, on rearrangement if

$$\delta\left(\overline{\theta}\right) > \sum_{\theta' \in \Theta} \Pr\left(s \in \underline{S}\left(\theta'\right) \mid \overline{\theta}\right) \left[v\left(a^*\left(\overline{\theta}\right), \overline{\theta}\right) - v\left(a^*\left(\theta'\right), \overline{\theta}\right)\right] = P\left(\overline{\theta}\right)$$

which is exactly what we needed to show.

Now, suppose that the highest q types who satisfy $\theta > \theta_P$ all play quiet in any equilibrium. Consider the $q+1$-th highest type θ^{q+1}. Since all types in $\left\{\theta^q, \theta^{q-1}, \ldots, \theta^1 = \overline{\theta}\right\}$ play messages in M^c by assumption, type θ^{q+1} is the highest type who could send any message $m \in M\left(\theta^{q+1}\right)$. Then, we can apply an identical argument to that following (15) to show that type θ^{q+1}'s expected payoff from disclosing is bounded above by

$$\sum_{s \in S} g\left(s \mid \theta^{q+1}\right) . v\left(a^*_{\sigma^*_S}\left(m, s\right), \theta^{q+1}\right) - \delta\left(\theta^{q+1}\right) \leq v\left(a^*\left(\theta^{q+1}\right), \theta^{q+1}\right) - \delta\left(\theta^{q+1}\right) \quad (17)$$

Moreover, an almost identical argument establishes that the lower bound on type θ^{q+1}'s payoff from playing a message $m \in M^c$ is

$$\sum_{s \in S} g\left(s \mid \theta^{q+1}\right) . v\left(a^*_{\sigma^*_S}\left(m', s\right), \theta^{q+1}\right) \geq \sum_{\theta' \leq \theta^{q+1}} \Pr\left(s \in \underline{S}\left(\theta'\right) \mid \theta^{q+1}\right) . v\left(a^*\left(\theta'\right), \theta^{q+1}\right) \quad (18)$$

Comparing (17) and (18), a sufficient condition for type θ^{q+1} to strictly prefer playing $m \in M^c$ over any $m \in M\left(\theta^{q+1}\right) \setminus M^c$ *regardless* of the disclosure strategy of any type $\theta < \theta^{q+1}$ is, after rearrangement

$$\delta\left(\theta^{q+1}\right) > \sum_{\theta' \leq \theta^{q+1}} \Pr\left(s \in \underline{S}\left(\theta'\right) \mid \theta^{q+1}\right) . \left[v\left(a^*\left(\theta^{q+1}\right), \theta^{q+1}\right) - v\left(a^*\left(\theta'\right), \theta^{q+1}\right)\right] = P\left(\theta^{q+1}\right)$$

which completes the proof. $\qquad\square$

Proof of Proposition 2

Proposition. *Suppose that $\theta_P < \overline{\theta}$ and that types above θ_P are minimally persuasive. If the state space is sufficiently dense, then there is a non-empty set of types $\theta \leq \theta_P$ who stay quiet with probability one in any equilibrium.*

Proof. We show that, if the density of the state space ζ, as defined in Equation (3), is sufficiently large, type θ_P must only play messages $m \in M^c$ in any equilibrium if a subset of types above θ_P is minimally persuasive. That is, any equilibrium disclosure strategy σ_S^* must satisfy $\sigma_S^*(\theta, \theta) = 1$ if $\theta > \theta_P$. Since type θ_P by definition does not have a dominant strategy to play $m = \emptyset$, this establishes non-trivial conditions under which reverse unraveling takes place.

Let #X denote the cardinality of an arbitrary set X. Recalling that S is a finite set, we establish the following Lemma:

Lemma 3. *Suppose $\zeta > 1/\varepsilon$ for some ε. Then, for two consecutive types $\theta^{t'}, \theta^{t'-1} \in \Theta$ where $\theta^{t'} = \theta_P$ and $\theta^{t'} < \theta^{t'-1}$, the Sender's equilibrium marginal payoff, $\Delta U(\theta)$, from playing $m \in M^c$ over some $m' \in M(\theta) \setminus M^c$ satisfies*

$$\Delta U\left(\theta^{t'-1}\right) \leq \Delta U\left(\theta^{t'}\right) + \varepsilon\left(1 + \#S.\overline{B}\right) \tag{19}$$

Additionally, (19) also holds for any strategy profile σ_S'' in which $\sigma_S''(\theta, \theta) = 1$ if $\theta > \theta_P$ and $a_{\sigma_S'}(m, s)$ is a best response to σ_S''.

Proof. By definition, we have

$$\Delta U\left(\theta^{t'-1}\right) = \sum_{s \in S} g\left(s \mid \theta^{t'-1}\right) \cdot \left[v\left(a_{\sigma_S^*}(m, s), \theta^{t'-1}\right) - v\left(a_{\sigma_S'}(m', s), \theta^{t'-1}\right)\right] \tag{20}$$
$$- \delta\left(\theta^{t'-1}\right)$$

for any $m' \in M\left(\theta^{t'-1}\right) \setminus M^c$. By ε-denseness, we know that for any $s \in S$, $g\left(s \mid \theta^{t'-1}\right) \leq g\left(s \mid \theta^{t'}\right) + \varepsilon$ and $v\left(a, \theta^{t'-1}\right) \leq v\left(a, \theta^{t'}\right) + \varepsilon$ for any action $a \in A$. Substituting these conditions into (20) yields

$$\Delta U\left(\theta^{t'-1}\right) \leq \sum_{s \in S} g\left(s \mid \theta^{t'}\right) \cdot \left[v\left(a(m, s), \theta^{t'}\right) - v\left(a\left(m'\left(\theta^{t'-1}\right), s\right), \theta^{t'}\right)\right]$$
$$- \delta\left(\theta^{t'}\right) + \varepsilon + \varepsilon(\#S).\overline{B} \tag{21}$$

Now, in any equilibrium it must be the case that $a_{\sigma_S^*}(m', s) \leq a^*\left(\theta^{t'}\right)$ for any $m' \in M\left(\theta^{t'-1}\right) \setminus M^c$. By contrast, type $\theta^{t'-1}$ can always secure $a^*\left(\theta^{t'-1}\right) \geq a^*\left(\theta^{t'}\right)$ by playing $m = \theta^{t'-1}$. Substituting

$$a^*\left(\theta^{t'-1}\right) \geq a_{\sigma_S^*}\left(m'\left(\theta^{t'}\right), s\right)$$

into (21) establishes that

$$\Delta U \left(\theta^{t'-1} \right) \leq \Delta U \left(\theta^{t'} \right) + \varepsilon \left(1 + \#S.\overline{B} \right)$$

The final claim of the proof can be established by similar logic to the above, noting that $a_{\sigma_S''} (m',s) \leq a^* \left(\theta^{t'} \right)$ must also hold for type $\theta^{t'} = \theta_P$ for any strategy in which types $\theta > \theta_P$ play the full disclosure message $\sigma_S'' (\theta, \theta) = 1$ if $\theta > \theta_P$. $\qquad \square$

We now show that *(i)* type θ_P's maximal net payoff to playing any cheap talk message is between 0 and $-\varepsilon \left(1 + \#S.\overline{B} \right)$ for the 'worst case' strategy profile $\tilde{\sigma}_S (\theta)$ in which $a = a^* (\theta')$ for $s \in \underline{S}(\theta')$; *(ii)* if types θ_P are minimally persuasive, then there is a cheap talk message which gets θ_P better than worst signals following some observations of s; *(iii)* there is an ε small enough for which the marginal payoff to playing this such message is positive when types $\theta \geq \theta_P$ stay quiet.

Find t' such that $\theta^{t'} = \theta_P$. First, note that by definition of type θ_P, we have

$$P(\theta_P) \geq \delta(\theta_P)$$

or on rearrangement

$$\sum \Pr \left(s \in \underline{S}(\theta') \right) . \left[v \left(a^* (\theta'), \theta_P \right) - v(a^* (\theta_P), \theta_P) \right] - \delta(\theta_P) \qquad (22)$$

Expression (22) is exactly the net payoff to type θ_P from playing $m = \theta_P$ under any 'worst case' strategy profile, $\tilde{\sigma}_S (\theta)$, $\tilde{a}(m,s)$, where $\tilde{\sigma}(m,\theta) > 0$ iff $m \in M^c$ if $\theta = \underline{\theta}$, $\tilde{\sigma}(\theta,\theta) = 1$ if $\theta > \underline{\theta}$ and $a(m,s) = a^* (\theta')$ for any $m \neq \{\theta\}$ and $s \in \underline{S}(\theta')$. Such a profile is feasible, and is not a dominated strategy for the Receiver (it is dominant after any s to play $a \geq a^* (\theta')$). [38] Thus, we have

$$\min_{\sigma_S(\theta),a(m,s)} \Delta U \left(\theta_P \mid \sigma_S (\theta), a(m,s) \right) \leq 0 \qquad (23)$$

among all disclosure profiles $\sigma_S \in \times_{\theta \in \Theta} \Delta (M(\theta))$ and undominated action profiles $a(m,s) \geq a^* (\theta')$ for $s \in \underline{S}(\theta')$, where the notation $\Delta U (\theta_P | \sigma_S (\theta), a(m,s))$ makes its dependence on other's play $m(\theta)$, $a(m,s)$ explicit for the sake of clarity.

However, we also know by definition of θ_P that for type $\theta^{t'-1}$

$$\Delta U \left(\theta^{t'-1} \mid \tilde{\sigma}_S (\theta), \tilde{a}(m,s) \right) > 0 \qquad (24)$$

But since $\tilde{\sigma}_S (\theta) = \theta$ for all $\theta > \underline{\theta}$ (and therefore for all $\theta > \theta_P$), we can apply Lemma 3 to Equation

---

[38] Off the equilibrium path, $a = a^* (\theta)$ can be sustained by the most pessimal belief consistent with s.

to yield

$$\Delta U\left(\theta_P \mid \tilde{\sigma}_S\left(\theta\right), \tilde{a}\left(m,s\right)\right) > -\varepsilon\left(1 + \#S.\overline{B}\right) \tag{25}$$

Having established that type θ_P has only a 'small' preference for disclosure in the worst case scenario, we now argue that he will always prefer to play some cheap talk message should types $\theta > \theta_P$ all play cheap talk. We begin by showing that, if types are minimally persuasive, then the 'worst case' marginal payoff from cheap talk must strictly improve for type θ_P after deletion of dominated strategies:

Lemma 4. *Suppose that types $\theta > \theta_P$ are minimally persuasive. For any message profile $\sigma'_S(\theta) \in \times_{\theta \in \Theta} \Delta\left(M\left(\theta\right)\right)$ satisfying $\sigma'_S\left(m, \theta\right) > 0$ iff $m \in M^c$ for all $\theta > \theta_P$, there must exist a signal s' and a cheap talk message $m^c \in M^c$ such that R's best response on observing m^c must satisfy*

$$a_{\sigma'_S}\left(m^c, s'\right) > a^*\left(\theta'\right)$$

where θ' is the lowest type who can send message s'.

Proof. Let

$$a_s^{**} := \arg\max_{a \in A} \mathbb{E}\left[u\left(a, \theta\right) \mid s, m \in M^c; \sigma'_S\right]$$

be the action which maximizes R's utility (on observing s) if R observed only that some cheap talk message was sent under message profile σ'_S (but not the specific realization).

