Inside and outside information: Fragility and Stress Test Design

Daniel Quigley and Ansgar Walther[1]

This draft: December 2017

[1] Quigley: Nuffield College, University of Oxford, Oxford OX1 1NF, UK. Email: daniel.quigley@economics.ox.ac.uk. Walther: Warwick Business School, Coventry CV4 7AL, UK. Email: ansgar.walther@wbs.ac.uk. We thank Heski Bar-Isaac, James Best, Vince Crawford, Martin Cripps, Peter Eso, Nicolas Inostroza, Ian Jewitt, Emir Kamenica, Navin Kartik, Alex Kohlhas, Paul Klemperer, Tim McQuade, Meg Meyer, Eduardo Perez-Richet, Larry Samuelson, Adrien Vigier, Selma Walther, Lucy White, Mungo Wilson, Peyton Young, and audiences at various seminars and conferences, for their comments. A previous version of this paper was circulated under the title “Crowding out disclosure”.

Abstract

We analyze a model of strategic communication where uninformed parties observe verifiable *inside information*, which is disclosed strategically by self-interested parties, as well as *outside information* that is beyond insiders' control. A central application is the design of financial stress tests (outside information) that are disclosed by financial regulators, and which interact with banks' endogenous choice of transparency (inside information). For a range of parameters, the classic "unraveling" spiral works in reverse, and information becomes fragile: Second-order changes in the distribution of outside information can trigger first-order reductions in inside disclosures. We show that optimal stress tests must satisfy a minimum standard of transparency. We further show that the importance of outside information hinges on the shape of insiders' payoff functions, which leads to new testable predictions for corporate disclosures.

1 Introduction

An enduring question in economic policy is whether governments should release more information to the public. For example, the financial crisis of 2008 triggered calls for more public transparency in banks, and public disclosures about banks' health have become a core tool of financial regulation.[2] The effect of public information on welfare is complex, especially when public signals influence how agents respond to additional private information about economic conditions. An active literature evaluates the trade-offs involved, focusing mostly on models where additional private information is dispersed among many agents.[3]

Another important case has received less attention: Private information is frequently concentrated in the hands of strategic *insider*s, who can decide whether or not to disclose verifiable evidence of what they know to other agents. In this paper, we argue that the incentive to disclose *inside information* depends critically on the availability and quality of public *outside information*. One of our main results is that outside information leads to fragility: Small changes in the distribution of outside signals can trigger large declines in inside disclosures. Because of this strong informational externality, the positive and normative consequences of better public (outside) information are markedly different in markets where inside information responds endogenously.

This is particularly salient in the context of financial crises, when financial stability seems to hinge on information about the health of a few systemically important banks, who can in principle decide how much inside information they wish to reveal. Recent data suggests that banks increased the rate and quality of inside disclosures following the 2008 crisis, but that the Fed's stress testing regime reduced the transparency of banks' financial accounting.[4] The growing academic literature on stress testing (e.g. Bouvard et al., 2015; Faria-e-Castro et al., 2016; Orlov et al., 2017; Goldstein and Leitner, 2017; Inostroza and Pavan, 2017) focuses mainly on the case where regulatory (outside) disclosures are the only credible source of information in a crisis. Applying our theory to a model of financial crises, we examine whether optimal stress tests are robust to the Lucas critique, in the sense that they remain optimal when banks' (inside) disclosures respond endogenously. A new insight that emerges

[2] See Goldstein and Sapra (2014) for a summary of arguments for and against public disclosures about banks, and Federal Reserve (2017) for an overview of the new regulatory framework.

[3] To name only a few key papers: Vives (1997) and Amador and Weill (2010) study the ambiguous relationship between public signals and the information content of prices; Hellwig (2002), Morris and Shin (2002), Angeletos and Pavan (2007) and Inostroza and Pavan (2017) evaluate the impact of public information on equilibria in coordination games.

[4] Bank of England (2013) shows that the quantity and quality of disclosures by international banks increased sharply in 2008, particularly regarding the valuation of their assets. Shahhosseini (2016) shows that stress-tested banks made fewer loan charge-offs and more frequently changed the classification of loan losses.

is that optimal financial stress tests must satisfy a minimum standard of transparency.

In Section 2, we set up a standard Sender-Receiver model of communication with binary responses. A short example illustrates our main points. Following our leading application, consider a large bank (Sender) whose assets are worth a random amount θ. Investors (Receiver) collectively make a binary choice: They roll over their loans to banks with expected asset value above a threshold p, and they run on banks worth less than p.[5] The bank would like to convince investors to roll over. Communication works as follows: First, the bank privately observes θ, and decides whether to verifiably disclose it. Disclosure comes at a cost,[6] but this cost is smaller than the benefit to the bank of avoiding a bank run. Second, a noisy outside signal s of θ, which we can think of as a stress test result, is observed by investors. Third, investors decide whether to run on the bank.[7]

Suppose that we are in a financial crisis: The expected value of assets is $E[\theta] < p$, and without any further information, there would be a bank run. If there is no outside information, then inside information 'unravels', as predicted by the classic literature (Grossman and Hart, 1980; Grossman, 1981; Milgrom, 1981): Banks with $\theta \geq p$ choose to avoid a run by disclosing θ. Banks with $\theta < p$ stay quiet, but face a run because investors understand that no news is bad news. As a result, equilibrium outcomes are as if investors had perfect information.

If public information is sufficiently precise, then incentives change dramatically. Impressive outside signals $s \geq s^\star$, for an endogenous threshold $s^\star$, now reveal high quality and entice investors to roll over, even when the bank does not disclose anything. Consequently, the very best banks prefer to stay quiet: They are confident that they will obtain an impressive stress test result $s \geq s^\star$ with high probability, so that the marginal benefit of disclosure is small. Moreover, this reaction sets off a feedback loop. When the best banks stay quiet, no news is ambiguous news, so that silence is interpreted more favorably by investors. Then, yet more high-quality banks prefer to stay quiet, silence becomes even better news, and so forth.[8]

We establish that this feedback loop, which we dub *reverse unraveling*, generates fragility

[5] We micro-found this setup in a classical coordination game between depositors as in Diamond and Dybvig (1983), using a global games refinement (Morris and Shin, 2000). Here, the run threshold p captures the illiquidity of bank assets. Our main insights do not concern the coordination game among investors; they are the same in the case where a single investor has incentives to withdraw when she is pessimistic.

[6] Leuz and Wysocki (2016) survey a large body of research documenting that disclosures are costly, both for technological reasons and because of concerns about releasing proprietary data to competitors. Our main results continue to hold in the case where disclosure costs are small, in the sense that they would not matter in a model without outside information.

[7] We focus on disclosures that are pre-emptive: When Sender decides whether to disclose θ, he cannot perfectly predict the realization of s. We discuss the foundations of this assumption in Section 2.

[8] The first step in this mechanism, i.e. the tendency of the best types to stay quiet, is reminiscent of the 'too cool for school' effect in the literature on signaling games (Feltovich et al., 2002; Daley and Green, 2014), which we discuss in detail below.

of information. A small improvement in the quality of outside signals can lead to a discontinuous decline in the amount of inside information that is revealed. This fragility is not a special case: In the model with binary responses, we show that discontinuities must arise along *any* continuous path of gradually improving outside signals under natural regularity conditions. Moreover, while the above intuition implicitly assumes that outside signals do not have full support (there is an $s^\star$ which reveals high quality beyond doubt), we show that the fragility of information extends in a natural sense to the case of full support.

Interestingly, both unraveling and reverse unraveling are driven by the same deeper feature, namely, that verifiable disclosures exhibit a form of strategic complementarity. If the best types of Sender are expected to disclose their type, Receiver rationally assumes that no news is bad news, which generates strong incentives for other types of Sender to also disclose. In models without outside information, the best are keen to disclose, and the complementarity leads to unraveling. In our model, by contrast, outside information weakens incentives to disclose among the best types, and the strategic complementarity can work in favor of opacity, which drives reverse unraveling.

Away from points of discontinuity, the relationship between inside and outside information is more nuanced. Better outside information tends to crowd out inside disclosures at the margin if Receiver's prior beliefs about θ are pessimistic, but crowd in disclosures if they are optimistic.

In Section 3, we consider the normative implications of our model. There is an informational externality: Better outside signals affect inside information in equilibrium. This externality is powerful due to the fragility of information, and optimal policy must take it into account.