Suppose for a contradiction that for any such message profile σ'_S that $a\left(m^c, s'\right) = a^*\left(\theta'\right)$ for all s and all cheap talk messages $m^c \in M^c$.[39] Then, we must have for all s that

$$\mathbb{E}\left[u\left(a^*\left(\theta'\right), \theta\right) \mid s, m = m^c; \sigma'_S\right] \geq \mathbb{E}\left[u\left(a_s^{**}, \theta\right) \mid s, m = m^c; \sigma'_S\right]$$

Taking expectations over m^c yields

$$\mathbb{E}\left[u\left(a^*\left(\theta'\right), \theta\right) \mid s, m \in M^c; \sigma'_S\right] \geq \mathbb{E}\left[u\left(a_s^{**}, \theta\right) \mid s, m \in M^c; \sigma'_S\right]$$

for all s, which is a contradiction to the definition of minimal persuasion, (4). Therefore, there must exist a $m^c \in M^c$ such that $a_{\sigma'_S}\left(m^c, s'\right) > a^*\left(\theta'\right)$. $\qquad\square$

Lemma 4 establishes that there must be a cheap talk message for which R plays actions strictly preferred to the 'worst case' for disclosure for any σ'_S for which $m \in M^c$ when $\theta > \theta_P$. Now, consider any such message profile and evaluate type θ_P net payoff from playing $m = m^c$ over any

---

[39] As we have argued above, any $a\left(m', s'\right) < a^*\left(\theta'\right)$ is dominated and therefore not admissible.

verifiable message $m^{\theta_P} \in M(\theta_P) \setminus M^c$. We have

$$\Delta U\left(\theta_P \mid \sigma'_S, a_{\sigma'_S}\right) = \sum_{s \in S} g(s \mid \theta_P) \cdot \left[v\left(a'(m^c, s), \theta_P\right) - v\left(a\left(m^{\theta_P}, s\right), \theta_P\right)\right] - \delta(\theta_P)$$

However, since no type $\theta > \theta_P$ plays any verifiable message, log-monotonicity of u implies that $a_{\sigma'_S}\left(m^{\theta_P}, s\right) \leq a^*(\theta_P)$ for any $m^{\theta_P} \in M(\theta_P) \setminus M^c$ that is played in equilibrium. Moreover, we know that $a_{\sigma'_S}(m^c, s) \geq a^*(\theta')$ for all s and there exists at least one s' for whom $a_{\sigma'_S}(m^c, s') > a^*(\theta'')$ where $s \in \underline{S}(\theta'')$. Moreover, for all ε sufficiently small, type θ_P must send this signal with strictly positive probability.[40] Thus,

$$\begin{aligned}
\Delta U\left(\theta_P \mid \sigma'_S, a_{\sigma'_S}\right) &\geq g\left(s' \mid \theta_P\right) \cdot \left[v\left(a'(m^c, s'), \theta_P\right) - v\left(a^*(\theta''), \theta_P\right)\right] + \\
&\quad \sum_{\theta' \leq \theta_P} \Pr\left(s \in \underline{S}(\theta')\right) \cdot \left[v\left(a^*(\theta'), \theta_P\right) - v\left(a^*(\theta_P), \theta_P\right)\right] - \delta(\theta_P) \\
&\geq g\left(s' \mid \theta_P\right) \cdot \underline{B} + \min_{m(\theta), a(m,s)} \Delta U\left(\theta_P \mid m(\theta), a(m,s)\right)
\end{aligned}$$

where the last line follows from the fact that there exists a lower bound $\underline{B} > 0$ such that for any two actions, $v(a, \theta) - v(a', \theta) \geq \underline{B}$, and noting that

$$\sum_{\theta' \leq \theta_P} \Pr\left(s \in \underline{S}(\theta')\right) \cdot \left[v\left(a^*(\theta'), \theta_P\right) - v(a^*(\theta_P), \theta_P)\right] - \delta(\theta_P) = \min_{m(\theta), a(m,s)} \Delta U\left(\theta_P \mid m(\theta), a(m,s)\right)$$

where the minimum on the RHS is defined as in (23). Thus, $\Delta U\left(\theta_P \mid \sigma'_S, a_{\sigma'_S}\right)$ is positive for any disclosure rule σ'_S with $m \in M^c$ when $\theta > \theta_P$ if

$$\frac{g\left(s' \mid \theta_P\right) \cdot \underline{B}}{1 + \#S \cdot \overline{B}} \geq \varepsilon$$

Since $g\left(s' \mid \theta_P\right)$ is strictly positive, while $\#S$, $\overline{B}$ and $\underline{B} > 0$ are finite constants, we have established that there exists an ε small enough that type θ_P will always prefer to send only cheap talk messages when types $\theta > \theta_P$ also exclusively play messages from M^c. $\qquad\square$

Proof of Proposition 3

A more rigorous statement of Proposition 3 is as follows: Let $\Gamma = (u, v, \Theta, f, S, g, A, M)$ denote some Sender-Receiver game, with extensive form as specified in Section 2. When we consider a perturbation of game Γ in which only the public signal structure changes from (S, g) to some (S', g') we will use the notation $\Gamma' = (u, v, \Theta, f, S', g', A, M)$ to refer to this alternative game.

---

[40]Since some types $\theta > \theta_P$ must send the signal with strictly positive probability in order for it to be minimally persuasive.

Proposition. *Suppose there is an equilibrium with full disclosure in some game Γ. There exists a public signal (S', g') which is (1) a Blackwell-improvement over (S, g) but for which (2) in any equilibrium of Γ', R's received information is Blackwell-deterioration relative to the most informative equilibrium of Γ.*

Proof. First, note that in game Γ (with public signal (S, g)), the equilibrium in which R receives the most information must be the full disclosure equilibrium, since R always learns Θ for sure before acting. Moreover, since the transparent equilibrium exists in game Γ, Proposition 1 implies

$$P_{\Gamma}\left(\overline{\theta}\right) > \delta\left(\overline{\theta}\right)$$

where we use $P_{\Gamma}(\theta)$ to denote type θ's maximal punishment under game Γ. In other words, for type $\overline{\theta}$

$$\sum_{\theta' < \overline{\theta}} \Pr\left(s \in \underline{S}\left(\theta'\right) \mid \theta\right) \times \left[v\left(a^*\left(\left(\overline{\theta}\right), \overline{\theta}\right) - v\left(a^*\left(\theta_P\right), \overline{\theta}\right)\right)\right] > \delta\left(\overline{\theta}\right)$$

Now, consider a perturbation of the public signal to some (S', g') given by:

$$g'\left(s \mid \theta\right) = \begin{cases} g\left(s \mid \theta\right) & \forall s \in S, \, if \, \theta < \overline{\theta} \\ p.g\left(s \mid \theta\right) & \forall s \in S, \, if \, \theta = \overline{\theta} \\ (1-p).\sum_{s \in S} g\left(s \mid \theta\right) & for \, some \, \tilde{s} \in \underline{S}\left(\overline{\theta}\right) \cup S' \setminus S, \, if \, \theta = \overline{\theta} \end{cases}$$

Public signal (S', g') leaves the signal generating process unchanged relative to (S, g) for any types $\theta < \overline{\theta}$. However, type $\overline{\theta}$'s signal structure now undergoes a 2-stage lottery. With probability p, the signal generating process remains as it was under (S, g). However, with probability $(1 - p)$, the new public signal forces type $\overline{\theta}$ to send a public signal that perfectly reveals his type.

It can be shown that (S', g') is a Blackwell-improvement of (S, g). Indeed, posteriors over θ under (S', g') are a mean preserving spread of those under (S, g):

$$\mathbb{E}_{g'}\left[\mathbb{E}_{g'}\left[\theta \mid s'\right] \mid s' \in \{s, s^*\}\right] = \mathbb{E}_{g}\left[\theta \mid s\right]$$

where $\mathbb{E}_{\gamma}$ denotes expectations under measure γ.

Now, consider type $\overline{\theta}$'s largest marginal benefit from disclosure when the signal structure is

(S', g')

$$P_{\Gamma'}(\overline{\theta}) - \delta(\overline{\theta}) = \sum_{\theta' < \overline{\theta}} \Pr\left(s \in \underline{S}(\theta') \mid \theta; g'\right) \times \left[v\left(a^*(\overline{\theta}, \overline{\theta}) - v\left(a^*(\theta_P), \overline{\theta}\right)\right)\right] - \delta(\overline{\theta})$$

$$= p. \sum_{\theta' < \overline{\theta}} \Pr\left(s \in \underline{S}(\theta') \mid \theta; g\right) \times \left[v\left(a^*(\overline{\theta}, \overline{\theta}) - v\left(a^*(\theta_P), \overline{\theta}\right)\right)\right] - \delta(\overline{\theta})$$

$$= p. P_{\Gamma}(\overline{\theta}) - \delta(\overline{\theta}) \tag{26}$$

where we denote probabilities generated under measure γ by $\Pr(\cdot \mid \cdot; \gamma)$ and the second line follows since $g'(s \mid \overline{\theta}) = p. g(s \mid \overline{\theta})$ for all $s \neq \tilde{s}$. By inspection of (26), clearly we can always find a $\overline{p} > 0$ small enough that

$$P_{\Gamma'}(\overline{\theta}) - \delta(\overline{\theta}) = \overline{p}. P_{\Gamma}(\overline{\theta}) - \delta(\overline{\theta}) < 0$$

But then we can apply Proposition 1 to argue in any equilibrium that $\sigma_S^*(\overline{\theta}) \in \Delta(M^c)$. Moreover, it is trivial to show that in any equilibrium $\sigma_S^*(\underline{\theta}) \in \Delta(M^c)$ and that, if $\overline{\theta}$ plays some message $m^c \in M^c$ with strictly positive probability, then so too does type $\underline{\theta}$.

For any signal $m' \in M$ we can write

$$\mathbb{E}_{\Gamma}\left[\mathbb{E}_{\Gamma}\left[\theta \mid \sigma_{\Gamma}^{FD}(\theta)\right]; \sigma_{\Gamma'}^* = m'\right] = \mathbb{E}_{\Gamma'}\left[\theta \mid \sigma_{\Gamma'}^* = m'\right]$$

where σ_{Γ}^{FD} denotes a full disclosure message profile in game Γ (assumed to be an equilibrium of Γ), σ_{Γ}^* is some arbitrary equilibrium message profile in game Γ', $\mathbb{E}_{\Gamma}$ denotes (with notational abuse) expectations with respect to probabilities induced by equilibrium play in game Γ. In other words, conditioning on a 'prior' induced by the set of types playing message m' in game Γ, R's posterior in the full disclosure equilibrium of game Γ is a MSP of that in Γ'.

Moreover, the reverse cannot be true since the full disclosure equilibrium induces degenerate posteriors while any equilibrium of Γ' necessarily does not induce degenerate priors over all types. Therefore, the most informative equilibrium of Γ Blackwell dominates any equilibrium of Γ'. $\qquad \square$

Proof of Lemma 1

Lemma. *In any equilibrium, there is a cutoff θ_1 such that the Sender discloses his type if and only if $\theta \in [c, \theta_1]$.*

Proof. We show first that R's best response to any arbitrary (potentially mixed) disclosure strategy $\sigma_S : \Theta \to \Delta M$,[41] $a_{\sigma_S}^*(m, s)$, satisfies:

1. *if $m = \theta$, $a = 1$ if and only if $\theta \geq c$.*[42]

---

[41] Which we restrict to obey the technological constraint that $m(\theta) \in \{\emptyset, \theta\}$.

[42] Strictly, R is indifferent between $a = 0$ or $a = 1$ when $m = c$. However, since such an event is measure 0 nothing

2. *if $m = \emptyset$, there exists $s^* \in \left[\underline{s}(\theta), \overline{s}(\theta)\right]$ such that R plays $a = 1$ if and only if $s \geq s^*$.*[43]

Suppose $m = \theta$. Then, since this verifiable message could only have been sent by type θ, R's net payoff from playing $a = 1$ is $\gamma(\theta)$. By assumption, $\gamma(\theta) \geq 0$ if and only if $\theta \geq c$ and the result follows immediately.

Now, suppose $m = \emptyset$ and consider R's posterior belief, $f(\theta \mid s, m = \emptyset; \sigma_S)$. For any (potentially mixed) disclosure strategy, this posterior can be written by Bayes' rule as

$$
\begin{aligned}
f(\theta \mid s, m = \emptyset; \sigma_S) &= f(\theta \mid m = \emptyset; \sigma_S) \cdot \frac{h(s \mid \theta, m = \emptyset; \sigma_S)}{\int_{\theta'} h(s \mid \theta', m = \emptyset; \sigma_S) \cdot f(\theta \mid m = \emptyset; \sigma_S) \cdot d\theta'} \\
&= f(\theta \mid m = \emptyset; \sigma_S) \cdot \frac{h(s \mid \theta)}{\int_{\theta'} h(s \mid \theta') \cdot f(\theta \mid m = \emptyset; \sigma_S) \cdot d\theta'}
\end{aligned}
\tag{27}
$$

where the second line follows because θ is sufficient for σ_S and $m = \emptyset$ (recall that S can only condition his disclosure strategy on θ). Written this way, (27) makes clear that we can treat R's posterior given $s, , m = \emptyset, \sigma_S$ as her updated belief following a 'prior' on θ given by $f(\theta \mid s, m = \emptyset; \sigma_S)$. But, since $h(s \mid \theta)$ satisfies MLRP by assumption, it follows from Milgrom (1981), Proposition 1 that for any signals s', s where $s' < s$, we have

$$
F(\theta \mid s, m = \emptyset; \sigma_S) \leq F\left(\theta \mid s', m = \emptyset; \sigma_S\right)
$$

for all $\theta \in \Theta$. In other words, R's posterior after s first order stochastically dominates that after s'. Therefore, since γ is an increasing function, R's net payoff to taking action $a = 1$ satisfies

$$
\mathbb{E}\left[\gamma(\theta) \mid s', m = \emptyset; \sigma_S\right] \leq \mathbb{E}\left[\gamma(\theta) \mid s, m = \emptyset; \sigma_S\right]
$$

which establishes the second claim above.

Now, given that R's best response to any disclosure strategy takes a cut-off form, where she sets $a = 1$ iff $s \geq s^*$ for some s^* (on observing $m = \emptyset$), S's net payoff from disclosure must satisfy

$$
\mathbb{E}[\beta(\theta) - \delta(\theta) \mid \theta] = \begin{cases} -\delta(\theta) & \text{if } \theta < c \\ \beta(\theta) \cdot \left(1 - H\left(\frac{s^* - \theta}{\sigma}\right)\right) - \delta(\theta) & \text{if } \theta \geq c \end{cases}
$$

Clearly, it is always optimal for types $\theta < c$ to play $m = \emptyset$. Finally, under our regularity condition, $\beta(\theta) \cdot \left(1 - H\left(\frac{s^* - \theta}{\sigma}\right)\right) - \delta(\theta)$ has at most a signal crossing on $\theta \in \left[c, \overline{\theta}\right]$ and this crossing is from above. This establishes the existence of a $\theta_1 \in \left[c, \overline{\theta}\right]$ such that S discloses if and only if $\theta \in [c, \theta_1]$. $\qquad\square$

---

significant in our model relies on the assumption that ties are broken in favor of S's favoured action.

[43] Again, while R will be indifferent between $a = 0$ or $a = 1$ when $s = s^*$, such an event is measure 0.

Proof of Proposition 4

We begin by showing that a transparent equilibrium exists if and only if (5) is satisfied. For necessity (only if), suppose that (5) is violated. If the Sender stays quiet, the Receiver will believe that $\theta \geq c$ with probability 1, and thus take the high action, if the public signal $s > \bar{s}(c)$. Normalizing $u(0, \theta) = 0$ without loss of generality, the payoff of type $\bar{\theta}$ from staying quiet is therefore at least

$$\left[1 - H\left(\frac{\bar{s}(c) - \bar{\theta}}{\sigma}\right)\right]\beta(\bar{\theta}), \tag{28}$$

while disclosure gives this type a payoff of $\beta(\bar{\theta}) - \delta(\bar{\theta})$. Staying quiet yields a higher payoff than disclosure for type $\bar{\theta}$, implying that transparency cannot be an equilibrium, as long as

$$\left[1 - H\left(\frac{\bar{s}(c) - \bar{\theta}}{\sigma}\right)\right]\beta(\bar{\theta}) - (\beta(\bar{\theta}) - \delta(\bar{\theta})) > 0$$
$$\Leftrightarrow \delta(\bar{\theta}) - \beta(\bar{\theta})L(\bar{\theta}) > 0,$$

which is true since (5) is violated by assumption.

For sufficiency (if), suppose that (5) is satisfied. We establish that transparency is an equilibrium by checking that no type $\theta \geq c$ has an incentive to deviate from disclosure. Our regularity condition ensures that it is sufficient to check that type $\bar{\theta}$ has no incentive to deviate. In a transparent equilibrium, Bayes' rule dictates that whenever $s \leq \bar{s}(c)$, the Receiver must believe that $\theta < c$ with probability 1 and take the low action. Thus the payoff of type $\bar{\theta}$ from deviating and staying quiet is at most the expression in (28), while his equilibrium payoff is again $\beta(\bar{\theta}) - \delta(\bar{\theta})$. He has no incentive to deviate if and only if

$$\left[1 - H\left(\frac{\bar{s}(c) - \bar{\theta}}{\sigma}\right)\right]\beta(\bar{\theta}) - (\beta(\bar{\theta}) - \delta(\bar{\theta})) \leq 0$$
$$\Leftrightarrow \delta(\bar{\theta}) - \beta(\bar{\theta})L(\bar{\theta}) \leq 0,$$

which is equivalent to our assumption that (5) is satisfied.

Next, we show that when (5) is violated, the highest disclosing type θ_1 is bounded strictly away from $\bar{\theta}$ in equilibrium. For this part we use the best response function defined in the text. More precisely, define $s^\star(\theta_1)$ as the highest public signal s satisfying $E[\theta|\theta \notin [c, \theta_1], s] \geq c$ (if this is impossible let $s^\star(\theta_1) = \bar{s}(\bar{\theta})$). Define $BR(\theta_1)$ as the highest type θ satisfying $H\left(\frac{s^\star(\theta_1) - \theta}{\sigma}\right)\beta(\theta) \geq \delta(\theta)$ (if this is impossible, let $BR(\theta_1) = c$). It is easy to see that a cutoff θ_1 induces an equilibrium if and only if

$$BR(\theta_1) = \theta_1. \tag{29}$$

Suppose that (5) is violated. From the first part of this proof, we know that $BR(\overline{\theta}) < \overline{\theta}$ (otherwise transparency would be an equilibrium). If $BR(\overline{\theta}) = c$ then $\theta_1 = c$ is the unique equilibrium and we are done. If $\overline{\theta} > BR(\overline{\theta}) > c$ then by continuity, $BR(\theta)$ is interior in a neighborhood of $\overline{\theta}$, and in this neighborhood it solves the implicit equation

$$BR(\theta_1) = s^\star(\theta_1) - \Psi(BR(\theta_1)),$$

where we have defined $\Psi(\theta) \equiv H^{-1}\left(\frac{\delta(\theta)}{\beta(\theta)}\right)$. The implicit function theorem yield the derivative of $BR(\theta_1)$ in this neighborhood,

$$BR'(\theta_1) = \frac{1}{1 + \Psi'(BR(\theta_1))} \times \frac{ds^\star(\theta_1)}{d\theta_1}.$$

To complete the proof it is sufficient to show that $\lim_{\theta_1 \to \overline{\theta}} BR'(\theta_1) = \infty$. If this is the case, then there is no fixed point solving (29) in a neighborhood of $\overline{\theta}$.

We know that $\lim_{\theta_1 \to \overline{\theta}} BR(\theta_1) = BR(\overline{\theta}) \in (c, \overline{\theta})$ so that the first factor $\frac{1}{1 + \Psi'(BR(\theta_1))}$ converges to a finite positive number. It is easy to see that $s^\star(\theta_1)$ converges to $\overline{s}(c)$. By continuity, $s^\star(\theta_1)$ is interior in a neighborhood of $\overline{\theta}$, and thus solves $E[\theta | \theta \notin [c, \theta_1], s^\star(\theta_1)] = 0$ which implies that

$$\int_{s^\star(\theta_1) - \sigma\overline{\varepsilon}}^{c} (\theta - c)f(\theta | s^\star(\theta_1))d\theta + \int_{\theta_1}^{\overline{\theta}} (\theta - c)f(\theta | s^\star(\theta_1))d\theta = 0,$$

where $f(\theta | s) = \frac{g(s | \theta)f(\theta)}{\int g(s | \theta')f(\theta')d\theta'}$ denotes the conditional density of θ given s, and the first integral uses the fact that $f(\theta | s^\star(\theta_1)) = 0$ for all θ such that $s^\star(\theta_1) > \overline{s}(\theta)$, i.e. for all $\theta < s^\star(\theta_1) - \sigma\overline{\varepsilon}$. Totally differentiating with respect to θ_1 we obtain

$$\frac{ds^\star(\theta_1)}{d\theta_1} = \theta_1 f(\theta_1 | s^\star(\theta_1)).$$

$$\times \left\{ \int_{[s^\star(\theta_1) - \sigma\overline{\varepsilon}, c] \cup [\theta_1, \overline{\theta}]} (\theta - c)\frac{\partial f(\theta | s^\star(\theta_1))}{\partial s}d\theta - [s^\star(\theta_1) - \sigma\overline{\varepsilon} - c]f(s^\star(\theta_1) - \sigma\overline{\varepsilon} | s^\star(\theta_1)) \right\}^{-1}$$

As $\theta_1 \to \overline{\theta}$, the second term on the RHS converges to zero since $s^\star(\theta_1) \to \overline{s}(c)$ is equivalent to $s^\star(\theta_1) - \sigma\overline{\varepsilon} \to c$. The first converges to $\overline{\theta}f(\overline{\theta} | \overline{s}(c))$, a finite number. Thus $\frac{ds^\star(\theta_1)}{d\theta_1} \to \infty$ which implies $BR'(\theta_1) \to \infty$.

Proof of Proposition 5

We fully define the concept of stability in Appendix B. Given that definition and the expositional assumption that $\frac{\delta(\theta)}{\beta(\theta)} = \rho$ for some $\rho \in (0,1)$, the proof of the Proposition proceeds as follows:

Proof. We prove sufficiency by arguing the contrapositive: if transparency is stable, then there must exist a $\tilde{K} \in \mathbb{R}$ that violates (7). Thus, suppose that transparency is a stable equilibrium. Then by Lemma 5, we can find a B such that $\forall \theta' \geq B$

$$BR(\theta') - \theta' > 0 \tag{30}$$

Now by definition $BR(\theta')$ is the best response of S to a disclosure strategy of θ', when R plays her best response $s^*(\theta')$. Therefore, it satisfies

$$\frac{\delta(BR(\theta))}{\beta(BR(\theta))} = H\left(\frac{s^*(\theta') - BR(\theta')}{\sigma_\varepsilon}\right)$$

or

$$BR(\theta') = s^*(\theta') - \sigma_\varepsilon H^{-1}\left(\frac{\delta(\theta')}{\beta(\theta')}\right) \tag{31}$$

Substituting (31) into (30) yields a lower bound on $s^*(\theta')$ as a function of θ' for any transparent equilibrium:

$$s^*(\theta') > \theta' + \sigma_\varepsilon H^{-1}\left(\frac{\delta(\theta')}{\beta(\theta')}\right) \tag{32}$$

Further, recall that $s^*(\theta')$ satisfies

$$\mathbb{E}\left[\gamma(\theta)|s^*(\theta'), \theta \notin [c, \theta']\right] = 0$$

or

$$\Pr\left(\theta \leq c \mid s^*(\theta')\right).\mathbb{E}\left[\gamma(\theta)|\theta \leq c, s^*(\theta')\right] + \Pr\left(\theta \geq \theta'|s^*(\theta')\right).\mathbb{E}\left[\gamma(\theta)|\theta \geq \theta', s^*(\theta')\right] = 0 \tag{33}$$

Now consider the left hand side of Equation (7) evaluated at $\tilde{K} = \sigma_\varepsilon H^{-1}(\rho)$, where $\rho = \frac{\delta(\theta)}{\beta(\theta)}$, $\forall \theta \in \Theta$. By (32), $s^*(\theta') > \theta' + \tilde{K}$. Then, it follows immediately from (33) and the MLRP assumption on signals s that

$$\Pr\left(\theta \notin [c, \theta'] \mid s = \theta' + \tilde{K}\right).\mathbb{E}\left[\gamma(\theta)|s = \theta' + \tilde{K}, \theta \notin [0, \theta']\right]$$
$$= \Pr\left(\theta \leq c \mid s = \theta' + \tilde{K}\right).\mathbb{E}\left[\gamma(\theta)|\theta \leq c, s = \theta' + \tilde{K}\right]$$
$$+ \Pr\left(\theta \geq \theta'|s = \theta' + \tilde{K}\right).\mathbb{E}\left[\gamma(\theta)|\theta \geq \theta', s = \theta' + \tilde{K}\right] < 0$$

Rearranging this expression yields, for any $\theta' \in \mathbb{R}$:

$$-\left(\frac{\Pr\left(\theta \geq \theta'|s=\theta'+\tilde{K}\right).\mathbb{E}\left[\gamma(\theta)|\theta \geq \theta', s=\theta'+\tilde{K}\right]}{\Pr\left(\theta \leq c \mid s=\theta'+\tilde{K}\right).\mathbb{E}\left[\gamma(\theta)|\theta \leq c, s=\theta'+\tilde{K}\right]}\right) < 1 \tag{34}$$

This final inequality shows our contrapositive claim that if transparency is a stable equilibrium then for $K \leq \tilde{K} = \sigma_\varepsilon H^{-1}(\rho)$, (7) is violated.[44]

(Necessity) We prove the argument by contradiction, in each of two cases. Suppose then that (7) does not hold, but that the transparent equilibrium is stable for all choices of $\rho = \frac{\delta(\theta)}{\beta(\theta)} > 0$. Then there exists some $\tilde{K}, \tilde{\theta}' \in \mathbb{R}$ such that[45]

$$-\frac{\Pr\left(\theta \geq \theta'|s=\theta'+\tilde{K}\right).\mathbb{E}\left[\gamma(\theta)|\theta \geq \theta', s=\theta'+\tilde{K}\right]}{\Pr\left(\theta \leq c \mid s=\theta'+\tilde{K}\right).\mathbb{E}\left[\gamma(\theta)|\theta \leq c, s=\theta'+\tilde{K}\right]} \leq 1 \tag{35}$$

$\forall \theta' \geq \tilde{\theta}'$. Note that (35) holds everywhere, not just in the limit, since the infimum is a non-decreasing function. Denoting for simplicity,

$$P_{wer}\left(\theta', \tilde{K}\right) = -\frac{\Pr\left(\theta \geq \theta'|s=\theta'+\tilde{K}\right).\mathbb{E}\left[\gamma(\theta)|\theta \geq \theta', s=\theta'+\tilde{K}\right]}{\Pr\left(\theta \leq c \mid s=\theta'+\tilde{K}\right).\mathbb{E}\left[\gamma(\theta)|\theta \leq c, s=\theta'+\tilde{K}\right]}$$

we now consider the following two exhaustive cases:

1. $\exists \tilde{\theta}' \in \mathbb{R}$ such that $P_{wer}\left(\theta', \tilde{K}\right) < 1, \forall \theta' \geq \tilde{\theta}'$;

2. $\limsup_{\theta' \to \infty} P_{wer}\left(\theta', \tilde{K}\right) \geq 1$

Case 1.