The objective function for optimal informational policy is context-specific. In some situations, for example in the market for used cars studied by Akerlof (1970), the first-best informational outcome is full transparency, and policy-makers should avoid crowding out inside disclosures. Then, there is a case for releasing very little outside information in order to encourage an unraveling outcome. In other situations, for example in insurance markets considered by Hirshleifer (1971), the first-best outcome involves some residual uncertainty, because uncertainty facilitates efficient risk-sharing. In this case, welfare can often be improved by releasing *more* outside information than would be optimal in the absence of externalities, in order to encourage reverse unraveling.

We show that our application to financial panics falls into the latter case. Intuitively, opacity by strong banks enhances welfare by generating implicit insurance for weak banks, who would otherwise face a bank run. An optimal policy in this context must exploit the informational externality by crowding out disclosures from the strongest banks. This is implemented by releasing outside signals, such as stress tests, that meet a *minimum standard of transparency*. This result is relevant in the context of a growing literature on stress test design and information disclosure during financial crises (Bouvard et al., 2015; Faria-e-Castro et al., 2016; Orlov et al., 2017; Goldstein and Leitner, 2017). Existing work focuses on a benchmark case where banks cannot make inside disclosures. A common result is that the optimal design of stress tests must be fine-tuned, and highly sensitive to investors' prior beliefs about bank quality. We complement this assessment by showing that, when inside information responds endogenously, a minimum degree of transparency should be part of any optimal policy as soon as investors' prior beliefs deteriorate beyond a certain threshold (in particular, as soon as $E[\theta] < p$ in the above example).

In Section 4, we consider a more general Sender-Receiver model where responses need not be binary. A new insight that emerges is that the impact of outside information depends not only on its quality, but also on the shape of Sender's payoffs. If Sender's payoffs are sufficiently *concave* as a function of his perceived type, then the marginal benefit of being perceived as the best is relatively low. Thus, the best types of Sender are happy to wait for outside information, and the reverse unraveling loop gains traction. The resulting equilibrium is either fully opaque, or features non-monotonic strategies with disclosures made only by mediocre Senders. If payoffs are sufficiently *convex*, on the other hand, we obtain monotone equilibria where only the best types disclose, as in games without outside information.

Our results on convex and concave payoffs deliver further empirical predictions. In an application to corporate disclosure, we show that high-quality firms are most likely to disclose when they are financed by equity (a convex claim on returns), but less likely to disclose when financed by debt (a concave claim). The existing literature emphasizes managers' desires to keep stock pries high (Verrecchia, 1983; Acharya et al., 2011) and to enhance market liquidity (Diamond and Verrecchia, 1991). Our model implies, in addition to these factors, capital structure and executive compensation play a key role in determining disclosure strategies.

A full characterization of equilibria in the general case is not tractable, but we demonstrate that similar mechanisms to the binary case come into play. In particular, due to a logic akin to 'reverse unraveling', equilibrium outcomes need not be continuous in the model's primitives. For any set of payoffs and prior beliefs, a small improvement in the quality of outside signals can take us from full disclosure to an opaque equilibrium where no inside disclosures are made at all. Moreover, for any equilibrium where Sender discloses with positive probability, more informative outside signals (in the sense of Blackwell, 1953) can leave Receiver worse informed overall (also in the Blackwell sense).

Related literature

Our work contributes to the theoretical literature on verifiable communication, as well as the applied literature on financial crises and stress tests.

Grossman and Hart (1980), Grossman (1981), Milgrom (1981) and Milgrom and Roberts (1986) point out the existence of full disclosure or unraveling equilibria in verifiable disclosure games.[9] Another strand of work shows that equilibria with limited disclosures arise when disclosure costs are significant (Jovanovic, 1982; Verrecchia, 1983) or when it is uncertain whether Sender has any private information (Dye, 1985; Shin, 1994, 2003). We complement this research by focusing on situations where little or no information is disclosed by insiders,[10] in a model where strategic complementarities work in favor of non-disclosure and information is fragile. Our focus on outside information connects our paper to Acharya et al. (2011), who study the link between (outside) public announcements and the endogenous timing of inside disclosures.

Feltovich et al. (2002) and Daley and Green (2014) study signaling games with outside information and two or three types of Sender. As in the first step of our reverse unraveling mechanism, the highest-quality Senders have weaker incentives to acquire signals if their quality is likely to be revealed. Local crowding out effects have also been studied in the case case of non-verifiable disclosures with outside information, for instance in the accounting literature by Dye (1983) and Gigler and Hemmer (1998). We complement this work by deriving the reverse unraveling loop, the fragility of information, and the importance of the shape of payoffs. We focus on verifiable disclosure, which is a special but more tractable case of signaling. This focus allows us to derive new insights in a setting with many types of Sender.