We argue that, under condition (35), $\exists \bar{\rho} > 0$ such that transparency is a stable outcome for all $\rho < \bar{\rho}$. Specifically, for all $\theta' \geq \tilde{\theta}'$, we know that

$$P_{wer}\left(\theta', \tilde{K}\right) < 1$$

which can be equivalently expressed as

$$\begin{aligned} 0 \quad > \quad & \Pr\left(\theta \leq c \mid s=\theta'+\tilde{K}\right).\mathbb{E}\left[\gamma(\theta)|\theta \leq c, s=\theta'+\tilde{K}\right] \\ & + \Pr\left(\theta \geq \theta'|s=\theta'+\tilde{K}\right).\mathbb{E}\left[\gamma(\theta)|\theta \geq \theta', s=\theta'+\tilde{K}\right] \end{aligned}$$

or

$$\mathbb{E}\left[\theta|\theta \notin [c, \theta'], s=\theta'+\tilde{K}\right] < 0 \tag{36}$$

---

[44]MLRP implies that if (34) holds for $\tilde{K}$, then it also holds for all $K \leq \tilde{K}$.

[45]Note by MLRP that if (35) holds for $\tilde{K}$, then it also holds for all $K \leq \tilde{K}$.

Therefore, since s satisfies the MLRP condition, (36) implies that $s^*(\theta') > \theta' + \tilde{K}$ for all $\theta' \geq \tilde{\theta}'$ or

$$s^*(\theta') - \theta' > \tilde{K} \tag{37}$$

Now, choose $\bar{\rho}$ that solves $\tilde{K} = \sigma_\varepsilon H^{-1}(\bar{\rho})$. But, for any $\rho \leq \bar{\rho}$, $BR(\theta')$ must satisfy

$$s^*(\theta') - BR(\theta') = \sigma_\varepsilon H^{-1}\left(\frac{\delta(\theta')}{\beta(\theta')}\right) \geq \tilde{K} \tag{38}$$

Comparing (37) and (38) establishes that for all $\theta' \geq \tilde{\theta}'$, $BR(\theta') > \theta$' - a contradiction to the assumed instability of the transparent equilibrium.

Case 2.

We argue that under condition (35), $\exists \bar{\rho} > 0$ (defined as above) such that transparency is a *neutrally* stable outcome for all $\rho \leq \bar{\rho}$: there exists $\tilde{\theta}'$ such that for any sequence $\{\theta'_n\}_{n=1}^\infty \to \infty$, $\theta_n > \tilde{\theta}'$, $\forall n$, there is a sequence of equilibria $\{\theta^*_n\}_{n=1}^\infty$ such that *(i)* starting from a perturbation θ'_n, best response dynamics converge to equilibrium θ^*_n; and *(ii)* $\lim_{n\to\infty} \theta^*_n = \infty$.

First, given $\tilde{K}$ from (35), we can find always find a sequence of values $\{\tilde{\theta}'_n\}_{n=1}^\infty \to \infty$ such that

$$P_{wer}\left(\tilde{\theta}'_n, \tilde{K}\right) \leq 1$$

$\forall n$. Likewise, we can find a similar sequence $\{\tilde{\theta}''_n\}_{n=1}^\infty \to \infty$ such that

$$P_{wer}\left(\tilde{\theta}'_n, \tilde{K}\right) \geq 1$$

Given these sequences, it is also always possible to construct subsequences $\{\tilde{\theta}'_q\}_{q=1}^\infty \subset \{\tilde{\theta}'_n\}_{n=1}^\infty$ and $\{\tilde{\theta}''_q\}_{q=1}^\infty \subset \{\tilde{\theta}''_n\}_{n=1}^\infty$ such that $\tilde{\theta}'_q \leq \tilde{\theta}''_q \leq \tilde{\theta}'_{q+1}$.[46] Now, the increasing sequence

$$\{\theta'_q : \theta'_q = \tilde{\theta}'_q \text{ if } q/2 \in \mathbb{Z}; \theta'_q = \tilde{\theta}''_q \text{ otherwise}\}_{q=1}^\infty \to \infty$$

defines intervals $\left[\theta'_q, \theta'_{q+1}\right]$. By the assumed continuity of $\mathbb{E}[\gamma(\theta) \mid s, \theta \notin [c, \theta']]$ in s, θ', there must exist at least one $\theta^*_q \in \left[\theta'_q, \theta'_{q+1}\right]$ such that $P_{wer}\left(\theta^*_q, \tilde{K}\right) = 1$, $\forall q$. In other words,

$$\mathbb{E}\left[\gamma(\theta) \mid \theta^*_q + \tilde{K}, \theta \notin \left[c, \theta^*_q\right]\right] = 0$$

or $s^*\left(\theta^*_q\right) = \theta^*_q + \tilde{K}$. Setting $\bar{\rho}$ to solve $\tilde{K} = \sigma_\varepsilon H^{-1}(\bar{\rho})$ establishes that these values of $\{\theta^*_q\}_{q=1}^\infty$ are equilibria when $\rho = \bar{\rho}$.

---

[46]The construction is simple. Take $\tilde{\theta}'_{q=1} = \tilde{\theta}'_{n=1}$. Then pick the lowest n' such that $\tilde{\theta}''_{n'} > \tilde{\theta}_{q=1}$ (which always exists, since $\lim \tilde{\theta}''_n = \infty$) and set $\tilde{\theta}''_{q=1} = \tilde{\theta}''_{n'}$. Proceeding iteratively yields the result.

From the proof of Lemma 5, we have established that for any $\theta_n \in \left[\theta'_q, \theta'_{q+1}\right]$, best response dynamics imply converegence from θ_n to an equilibrium $\theta_n^* \in \left[\theta^*_{q-1}, \theta^*_{q+1}\right]$. Since $\lim \theta'_q = \lim \theta^*_q = \infty$, any sequence $\theta_n \to \infty$ defines a sequence of equilibria $\theta_n^* \to \infty$ which satisfy the conditions required.

Finally, since $P_{wer}\left(\tilde{\theta}, K\right)$ is decreasing in K, then for any $K < \tilde{K}$ we are either in case 1. or case 2. The same arguments can then be made to show that the transparent equilibrium is at least neutrally stable for all $\rho < \bar{\rho}$. This contradicts out assumption that the transparent equilibrium was unstable. $\qquad\qquad\square$

Proof of Proposition 6

Proof. In any equilibrium, S's disclosure strategy is of the form $m = \theta$ iff $\theta \in [c, \theta_1]$ for some $\bar{\theta} \geq \theta_1 \geq c$. If equilibrium is unique, then the first claim of the Proposition holds trivially. Suppose there are $n \geq 1$ equilibria, represented by cut-offs $\left\{\theta_{1,1}, \ldots, \theta_{1,n}\right\}$ satisfying $\theta_{1,1} < \theta_{1,2} < \cdots < \theta_{1,n}$.

Consider two equilibria i and j such that $\theta_{1,i} < \theta_{1,j}$. In any equilibrium $\theta_{1,\cdot}$, R observes

$$I\left(\cdot\right) = \begin{cases} s. \text{ with probability } g\left(s.|\theta\right), & for\ \theta \notin [c, \theta_{1,\cdot}] \\ \theta, & for\ \theta \notin [c, \theta_{1,\cdot}] \end{cases}$$

Clearly, R gets the same information in equilibria i and j for any $\theta \in [c, \theta_{1,i}]$ since he observes θ in both cases. For any $\theta \in [\theta_{1,i}, \theta_{2,i}]$, R observes s under i and θ under j, while s is observed under both otherwise. R's posterior belief under s can be written

$$\begin{aligned}
\Pr\left[\theta|s, \theta \notin [c, \theta_{1,i}]\right] &= \Pr\left(\theta \in \left(\theta_{1,i}, \theta_{1,j}\right) |s\right) \Pr_i\left[\theta|s, \theta \notin \left[\theta_{1,i}, \theta_{1,j}\right]\right] \\
&+ \Pr\left(\theta \in \left(\theta_{1,i}, \theta_{1,j}\right) |s\right) \Pr_i\left[\theta|s, \theta \notin \left[\theta_{1,i}, \theta_{1,j}\right]\right]
\end{aligned}$$

while R's posterior belief under equilibrium j is $\Pr_j\left(\theta|s, \theta\right) = 1$ for $\theta \in \left(\theta_{1,i}, \theta_{1,j}\right)$ and otherwise, $\Pr_i\left[\theta|s, \theta \notin \left[\theta_{1,i}, \theta_{1,j}\right]\right]$. By inspection, it should be clear that R's posterior belief under equilibrium j is a MPS of that under i, which establishes the first part of the claim.

To establish the second part of the claim, we construct an example. Find the most informative equilibrium $[c, \theta_1]$, which exists given the the argument above. Suppose that $\theta_1 < \infty$ and consider the following perturbation to the public signal $g\left(s|\theta\right)$ following a message $m = \emptyset$.[47] Namely, for types in $[\theta_1 - \varepsilon, \theta_1]$, $0 < \varepsilon < \theta_1 - c$, we allow R to observe the outcome of the following two-stage lottery: In round one, θ is revealed with probability p. If θ was not revealed in round 1, then R observes s according to the probability distribution $g\left(s|\theta\right)$.

---

[47]The case $\theta_1 = \infty$ is a simple extension and therefore omitted.

As in the proof of Proposition 3, it is trivial to show that there exists $\overline{p} < 1$ such that all types $\theta \in [\theta_1 - \varepsilon, \theta_1]$ will play a best response of $m = \emptyset$ as a dominant strategy. Consider R's best response given this change. R plays $a = 1$ iff

$$\Pr(\theta \notin [c, \theta_1]) \mathbb{E}[\gamma(\theta)|s; \theta \notin [c, \theta_1]] + (1-p) \Pr(\theta \in [\theta_1 - \varepsilon, \theta_1]) \mathbb{E}[\gamma(\theta)|s; \theta \in [\theta_1 - \varepsilon, \theta_1]] \geq 0 \tag{39}$$

But on observing signal $s^*(\theta_1)$, (39) must hold as a strict inequality since $\mathbb{E}[\gamma(\theta)|s; \theta \in [c, \theta_1)] = 0$ while $\gamma(\theta) > 0$ on $[\theta_1 - \varepsilon, \theta_1]$. By MLRP of signal s, it must be that R's best response in the perturbed game,[48] denoted $\tilde{s}^*(\theta_1)$, satisfies $\tilde{s}^*(\theta_1) < s^*(\theta_1)$. Moreover the same argument can be employed to show that $\tilde{s}^*(\theta) < s^*(\theta)$ for any $\theta \geq \theta_1$.

Since θ_1 is the most informative equilibrium, we must have $BR(\theta) < \theta$ for all $\theta \geq \overline{\theta}$ - otherwise, there would exist a more transparent equilibrium (potentially the full disclosure equilibrium), which would contradict the definition of θ_1. Since BR is increasing in s^*, clearly any equilibrium of the pertubed game $\tilde{\theta}_1$ must satisfy $\tilde{\theta}_1 < \theta_1$.

Finally, an almost identical argument to that used to prove the first part of the Proposition establishes that $\tilde{\theta}_1$ is a Blackwell deterioration of R's equilibrium information, relative to θ_1. $\square$

Proof of Proposition 7

Proof. In any equilibrium, we have by definition that S's best response satisfy the fixed point problem

$$BR(\theta_1; \sigma) = \theta_1 \tag{40}$$

Totally differentiating both sides of (40) with respect to σ yields (8) upon rearrangement.

Lemma 5, Appendix B implies that in any stable equilibrium $\frac{\partial BR(\theta_1; \sigma)}{\partial \theta_1} < 1$. Therefore, (8) has the same sign as $\frac{\partial BR(\theta_1; \sigma)}{\partial \sigma}$. To evaluate the sign of this derivative, we first rearrange and then partially differentiate (11) with respect to σ to yield

$$H^{-1}\left(\frac{\delta(\theta)}{\beta(\theta)}\right) = \frac{\partial s^*}{\partial \sigma}(\theta_1; \sigma) - \frac{\partial BR(\theta_1; \sigma)}{\partial \sigma} \tag{41}$$

Using (11) once more, we can substitute $H^{-1}\left(\frac{\delta(\theta)}{\beta(\theta)}\right) = \frac{s^* - BR}{\sigma} = \frac{s^* - \theta_1}{\sigma}$ into (41), which yields the result. $\square$

---

[48]Recall from the proof of Lemma 1 that R's best response is always a cut-off, regardless of whether some types mix over disclosure. From R's perspective, this is equivalent to the perturbartion we have chosen

Proof of Lemma 2

Proof. To establish the first part of the Lemma, note that the characterization from Appendix C allows us to write

$$s^\star(\theta_1, \sigma) = \frac{1}{1-\alpha} \cdot \left[\sigma_{s^*} \cdot y\left(\frac{\theta_1}{\sigma_s}\right) + c - \alpha \cdot \mu \right],$$

where $\sigma_s^2 = \frac{\sigma_\theta^2 \sigma^2}{\sigma_\theta^2 + \sigma^2}$ is the posterior variance of θ given s, $\alpha = \frac{\sigma^2}{\sigma^2 + \sigma_\theta^2}$ is the weight placed on the prior in Bayesian updating and $y(x)$ is defined as in (52). Differentiating, using (9) and substituting the equilibrium condition that $\frac{\theta_1 - s^\star(\theta_1, \sigma)}{\sigma} = \Phi^{-1}\left(\frac{\delta}{\beta}\right)$, we have

$$\frac{d\theta_1}{d\sigma} \overset{sign}{=} \frac{2\sigma}{\sigma^2 + \sigma_\theta^2} s^\star(\theta_1, \sigma) + \frac{1}{1-\alpha}\left[\left(y\left(\frac{\theta_1}{\sigma_s}\right) - \left(\frac{\theta_1}{\sigma_s}\right) y'\left(\frac{\theta_1}{\sigma_s}\right)\right) \frac{\sigma_\theta}{2\sqrt{\alpha}} - \mu\right] \frac{d\alpha}{d\sigma}$$
$$- \Phi^{-1}\left(\frac{\delta}{\beta}\right). \tag{42}$$

When $\mu < 0$ then we have $s^\star(\theta_1, \sigma) > 0$ so that the first term is positive. Letting $\theta_1 / \sigma_s = x$, the second term is guaranteed to be positive as long as $y(x) > xy'(x)$, or equivalently if the slope of a ray from the origin to the point $(x, y(x))$ is greater than $y'(x)$. This follows from Proposition 10, where we show that each ray from the origin crosses the function $y(x)$ once from above, which establishes that it must be steeper than $y(x)$ at the point of crossing. Finally, the third term is negative by assumption since $\Phi^{-1}(z) < 0$ for all $z < 1/2$.

For the second part, consider any equilibrium with cutoff θ_1, and a simultaneous change in μ and δ which ensures that θ_1 remains an equilibrium cutoff. It is more convenient to represent the change in δ by a change in $\Psi \equiv \Phi^{-1}\left(\frac{\delta}{\beta}\right)$. Equilibrium requires that $s^\star(\theta_1) - \theta_1 = \sigma\Psi$ and so we must have

$$\frac{d\Psi}{d\mu} = \frac{1}{\sigma}\frac{ds^\star(\theta_1, \sigma)}{d\mu}.$$

Note further that $\frac{ds^\star(\theta_1, \sigma)}{d\mu} = \frac{-\alpha}{1-\alpha}$. We now consider the right-hand side of (42), which changes in proportion to $d\mu$ by

$$\frac{d}{d\mu}\left\{\frac{2\sigma}{\sigma^2 + \sigma_\theta^2}[s^\star(\theta_1, \sigma) - \mu] - \Psi\right\} = \frac{2\sigma}{\sigma^2 + \sigma_\theta^2}\left[\frac{ds^\star(\theta_1, \sigma)}{d\mu} - 1\right] - \frac{1}{\sigma}\frac{ds^\star(\theta_1, \sigma)}{d\mu}.$$
$$= \frac{-2\sigma}{\sigma^2 + \sigma_\theta^2}\frac{1}{1-\alpha} + \frac{1}{\sigma}\frac{\alpha}{1-\alpha}$$
$$= \frac{-2\sigma}{\sigma_\theta^2} + \frac{1}{\sigma}\frac{\sigma^2}{\sigma_\theta^2} = -\frac{\sigma}{\sigma_\theta^2} < 0.$$

Thus the right-hand side of (42) changes linearly with $d\mu$ and is guaranteed to be negative whenever

$d\mu$ is large enough, which completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad$ □

B Analysis of stability with binary actions

Definition 2. An $\varepsilon-$perturbation of S's (pure) equilibrium disclosure strategy $m^*(\theta) = \theta \iff \theta \in [0, \theta^*]$ is a pure disclosure strategy of the following form:

$$m^\varepsilon(\theta) = \begin{cases} \emptyset & if\ \theta < 0 \\ \theta & if\ 0 \leq \theta \leq \theta^* + \varepsilon \\ \emptyset & if\ \theta > \theta^* + \varepsilon \end{cases}$$

if $\theta^* < \infty$, and

$$m^\varepsilon(\theta) = \begin{cases} \emptyset & if\ \theta < 0 \\ \theta & if\ 0 \leq \theta \leq \varepsilon^{-1} \\ \emptyset & if\ \theta > \varepsilon^{-1} \end{cases}$$

for some $\theta' \in \mathbb{R}_+$ if $\theta^* = \infty$.

Our definition of ε-perturbations restricts attention to small deviations from an equilibrium interval strategy to a 'nearby' interval disclosure strategy. However, since we have shown that S's best response to *any* (possibly mixed) disclosure strategy is to play an interval strategy, our definition will be without loss for considering best response dynamics. While our definition of ε-perturbations in the case of $\theta^* = \infty$ looks particular, we make it for expositional convenience. In particular, the definition is without loss for our purposes as nothing in our results relies on the 'speed' at which perturbations vary in ε.[49]

Definition 3. In the binary action model, we describe a cut-off equilibrium $(\theta^*, s^*(\theta^*))$ as *stable* if there exists $\xi > 0$ such that for all $\varepsilon-$perturbations of S's disclosure strategy, $|\varepsilon| < \xi$, are stable under best response dynamics. That is, the sequences of iterated best response functions $\{BR^n(\theta)\}_{n=1}^\infty$, $\{s^n\}_{n=1}^\infty$ satisfy

$$\lim_{n\to\infty} BR^n = \theta^* \tag{43}$$

$$\lim_{n\to\infty} s^n = s^*(\theta^*) \tag{44}$$

where $BR^1 = \theta^* + \varepsilon$, and $BR^n := BR(BR^{n-1})$, $s^n := s^*(BR^{n-1})$ for $n > 1$. If (43) does not hold,

---

[49]In other words, we could have just as easily defined a perturbation when θ^* as a bound B such that $m'(\theta) = \theta \iff \theta \in [0, B]$. Clearly, this is equivalent to Definition 1 with B=ε^{-1}.

then the equilibrium is unstable.

We now present a useful Lemma which simplifies the task of checking for instability of an equilibrium.

Lemma 5. *A cut-off equilibrium* (θ^*, s^*) *is (un)stable if* $\exists \xi' > 0$ *such that,* $\forall \theta' \in [\theta^* - \xi', \theta^* + \xi']$, $\theta' \neq \theta^*$

$$(BR(\theta') - \theta').(\theta' - \theta^*) < (>)0 \tag{45}$$

if $\theta^* < \infty$. *If* $\theta^* = \infty$, *then equilibrium is (un)stable if* $\exists B > 0$ *such that,* $\forall \theta' > B$

$$BR(\theta') - \theta' > (<)0$$

Proof. Suppose that θ^* is finite.[50] We establish that transparency is stable if

$$(BR(\theta') - \theta').(\theta' - \theta^*) < 0 \tag{46}$$

on $\theta' \in [\theta^* - \xi', \theta^* + \xi']$, $\theta' \neq \theta^*$. Thus, select some $\theta^{(1)} \in (\theta^* - \xi', \theta^*)$ and set $m^1(\theta) = \theta \iff \theta \in [0, \theta']$. Recall that R's best response to m^1 is to play $a = 1$ if and only if $s \geq s^*(\theta')$. By definition of $BR(\theta')$, S's best response when R plays cut-off strategy $s^*(\theta')$ is $BR^2 := BR(\theta^{(1)})$. Letting $\theta^{(2)} = BR(\theta^{(1)})$, we can continue iteratively to define sequences $\{\theta^{(n)}\}_{n=1}^{\infty}$, $\{s^*(\theta^{(n)})\}_{n=1}^{\infty}$.

We now argue by induction that the elements of these sequences satisfy: *(i)* $\theta^{(n)} \leq \theta^*$; *(ii)* $\theta^{(n)} \leq \theta^{(n+1)}$, $\forall n \in \mathbb{N}$; and *(iii)* $\lim_{n \to \infty} \theta^{(n)} = \theta^*$. First, note that $\theta^{(1)} \leq \theta^*$ by construction and that from condition (46), $\theta^{(2)} \geq \theta^{(1)}$. Now suppose that $\theta^{(n)} \leq \theta^*$ for all $n \leq N$. In particular, this is then true for $\theta^{(N)}$. But we have already established that $BR(\theta')$ is a strictly increasing function of θ'. So $\theta^{(N+1)} = BR(\theta^{(N)}) < BR(\theta^*) = \theta^*$, establishing *(i)*. An almost identical inductive argument establishes *(ii)*.

Since $\{\theta^{(n)}\}_{n=1}^{\infty}$ is a bounded, increasing sequence it must have some limit. To establish (iii), we suppose for a contradiction that *(iii)* $\lim_{n \to \infty} \theta^{(n)} = \theta'' < \theta^*$. But since $\theta'' < \theta^*$, it follows from (46) that $BR(\theta'') > \theta''$. But since $BR(\theta)$ is a continuous function, there exists an open ε-neighborhood around θ'' such that $|BR(\theta'') - BR(\theta)| \leq k := \frac{BR(\theta'') - \theta''}{2}$ for all $\theta \in (\theta'' - \varepsilon, \theta'' + \varepsilon)$. Further, since $\lim \{\theta^{(n)}\}_{n=1}^{\infty} = \theta''$, we can find N such that $|\theta^{(n)} - \theta''| \leq \varepsilon$, for all $n \geq N$. Taking any $n' \geq N$, we must have that

$$
\begin{aligned}
\theta^{(n'+1)} = BR(\theta^{(n')}) &\geq BR(\theta'') - \frac{BR(\theta'') - \theta''}{2} \\
&= \frac{BR(\theta'') + \theta''}{2} \\
&> \theta''
\end{aligned}
$$

---

[50] The infinite case is analogous to the steps we show below for $\theta' < \theta^*$ and therefore omitted).

- a contradiction to the assumption that θ'' is the limit of the increasing function $\{\theta^{(n)}\}_{n=1}^{\infty}$. This establishes *(iii)*.

Finally, notice that for any $\theta^{(1)} \in (\theta^* - \xi', \theta^*)$ we have constructed sequences of best responses $\{BR^n\}_{n=1}^{\infty}$, $\{s^*(\theta^{(n-1)})\}_{n=1}^{\infty}$ which all have the property that $\lim_{n\to\infty} BR^n = \theta^*$ and $\lim_{n\to\infty} s^n = s^*(\theta^*)$. This establishes the claim of the Lemma.

As the details of the proof are again very similar, we briefly outline the argument that transparency is unstable when $(BR(\theta') - \theta') \cdot (\theta' - \theta^*) \geq 0$. To fix ideas, select again some $\theta' \in [\theta^* - \xi', \theta^* + \xi]$. The proof proceeds by showing that any $\theta^{(1)} < \theta^*$ creates a profile of iterated best responses $\theta^{(n)} = BR(\theta^{(n)})$ which are decreasing on $\theta^{(n)} \in [\theta^* - \xi', \theta^*]$ and satisfy $BR(\theta^{(n)}) < \theta^* - \xi'$ for any $\theta^{(n)} < \theta^* - \xi'$. Similarly, for any $\theta^{(1)} > \theta^*$ analogous steps show that the iterated best responses $\theta^{(n)} = BR(\theta^{(n)})$ are increasing on $\theta^{(n)} \in [\theta^* - \xi', \theta^*]$ and satisfy $BR(\theta^{(n)}) > \theta^* - \xi'$ for any $\theta^{(n)} > \theta^* - \xi'$. $\qquad\square$

C Analysis of normal distributions

This appendix characterizes the potential multiplicity of equilibria in the financial crises model, assuming that types and signals are normally distributed.

Let $\theta \sim N\left(\mu, \sigma_\theta^2\right)$ and $s = \theta + \varepsilon$, where $\varepsilon \sim N\left(0, \sigma_\varepsilon^2\right)$ and θ, ε are independent. We are interested in characterizing the solutions $(\theta^*, s^*(\theta^*))$ to the system

$$\Phi\left(\frac{\theta^* - s^*(\theta^*)}{\sigma_\varepsilon}\right) = \frac{\delta}{\beta} \tag{47}$$

$$\mathbb{E}\left[\theta | s^*(\theta^*), \theta \notin [c, \theta^*]\right] = c \tag{48}$$

We approach the problem in two steps: First, we consider an arbitrary disclosure strategy $m(\theta) = \theta \iff \theta \in [0, \theta']$ and develop the properties of R's best response upon observing signal $m = \emptyset$. In other words, we characterize the solution $s^*(\theta')$ to the implicit function[51]

$$\mathbb{E}\left[\theta | s^*(\theta'), \theta \notin [c, \theta']\right] = c \tag{49}$$

and second, rewriting (47), we solve the fixed point problem $BR(\theta') = \theta'$ to

$$BR(\theta') = s^*(\theta') - \sigma_\varepsilon \cdot \Phi^{-1}\left(\frac{\delta}{\beta}\right) \tag{50}$$

Note that given parameters δ, β, the function $BR(\theta')$ is simply an affine transformation of $s^*(\theta')$ - a point which will be important in what follows.

---

[51] In general, the solution $s^*(\theta')$ will also depend on the parameters $\{\mu, \sigma, \sigma_\varepsilon, \delta, \beta\}$. We suppress that explicit dependence for notational ease here, but will reintroduce it when looking at comparative statics.

Thus, we focus first on (49). A well-known property of truncated normal distributions (which we present without proof) is that their conditional mean can be expressed as follows:

Lemma 6. *Let* $u \sim N\left(\mu', (\sigma')^2\right)$ *and suppose that* $u \notin [a,b]$. *Then*

$$\mathbb{E}\left[u | u \notin [c,d]\right] = \mu' - \frac{\phi\left(\frac{c-\mu'}{\sigma'}\right) - \phi\left(\frac{d-\mu'}{\sigma'}\right)}{1 - \Phi\left(\frac{d-\mu'}{\sigma'}\right) + \Phi\left(\frac{c-\mu'}{\sigma'}\right)} \cdot \sigma' \tag{51}$$

In our environment, $\theta - c | s \sim N\left(\alpha.\mu + (1-\alpha).s - c, \alpha.\sigma_\theta^2\right)$, where $\alpha = \frac{\sigma_\varepsilon^2}{\sigma^2 + \sigma_\varepsilon^2}$. Letting $\mu_s = \alpha.\mu + (1-\alpha).s - c$, $\sigma_s = \alpha = \frac{\sigma_\varepsilon^2}{\sigma^2 + \sigma_\varepsilon^2}$ denote the conditional mean and variance of θ on observing s, we can use Lemma 6 to write (49) as

$$\frac{\mu_{s^*}}{\sigma_{s^*}} = \frac{\phi\left(-\frac{\mu_{s^*}}{\sigma_{s^*}}\right) - \phi\left(\frac{\theta'-\mu_{s^*}}{\sigma_{s^*}}\right)}{1 - \Phi\left(\frac{\theta'-\mu_{s^*}}{\sigma_{s^*}}\right) + \Phi\left(-\frac{\mu_{s^*}}{\sigma_{s^*}}\right)}$$

or, defining $x = \frac{\theta'}{\sigma'}$, $y(x) = \frac{\mu_{s^*}}{\sigma_{s^*}}$ (recall that s^*, and therefore μ_{s^*} are functions of θ'):

$$y(x) = \frac{\phi\left(-y(x)\right) - \phi\left(x - y(x)\right)}{1 - \Phi\left(x - y(x)\right) + \Phi\left(-y(x)\right)} \tag{52}$$

Since $s^*(\theta_1) = \frac{1}{1-\alpha}.(\mu_{s^*} + c - \alpha.\mu) = \frac{1}{1-\alpha}.(\sigma_{s^*}.y(x) + c - \alpha.\mu)$, where $\frac{\sigma_{s^*}}{1-\alpha}$ is a constant in parameters, we develop the properties of the function $y(x)$ and apply them to $s^*(\theta')$. A first useful Lemma on the properties of $y(x)$ is the following:

Lemma 7. *For all* $x \in \mathbb{R}_+$, *(52) has a unique solution* $y(x)$, *an increasing, continuous function with:*

$$0 \leq y(x) \leq \frac{x}{2}$$

with strict inequality for all $x > 0$.

Proof. We first claim that for any $x > 0$, there is no solution $y \geq \frac{x}{2}$ or $y \leq 0$. If $y \leq 0$, then the LHS of (52) is trivially nonpositive while the RHS is strictly positive because $\phi(-y) > \phi(x-y)$ for all $x > 0$, $y \leq 0$ by the single-peakedness of $\phi(u)$, with peak at $u = 0$.[52] Thus, $y \leq 0$ cannot solve (52). Similarly, suppose $y \geq \frac{x}{2}$. Then, the LHS is trivially positive (since $x > 0$), while the RHS is nonpositive since $\phi(-y) \leq \phi(x-y)$ if $y \geq x - y$ by symmetry and single-peakedness of ϕ.

Multiplying (52) through by $1 - \Phi(x-y) + \Phi(-y)$ (strictly positive for $x > 0$, $\frac{x}{2} \geq y \geq 0$),

---

[52]The denominator, $1 - \Phi(x-y) + \Phi(-y)$, is positive, for all $x, y \in \mathbb{R}$.

imposing $\Phi(-y) = 1 - \Phi(y)$ and $\phi(-y) = \phi(y)$ and taking differences gives

$$g(y;x) = (1 - \Phi(x-y) + 1 - \Phi(y)) \cdot y - (\phi(y) - \phi(x-y))$$

By an identical argument to to one above, $g(0;x) < 0$, $g\left(\frac{x}{2};x\right) > 0$. Further,

$$
\begin{aligned}
\frac{\partial g}{\partial y} &= (1 - \Phi(x-y) + 1 - \Phi(y)) + y \cdot (\phi(x-y) - \phi(y)) - \phi'(y) - \phi'(x-y) \\
&= (1 - \Phi(x-y) + 1 - \Phi(y)) + x\phi(x-y) \\
&> 0
\end{aligned}
$$

where the second line uses the property of normal densities that $\phi'(u) = -u\phi(u)$. Since $g(y;x)$ is continuous and increasing, there must exist exactly one $y' \in \left[0, \frac{x}{2}\right]$ which solves

$$g(y';x) = 0$$

for each $x > 0$. Further, it is easy to show that g is continuous in both x and y and decreasing in x. Thus, the implicit function theorem implies that $y(x)$ is everywhere continuous and increasing in x. $\qquad\square$

Finally, we present two useful technical Lemmata that we will rely on heavily in the proof of Proposition 10.

Lemma 8. *Consider a C^1 function $\eta : A \to \mathbb{R}$ for some domain $A = [k,l]$, $k, l \in \mathbb{R} \cup \{-\infty, \infty\}$. Let the number of roots of η be denoted $R(\eta) := \#\{x : \eta(x) = 0\}$. Then*

$$R(n) \leq R(\eta') + 1$$

where $\eta' : A \to \mathbb{R}$ is the derivative of η, i.e. $\eta' = \frac{d\eta}{dx}$. Further, if $l = \infty$ and $\lim_{x \to \infty} \eta(x) = 0$, then

$$R(n) \leq R(\eta')$$

Proof. We focus on the case where η' has finitely many roots, since this will be what matters for our purposes. Suppose that η' has m distinct roots and denote these by $\left\{x_1^{\eta'}, x_2^{\eta'}, \ldots, x_m^{\eta'}\right\}$, where $x_1^{\eta'} < x_2^{\eta'} < \cdots < x_m^{\eta'}$. By the continuity of η', these roots divide A into $m+1$ intervals, $(k, x_1^{\eta'}]$, $(x_1^{\eta'}, x_2^{\eta'}], \ldots, (x_m^{\eta'}, l]$ on each of which η is a monotonic function. Since a monotonic function can have at most one root on its domain, η can have at most one root for each of the intervals above and therefore $R(\eta) \leq m+1 = R(\eta') + 1$.

Now suppose that $\lim_{x \to \infty} \eta(x) = 0$. We argue that there cannot be a root of η on the final

interval $\left(x_m^{\eta'},\infty\right)$. Suppose not - then $\exists x' \in \left(x_m^{\eta'},\infty\right)$ such that $\eta(x') = 0$. We consider three cases: First, suppose $\exists \varepsilon'$ s.t $\eta(x'-\varepsilon) < 0$ and $\eta(x'+\varepsilon) > 0$ for all $\varepsilon > 0$, then $\eta'(x') > 0$ and $\eta(x) > 0$ for some $x > x'$ (namely, any $x \in (x',x'+\varepsilon']$). Since there is no root of $\eta'(x)$ on $\left(x_m^{\eta'},\infty\right)$ by construction, $\eta'(x) > 0$ and

$$\lim_{x\to\infty} \eta(x) > \eta(x'+\varepsilon') > 0$$

contradicting $\lim_{x\to\infty} \eta(x) = 0$. An analogous argument establishes the claim for the case $\exists \varepsilon'$ s.t $\eta(x'-\varepsilon) > 0$ and $\eta(x'+\varepsilon) < 0$. Finally, if no such ε' exists to satisfy cases 1 or 2, then it must either be that there exists an ε' such that $\eta(x') \geq \eta(x)$, $\forall |x - x'| \leq \varepsilon'$ or $\eta(x') \leq \eta(x)$, $\forall |x - x'| \leq \varepsilon'$. In either case, $x' \in \left(x_m^{\eta'},\infty\right)$ must be a local extremum of η and therefore satisfy $\eta'(x') = 0$ - a contradiction to $x_m^{\eta'}$ being the largest root of η'. $\qquad\square$

Lemma 9. *Let $\tilde{y} : \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous, increasing function of x. There exists a unique $x^t \in \mathbb{R}_+$ such that $\tilde{y}(x)$ is convex (concave) on $[0,x^t)$ and concave (convex) on $[x^t,\infty]$ if and only if $\forall a, b \in \mathbb{R}$, the system*

$$
\begin{aligned}
y &= \tilde{y}(x) \\
y &= ax+b
\end{aligned}
$$

has at most 3 solutions and has exactly 3 solutions for some $a, b \in \mathbb{R}$.

Proof. Strategy - only if is easy. If done by contradiction - taking the convex hull of a non-standard region & arguing this is a straight line. Then shift the intercept a little and you generate 4 solutions. $\qquad\square$

We are now ready to state and prove our main result characterizing the implicit function, $y(x)$.

Proposition 10. *The implicit function $y(x)$ that solves (52):*

1. *is first convex and then concave: i.e. $\exists x^t \in \mathbb{R}_+$ such that $y(x)$ is convex on $[0,x^t)$ and concave on $[x^t,\infty)$;*

2. *satisfies*

$$\lim_{x\to\infty} y(x) - \frac{x}{2} \to 0$$

Proof. We prove 1. and 2. by counting the number of solutions to the nonlinear system of Equations

$$
\begin{aligned}
y &= y(x) && (53) \\
y &= ax+b && (54)
\end{aligned}
$$

for all choices of constants, $a, b \in \mathbb{R}$. Based on Lemma 9, Proposition 10.1 follows if (53) - (54) has no more than 3 solutions. Proposition 10.2 will follow below when we consider the case $a = \frac{1}{2}$, $b = -\varepsilon$, for $\varepsilon \to 0$.

Consider the following function

$$G(x;a,b) := 1 - \Phi((1-a)x - b) + 1 - \Phi(ax+b) - \frac{\phi(ax+b) - \phi((1-a)x-b)}{ax+b} \tag{55}$$

On inspection of (55), it should be clear that

$$G(x;a,b) = \frac{g(y;x)}{y} \Big|_{y=ax+b}$$

and consequently for any $y = ax + b > 0$, $G(x';a,b) = 0$ if and only if $(x', ax' + b)$ solves (53)-(54).[53] Therefore, to analyze the solutions to (53)-(54), it is sufficient to find the roots to

$$G(x;a,b) = 0 \tag{56}$$

Some immediate and useful properties of G are summed up in the following Lemma:

Lemma 10. $G(x;a,b)$ satisfies:

(i) If $a > 1$ ($a \leq 0$), $G(x;a,b)$ has exactly one solution on $x \geq 0$ if $b \leq 0$ ($b \geq 0$). Otherwise, it has none;

(ii) If $a \geq \frac{1}{2}$, $G(x;a,b)$ has no solutions when $b > 0$;

(iii) $\forall a \in (0,1)$, $\lim_{x \to \infty} G(x;a,b) = 0$;

(iv) $\forall b > 0$, $a \geq 0$, $G(x;a,b)$ is continuous on $x \geq 0$ with

$$G(0;a,b) = 1$$

(v) $\forall b \leq 0$, , $a \geq 0$, $G(x;a,b)$ has no roots on $x \in [0, -\frac{b}{a}]$, is continuous on $x > -\frac{b}{a}$ with

$$\lim_{x \downarrow -\frac{b}{a}} G(x;a,b) = -\infty$$

and

$$G(0;a,b) = 1$$

---

[53] We do not need to worry about cases where $y \leq 0$ as we have already seen that (i) $y < 0$ is never a solution to (53)-(54); (ii) $y = 0$ solves (53)-(54) if and only if $x = 0$. In other words, there is only an additional solution to (53)-(54) not captured by (56) when $b = 0$. We will account for this additional solution when we consider cases where $b = 0$.

(vi) The derivative of G with respect to x, G', is continuous on $x > -\frac{b}{a}$ and satisfies

$$G'(x;a,b) = \frac{a\phi\left((1-a)x-b\right)}{(ax+b)^2} \cdot \left[e^{\left(\frac{1}{2}-a\right)x^2 - bx} - \left((1-a)x^2 + \frac{1-a}{a}bx + 1 \right) \right]$$

Proof. (i) If $a < 0$, the claim follows immediately by noting that (54) is nonincreasing, while (53) is increasing with $y(0) = 0$. When $a > 1$, the claim is established by applying the IFT to implicitly differentiate $g(y(x);x) = 0$ w.r.t. x. After some algebra, this yields

$$y'(x) = \frac{x\phi(x-y)}{(1-\Phi(x-y)+1-\Phi(y))+x\phi(x-y)} \leq 1$$

(i) then follows immediately. *(ii)* follows from the bound $y(x) < \frac{x}{2}$, since $a \geq \frac{1}{2}$, $b > 0$ implies $ax + b \geq \frac{x}{2} > y(x)$, $\forall x \geq 0$. *(iii)* follows from $\lim_{u\to\infty} \Phi(u) = 1$, $\lim_{u\to\infty} \phi(u) = 0$. When $b \geq 0$, *(iv)* follows since G is clearly continuous as it is the sum of continuous functions (where the denominator of the second term is always positive for $x \geq 0$ when $b > 0$).

To establish *(v)*, note first that for $b \leq 0$, there is a single discontinuity in G at $ax + b = 0$, or $x = -\frac{b}{a}$ - which represents an asymptote of the function. Indeed, we can verify directly that

$$\lim_{x\downarrow-\frac{b}{a}} G(x;a,b) = 1 - \lim_{x\downarrow-\frac{b}{a}} \frac{\phi(ax+b) - \phi((1-a)x-b)}{ax+b} = -\infty$$

where the final step holds because the numerator of the second term converges to $\phi(0) - \phi\left(\frac{-b}{a}\right)$ > 0 for $a \geq 0$ while the denominator converges to 0 from above.[54]

However, for $x \in \left[0, -\frac{b}{a}\right]$, $a \geq 0$, we have $y(x) > 0$ (by Lemma 7) whenever $ax + b < 0$. Therefore, (53)-(54) cannot have a solution on this range so long as $a > 0$.

Finally, differentiating (55) with respect to x yields

$$
\begin{aligned}
G'(x;a,b) &= -(1-a)\phi\left((1-a)x-b\right) - a\phi(ax+b) - \frac{1}{(ax+b)^2}. \\
&\quad \left((ax+b) . \left[a\phi'(ax+b) - (1-a) . \phi'((1-a)x-b) \right] - \right. \\
&\quad \left. a . \left[\phi(ax+b) - \phi((1-a)x-b) \right] \right) \\
&= \frac{1}{(ax+b)^2} . \left[a\phi(ax+b) - \left((1-a)ax^2 + (1-a)bx + a \right) . \phi((1-a)x+b) \right] \\
&= \frac{a\phi\left((1-a)x-b\right)}{(ax+b)^2} . \left[\frac{\phi(ax+b)}{\phi((1-a)x-b)} - \left((1-a)ax^2 + (1-a)bx + a \right) \right] \\
&= \frac{a\phi\left((1-a)x-b\right)}{(ax+b)^2} . \left[e^{\left(\frac{1}{2}-a\right)x^2 - bx} - \left((1-a)x^2 + \frac{1-a}{a}bx + 1 \right) \right] \quad (57)
\end{aligned}
$$

---

[54]Similarly, it is easy to show that $\lim_{x\uparrow-\frac{b}{a}} G(x;a,b) = \infty$.

where line 2 follows after using the property of normal densities that $\phi'(u) = -u\phi(u)$, $\forall u \in \mathbb{R}$, and some tedious algebra and line 4 uses the property of normal densities that $\frac{\phi(u)}{\phi(v)} = e^{\frac{1}{2}(v+u).(v-u)}$. Since this function is composed of sums and products of continuous functions for all $x \neq -\frac{b}{a}$, it is continuous on $x > -\frac{b}{a}$. As with G, G' has an asymptote at $x = -\frac{b}{a}$. However, since all solutions to $G(x;a,b) = 0$ are on $x > -\frac{b}{a}$, this is of no concern. $\quad\square$

Lemma 10 establishes that when $a < 0$ or $a > 1$, then system (53)-(54) has at most one solution. Further, when $b > 0$, the system has no solutions if $a \geq \frac{1}{2}$. Conditions (iii) - (vi) of the Lemma will be helpful in establishing the result in the remaining cases:

1. $0 < a \leq \frac{1}{2}$, $b \geq 0$;

2. $0 < a \leq \frac{1}{2}$, $b < 0$; and

3. $1 > a > \frac{1}{2}$, $b \leq 0$.

In particular, across these remaining two cases, $a \in (0,1)$. Since condition *(iii)* of Lemma 10 shows that $\lim_{x\to\infty} G(x;a,b) = 0$ conditions *(v)-(vi)* that G is a C^1 function on $x > \max\left\{0, -\frac{b}{a}\right\}$. Then, we can apply Lemma 8 to argue that

$$R(G) \leq R(G')$$

for $x > \max\left\{0, -\frac{b}{a}\right\}$.

Furthermore, since

$$\frac{a\phi((1-a)x - b)}{(ax+b)^2} > 0$$

for all $a \in (0,1)$, $x > \max\left\{0, -\frac{b}{a}\right\}$, we can see from Equation (57) that $G'(x;a,b) = 0$ if and only if

$$\gamma(x;a,b) := e^{\left(\frac{1}{2}-a\right)x^2 - bx} - \left((1-a)x^2 + \frac{1-a}{a}bx + 1\right) = 0 \tag{58}$$

Equation (58) is the difference of an exponential with a quadratic power and a quadratic. To characterize its roots in cases 1-3, we find it convenient to take derivatives of γ and apply Lemma 8 repeatedly in each case. Calculating the derivatives directly, we have:

$$\gamma'(x;a,b) = ((1-2a)x - b)e^{\left(\frac{1}{2}-a\right)x^2 - bx} - 2(1-a)x - \frac{1-a}{a}b \tag{59}$$

$$\gamma''(x;a,b) = \left[((1-2a)x - b)^2 + (1-2a)\right].e^{\left(\frac{1}{2}-a\right)x^2 - bx} - 2(1-a) \tag{60}$$

$$\gamma'''(x;a,b) = ((1-2a)x - b).\left[3(1-2a) + ((1-2a)x - b)^2\right].e^{\left(\frac{1}{2}-a\right)x^2 - bx} \tag{61}$$

We now proceed to prove the result for the three remaining cases, 1-3:

Case 1: $0 < a \leq \frac{1}{2}$, $b \geq 0$ - $R(G) = 1$

First, note that $\gamma'(x;a,b)$ cannot have a root on $x \in \left[0, \frac{b}{1-2a}\right]$ since $b > 0$ and $a \leq \frac{1}{2}$ implies that the first term of (59) is negative (clearly, so too is the second term, $2(1-a)x - \frac{1-a}{a}b$, $\forall x$).

Consider the roots of $\gamma''(x;a,b)$ on $x \in [\frac{b}{1-2a}, \infty)$. Note that, for $x \geq \frac{b}{1-2a}$, the first term in (60)

$$\left[((1-2a)x - b)^2 + (1-2a)\right].e^{\left(\frac{1}{2}-a\right)x^2 - bx}$$

is the product of two increasing functions and therefore increasing itself.[55] Evaluating $\gamma''\left(\frac{b}{1-2a}; a, b\right)$ $= 1 - 2a - 2(1-a) = -1 < 0$ and $\lim_{x\to\infty} \gamma''(x;a,b) = \infty$, it is clear that $\gamma''(x;a,b)$ must have exactly 1 root on $[\frac{b}{1-2a}, \infty)$.

Applying Lemma 8, $\gamma'(x;a,b)$ can have no more than 2 roots on $[\frac{b}{1-2a}, \infty)$. However, since

$$\gamma'\left(\frac{b}{1-2a}; a, b\right) = -b(1-a)\left(\frac{3-2a}{a(1-2a)}\right) < 0$$

and since the exponential function always dominates any finite polynomial,

$$\lim_{x\to\infty} \gamma'\left(\frac{b}{1-2a}; a, b\right) = \infty$$

then γ' has an odd number of sign changes and therefore has exactly one root on $[\frac{b}{1-2a}, \infty)$. Further, since γ' has no roots on $\left[0, \frac{b}{1-2a}\right]$, $R(\gamma') = 1$.

Re-applying Lemma 8, γ can have at most 2 roots on $\mathbb{R}_+$. By a similar argument for γ', it is easy to show that[56]

$$\lim_{x\downarrow 0} \gamma(x;a,b) = 0_- \tag{62}$$

since $\gamma(0;a,b) = 0$ and $\gamma'(0;a,b)$. Further,

$$\lim_{x\to\infty} \gamma\left(\frac{b}{1-2a}; a, b\right) = \infty$$

Therefore, γ too must have an odd number of sign changes so that $R(\gamma) = 1$.

Finally, $R(\gamma) = R(G')$ by definition, and by Lemma 8 $R(G) \leq R(G')$. However, since $G(0;a,b) = 0$, $G'(x;a,b) < 0$ for $x \approx 0$ (recall (62)) and $\lim_{x\to\infty} G(x;a,b) = 0$ the unique turning point of G, some $x' \in \mathbb{R}$, must satisfy $G(x';a,b) < 0$. Otherwise, we would have to have $\lim_{x\to\infty} G(x;a,b) >$

---

[55]Notice that the quadratic functions $(1-2a)x - b^2 + (1-2a)$ and $\left(\frac{1}{2}-a\right)x^2 - bx$ both achieve their minima at $x = \frac{b}{1-2a}$.
[56]we use the subscript '$-$' to denote convergence from below

$G(x';a,b) \geq 0$ - a contradiction to $\lim_{x \to \infty} G(x;a,b) = 0$.

Therefore, there always exists exactly one root to $G(x;a,b) = 0$ for $0 < a < \frac{1}{2}, b > 0$

Case 2: $0 < a \leq \frac{1}{2}$, $b < 0$ - $R(G) \leq 2$

Since both $((1-2a)x-b)^2$ and $(\frac{1}{2} - a)x^2 - bx$ achieve their minima at $x = \frac{b}{1-2a} < 0$, inspection of (60) reveals that $\gamma''(x;a,b)$ is an increasing function of x for all $x \geq 0$. Therefore, it can have at most one root on $x \geq 0$. Evaluating

$$\gamma''(0;a,b) = b^2 + (1-2a) - 2(1-a)$$

As in case 1, it can be verified that $\lim_{x \to \infty} \gamma''(x;a,b) = \infty > 0$. Therefore, we must have[57]

$$R(\gamma'') = \begin{cases} 1 & , if \, b < -1 \\ 0 & if \, 0 > b \geq -1 \end{cases}$$

Applying Lemma 8 , we can write

$$R(\gamma') \leq \begin{cases} 2 & , if \, b < -1 \\ 1 & if \, 0 > b \geq -1 \end{cases}$$

Furthermore, direct calculation verifies that $\gamma'(0;a,b) = -\frac{b}{a} > 0$ and $\lim_{x \to \infty} \gamma'(x;a,b) = \infty$. Therefore, $\gamma'(x;a,b)$ must generically have an even number of roots[58] so that

$$R(\gamma') = \begin{cases} 2 & , if \, b < -1 \\ 0 & if \, 0 > b \geq -1 \end{cases}$$

generically. Now, iterating the argument of Lemma 8 once more to γ, we have

$$R(\gamma) \leq \begin{cases} 3 & , if \, b < -1 \\ 1 & if \, 0 > b \geq -1 \end{cases}$$

However, it can easily be verified that $\lim_{x \downarrow 0} \gamma(x;a,b) = 0_+$, where the '$+$' subscript denotes convergence from above,[59] while $\lim_{x \to \infty} \gamma(x;a,b) = \infty$. Thus, again, $\gamma(x;a,b)$ must have an even

---

[57]The argument for the case $b = -1$ follows by showing that $\gamma'''(0;a,b) > 0$. Thus, when $\gamma''(0;a,b) = 0$, there exists ε such that $\gamma''(x;a,b) > 0$, for all $0 < x \leq \varepsilon$.

[58]Genericity follows because γ' is strictly decreasing in b. If there were ever a single root to the Equation for some (a,b), then a marginal deviation to any b' close enough to b would create a function with 0 or 2 roots.

[59]This follows from direct calculation of $\gamma(0;a,b) = 0$, $\gamma'(0;a,b) > 0$

number of roots generically so that

$$R(\gamma) = \begin{cases} 0 \, or \, 2 & , if \, b < -1 \\ 0 & if \, 0 > b \geq -1 \end{cases} \tag{63}$$

Moreover, γ can never have 3 roots, even on sets of measure 0.[60] Now from (57)-(58), $R(\gamma) = R(G')$. Thus, by Lemma 8 applied to G, we have $R(G) \leq R(G')$.

We show that G actually achieves its upper bound of 2 roots for some choices of a, b. As a Corollary, we will also prove Proposition 10.2. First note that γ never takes its roots on $0 < x < -\frac{b}{a}$: noting that *(i)* $\gamma(0;a,b) = 0$; *(ii)* $\left(\frac{1}{2} - a\right)x^2 - bx$ is strictly increasing on $x \in \left[0, -\frac{b}{a}\right]$; and *(iii)* $(1-a)x^2 + \frac{1-a}{a}bx + 1 \leq 1 \equiv (1-a)x^2 + \frac{1-a}{a}bx + 1|_{x=0}$ for all $x \in \left[0, -\frac{b}{a}\right]$, (58) immediately implies that $\gamma(x;a,b) > 0$ on $0 < x \leq -\frac{b}{a}$. Then given (57)-(58), G' must also take its roots on $x > -\frac{b}{a}$.

Suppose that $a = \frac{1}{2}$. From Lemma 7, when $b = 0$ system (53)-(54) has a solution at $(x = 0, y = 0)$; and no additional solution on $x > 0$. By continuity of (54) in b, there must exist a $0 > \bar{b} \geq -1$ and an $\tilde{x} > 0$ such that $\forall 0 > b' \geq \bar{b}$:

$$y(0) = 0 > b'$$

and

$$y(\tilde{x}) \quad < \quad a\tilde{x} + b'$$

By continuity of (53) and (54) in x, there must therefore exist some $\tilde{x}_R \in (0, \tilde{x})$ such that $y(\tilde{x}_R) = a\tilde{x}_R + b'$ - i.e.

$$R(G) \geq 1$$

By Lemma 8, $R(G) \leq R(G') = R(\gamma)$. Moreover, from (63), $R(\gamma) \in \{0, 2\}$. Thus, $R(G) \geq 1 \implies R(G') = 2$. Furthermore, the argument of the preceding paragraph implies both roots of G' must occur on $x \geq -\frac{b'}{a}$. Under these conditions, we argue that G must have exactly two roots on $x \geq -\frac{b'}{a}$.

Suppose for a contradiction that G only had one root, $\tilde{x}_R$, on $x > -\frac{b'}{a}$. Since $\lim_{x\downarrow-\frac{b'}{a}} G(x;a,b') = -\infty$, $\tilde{x}_R$ must occur either before the first turning point of G or after the second turning point. Suppose we are in the former case. Then the second turning point of G, which we denote $\tilde{x}''$, must

---

[60]Otherwise, the one of the three roots of γ, say x', would also have to also be a root of γ' (to preserve $sign(\gamma(0;a,b)) = sign\left(\lim_{x\to\infty} \gamma'(x;a,b)\right) > 0$). Suppose that x' is the smallest root of γ(essentially the same argument goes through for the other roots). Then, since x' is a local extremum, there must exist some $x'' > x'$ such that γ does not change sign on $[0, x'']$ and $\gamma'(x;a,b) > 0$ for $x \in (x', x'')$. But since γ has two more roots, there must exist two more turning points of γ. This implies that γ' has 3 roots - a contradiction.

satisfy

$$G\left(\tilde{x}'';a,b'\right) > 0$$

since by assumption, $\tilde{x}_R$ is the unique root of G and

$$G'\left(x;a,b'\right) \;>\; 0$$

for $x > \tilde{x}''$ (since $\lim_{x\downarrow -\frac{b}{a}} G'\left(x;a,b'\right) = \infty > 0$ and $sign\left(G'\right)$ must not change after the function has been through two strict roots). However, this implies $\lim_{x\to\infty} G\left(x;a,b'\right) > G\left(\tilde{x}'';a,b'\right) > 0$ - a contradiction to $\lim_{x\to\infty} G\left(x;a,b'\right) = 0$. The latter case follows a similar argument and is therefore omitted.

We are now in a position to prove Proposition 10.2. We have established above that for all $b' \in (0,\bar{b}]$, the system

$$y \;=\; y(x) \tag{64}$$
$$y \;=\; \frac{1}{2}x + b' \tag{65}$$

has exactly 2 roots. Moreover, since $y(0) = 0 > b'$, (65) must intersect (64) at below for the first root. Similarly, (65) must therefore intersect (64) from above at its second root. Denoting this second root of the system as $x^*\left(b'\right) \in \mathbb{R}_+$, it must therefore be that

$$\frac{x}{2} > y(x) > \frac{x}{2} + b'$$

for all $x > x^*\left(b'\right)$. Taking $b' \uparrow \infty$ establishes that:

$$\lim_{x\to\infty} y(x) - \frac{x}{2} = 0$$

Finally, a similar strategy can be used to prove that $\exists \hat{b} \in (0,-1)$ such that for all $b'' \in [0,\hat{b}]$, system (53)-(54) has exactly two solutions. The proof again applies a continuity argument, starting with the finding of case *1* that there are 2 solutions to (53)-(54) when $a < \frac{1}{2}$, $b = 0$ and using $R(G) \leq 2$ to rule out the Introduction of additional equilibria as we move to some b'' below 0.

Case 3: $1 > a > \frac{1}{2}, b \leq 0$ - $R(G) \leq 3$

We first note that for $b \leq 0$, $a > \frac{1}{2}$, $\kappa(x) \equiv \left(\frac{1}{2} - a\right)x^2 - bx$ is a negative-definite quadratic which attains its maximum at $x_\kappa^* = -\frac{b}{2a-1}$, where $x_\kappa^* \geq 0$. By contrast, $\chi(x) \equiv (1-a)x^2 + \frac{1-a}{a}bx + 1$ is

positive definite and achieves its minimum at $x_\chi^* = -\frac{b}{2a}$, where $x_\kappa^* > x_\chi^* \geq 0$.

In particular, $x_\kappa^* > x_\chi^*$ implies that $\gamma(x;a,b) \equiv e^{\kappa(x)} - \chi(x)$ is a strictly decreasing function for all $x \geq x_\kappa^*$ since it is the difference of a decreasing function and an increasing function of x on this range. In other words, we must have $\gamma'(x;a,b) < 0$ for all $x \geq x_\chi^*$ so that

$$R\left(\gamma' \mid x \in [x_\kappa^*, \infty)\right) = 0 \qquad (66)$$

where we use the notation $R(\eta | x \in [a,b])$ to denote the roots on η on the restricted domain $x \in [a,b]$.

Now, consider the roots of $\gamma'''(x;a,b)$. Recalling from (61) that

$$\gamma'''(x;a,b) = ((1-2a)x - b) \cdot \left[3(1-2a) + ((1-2a)x - b)^2\right] \cdot e^{\left(\frac{1}{2}-a\right)x^2 - bx}$$

γ''' has 3 roots: (1) When $x = x_\kappa^*$, the first term is $(1-2a)x_\kappa^* - b = 0$; (2) the second term is a quadratic of the form $(z(x))^2 + k$ where $z(x) = (1-2a)x - b$, so that $z(x) = 0$ at $x = x_\kappa^*$, and $k = 3(1-2a) < 0$ for $a > \frac{1}{2}$. Thus, it must have two roots, x_{R_1}, x_{R_3} which satisfy $x_{R_1} < x_\kappa^*$ and $x_{R_3} > x_\kappa^*$.

Since $x_\kappa^* > -\frac{b}{a}$ for $1 > a$, we can split the interval $(-\frac{b}{a}, \infty)$ into the union of two intervals, $(-\frac{b}{a}, x_\kappa^*)$ and $[x_\kappa^*, \infty)$ and apply the result of Lemma 8 iteratively to each interval separately and also to their union. We collect the results of this exercise in Table 1, additionally applying our finding that $R(\gamma' \mid x \in [x_\kappa^*, \infty)) = 0$ when $1 > a > \frac{1}{2}$, $b \leq 0$. Each row of Table 1 represents upper bounds on the roots of the function listed in that row on different domains, where each column represents the interval domain listed in that column.

Lemma 8 implies that as we move down any given column in the table, entries in successive rows cannot increase by more than 1. Further, since $(-\frac{b}{a}, \infty) = (-\frac{b}{a}, x_\kappa^*) \cup [x_\kappa^*, \infty)$, the least upper bound on the roots of γ^n on $(-\frac{b}{a}, \infty)$ must be no greater the sum of the upper bounds we find for γ^n's roots on each of $(-\frac{b}{a}, x_\kappa^*)$, $[x_\kappa^*, \infty)$. In other words, in any given row of Table 1, entries in the third column must be no greater than the sum of corresponding entries in the first two columns.

There are two key entries of Table 1. First, the bound on $R(\gamma' | x \in [x_\kappa^*, \infty))$ is established by reference to (66). Second, the bound on $R\left(\gamma' | x \in (-\frac{b}{a}, \infty)\right)$ is determined by the horizontal sum of the bounds on $R(\gamma' | x \in [x_\kappa^*, \infty))$ and $R\left(\gamma' | x \in (-\frac{b}{a}, x_\kappa^*)\right)$. All other entries are determined using Lemma 8 - i.e. by adding 1 to the entry immediately above.

As Table 1 shows, this exercise shows that

$$R(\gamma(x;a,b)) := R\left(\gamma(x;a,b) | x \in (-\frac{b}{a}, \infty)\right) \leq 4$$

$R(\eta\,|\,x\in A) \leq$			
$\eta=/A=$	$\left(-\frac{b}{a}, x_\kappa^*\right)$	$[x_\kappa^*, \infty)$	$\left(-\frac{b}{a}, \infty\right)$
γ'''	1	2	3
γ''	2	3	4
γ'	3	0	3
γ	4	1	4

Table 1: Upper bound on roots of γ and its derivatives on sub-intervals of $\left(-\frac{b}{a}, \infty\right)$

Again, Lemma 8 immediately implies that

$$R(G) \leq R\left(G'\right) = R(\gamma)$$

Given this, we can now strengthen the argument to show that

$$R(G) \leq 3$$

for $1 > a > \frac{1}{2}$, $b \leq 0$. Recall from Lemma 10.(v) that

$$\lim_{x\downarrow -\frac{b}{a}} G(x;a,b) = -\infty < 0 \tag{67}$$

Further, since $a > \frac{1}{2}$, $b < 0$, there must exist some $x_t > -\frac{b}{a} > 0$ such that $ax+b > \frac{x}{2}$ for all $x > x_t$, or $ax+b > (1-a)x-b$. By the properties of the standard normal density, we must therefore have

$$\phi\left((1-a)x-b\right) > \phi\left(ax+b\right) \tag{68}$$

for all $x > x_t$. On inspection of (55), Equation (68) implies that for all $x > x_t > -\frac{b}{a}$

$$G(x;a,b) > 0 \tag{69}$$

Therefore, (67) - (69) imply that G must generically have an odd number of roots. Moreover, a very similar argument to the one made in case 2 shows that G can never have 4 roots, even on sets of measure 0.

We have established: (1) $R(G(x;a,b)) \leq 3$ for all $x \geq 0$, a, $b \in \mathbb{R}$; (2) $\lim_{x\to\infty} y(x) - \frac{x}{2} = 0$; (3) $y'(0) = 0$. Since $y(x)$ is strictly increasing, (3) implies that there is an interval $[0, x_1]$ on which $y(x)$ is convex. However, since $y(x) < \frac{x}{2}$, (2) implies there must exist some x_2 such that $y'(x_2) > \frac{1}{2}$ and an $x_3 > x_2$ such that $y'(x) < \frac{1}{2}$ for all $x \geq x_3$. In other words, $y(x)$ cannot be a convex function on the whole of $\mathbb{R}_+$ - which establishes that $R(G(x;a,b)) = 3$ for some choice of a, b (since all convex

functions satisfy $R(\eta;a,b) \leq 2, \forall a, b \in \mathbb{R}$). Appealing to Lemma 9 finishes the proof. $\qquad\square$