In the applied literature on stress tests, recent work has focused on the optimal design of regulatory (outside) information disclosure when this is the only signal available to markets. Goldstein and Leitner (2017) characterize optimal stress tests in a lemons market. Bouvard et al. (2015) study the credibility of stress testing policy, Faria-e-Castro et al. (2016) analyze the interaction between bailout policies and stress test regimes, and Orlov et al. (2017) focus on macro-prudential stress tests that inform on the correlation of risk across banks. Inostroza and Pavan (2017) characterize the optimal design of information, with applications to stress tests, in a global game of regime change. We complement this literature by analyzing the constraints that endogenous inside disclosures place on stress test design.

More generally, we propose a model where asymmetric information in the financial system

[9] Hagenbach et al. (2014) extend this line of work to a more general class of games with pre-play certifiable communication.

[10] Mathios (2000) and Jin and Leslie (2003) provide empirical evidence of incomplete disclosure.

persists in bad times due to a lack of disclosures, consistently with the stylized fact that asymmetric information persists during financial crises (Mishkin, 1990; Gorton, 2008). Thus, in addition to the literature on stress tests, our results complement research which uses asymmetric information to generate persistent downturns (e.g. Mankiw, 1986; Boissay et al., 2015; Heider et al., 2015), and which studies optimal policy responses to 'lemons' problems (Philippon and Skreta, 2012; Tirole, 2012).

2 Inside and outside information: Binary actions

We study a game between a *Sender (he),* who has the opportunity to disclose verifiable inside information, and a *Receiver (she),* who decides on a binary action $a \in \{0, 1\}$, based on Sender's disclosures and on outside information.

We invite the reader to think of this abstract setup in terms of our leading example: The action a can capture the collective decision of investors to run on their bank ($a = 1$) or not ($a = 0$). The bank can disclose verifiable information about the quality of its assets, and outside outside information is made available by policymakers, for example, in the form of stress tests. In Section 3, we derive an exact micro-foundation of this interpretation.

Inside information Sender privately observes his 'type' $\theta \in [\underline{\theta}, \bar{\theta}]$, which is drawn from a commonly known prior distribution $F(\theta)$, with smooth density $f(\theta) > 0$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$. Sender can send a message $m \in \{\theta, \emptyset\}$ to Receiver: $m = \theta$ verifiably reveals his type and incurs a utility cost $c(\theta)$, while $m = \emptyset$ conveys no verifiable information, since it can be sent by any type, but is costless. This binary message space makes for a particularly clean exposition of our main ideas. We return to more general message spaces in Section 4 and Online Appendix F.3.

Outside information In addition to inside information m, Receiver observes an outside signal s, which has compact, convex support $S(\theta)$ and is drawn from a conditional distribution $H(s|\theta)$, with smooth density $h(s|\theta) > 0$ and a bounded first derivative $h_s(s|\theta)$ for all $s \in S(\theta)$. High outside signals are good news in the sense of Milgrom's (1981) Monotone Likelihood Ratio Property (MLRP): Regardless of the prior distribution of θ, the conditional expectation of any increasing function of θ is strictly increasing in s.

Preferences Sender's utility is $a - c(\theta) \times \mathbf{1}_{m=\theta}$; he enjoys high actions but suffers the cost of disclosure. Receiver's utility is $a(\theta - p)$, so that she chooses $a = 1$ if and only if she believes that $\mathbb{E}[\theta] \geq p$. Here, p parametrizes Receiver's prior propensity to take the high action.[11] We focus on the non-trivial case, in which *(i)* $\mathbb{E}[\theta] < p$: the prior is pessimistic enough so that Receiver would take the low action without further information, and *(ii)* $c(\theta) < 1$: disclosure costs are smaller than the benefits to Sender of inducing a high action. This simple setup nests a much wider class of games with binary responses, subject only to the standard restrictions that (i) Sender prefers high to low actions, and (ii) Receiver prefers high actions only if θ is high.[12]

Game timing and equilibrium definition We consider the following game of communication:

1. Sender privately observes θ, and chooses a message m.

2. Receiver observes m, as well as the outside signal s.

3. Receiver chooses an action $a \in \{0, 1\}$.

We consider Perfect Bayesian Equilibria: Sender and Receiver choose messages and actions to maximize expected payoffs, and the Receiver's posterior beliefs about θ are derived using Bayes' law on the equilibrium path. Off the equilibrium path, we require that Receiver places zero probability on type θ' if she observes an outside signal $s \notin S(\theta')$.[13]

An key assumption is that Sender commits to a pre-emptive disclosure before he knows the realization s of outside information. One of our key intuitions will be that the best types sometimes have *weaker* incentives to disclose if they anticipate a favorable realization of s. In an alternative model where verifiable messages can be sent *between* the realization of s and Receiver's action a, the best types would have *stronger* incentives to make such a disclosure. However, the case of pre-emptive disclosures is relevant in many applications. In financial markets, investors respond very quickly to bad news, such as a failed stress test or credit downgrades, and irreparable damage to firms' prospects may be done if they wait until

[11] In our model of bank runs in Section 3, investors' propensity to run p measures the degree of illiquidity of the bank's long-term assets. Intuitively, as p increases, the coordination among investors becomes stronger, and bank runs are more likely to occur in equilibrium.

[12] Suppose that Receiver's utility is $v(a, \theta)$, assuming only that the net benefit $\Delta(\theta) = v(1, \theta) - v(0, \theta)$ of the high action is increasing in θ. Without loss of generality, we can re-define Sender's type as $\tilde{\theta} = \Delta(\theta) - p$, yielding a game that is equivalent to our setup. Furthermore, suppose that Sender's utility is $u(a, \theta)$, assuming only that $B(\theta) = u(1, \theta) - u(0, \theta) > 0$. Our arguments below imply that equilibrium play is fully determined by Sender's benefit-cost ratio $B(\theta)/c(\theta)$. Thus we can translate Sender's preferences as $\tilde{u}(a, \theta) = a$ and $\tilde{c}(\theta) = c(\theta)/B(\theta)$ without loss, consistently with our setup. Seidmann and Winter (1997) and Giovannoni and Seidmann (2007) study verifiable message games where the above conditions on preferences are relaxed.

[13] The latter refinement is natural, and common in the applied literature (e.g. Angeletos et al., 2006). Moreover, it is inconsequential when outside signals full support (i.e. $S(\theta) = S$ for all θ). When full support fails, the refinement allows for a clean characterization of equilibria. We will see that our main results remain valid in the case of full support, and therefore do not hinge on this refinement.

after the event to prove their quality. This is especially relevant where verifiable information takes time to prepare and circulate, for example due to delays in preparation and external auditing. More generally, if economic agents have limited capacity for processing information as in Sims (2003), Receiver may be unable to (or rationally choose not to) process further communications by Sender once the outside signal s has resolved a significant portion of the uncertainty.[14]

Regularity condition In this Section, we impose a mild regularity condition. For any given realization s of the outside signal, we assume that the function

$$J(\theta) = H(s|\theta) - c(\theta) \tag{1}$$

crosses zero at most once. If it crosses once, then it must cross from above. The function $J(\theta)$ compares two terms. The first term 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 by MLRP. 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 holds when disclosure costs do not decrease too quickly with θ. In particular, it is guaranteed to hold when disclosure costs are fixed or increasing in θ, and when outside information s is precise enough.[15]

2.1 Fragility of information

It is helpful to begin with a well-known benchmark. Suppose Receiver had access to no outside information, and therefore had to rely exclusively on Sender's disclosures. Given our assumptions about prior quality, it is easy to see that in any equilibrium the best type $\bar{\theta}$ must disclose ($m = \overline{\theta}$). Moreover, the classic unraveling argument (Grossman, 1981) applies to all $\theta \geq p$, and therefore the *unique* equilibrium of the game is one in which Sender discloses whenever $\theta \geq p$. Meanwhile, types $\theta < p$ have a dominant strategy to stay quiet, but in equilibrium, their silence reveals that $\theta < p$. Receiver therefore takes the high action if and

[14] A potential variation on our model is a setting where the verifiable report $m = \theta$ takes time to prepare, but where Sender can prepare it *in advance* and decide whether to release it once s has been observed. In this environment, Sender has stronger incentives to prepare the report than in our model, because he retains the option to keep it to himself in case s turns out to be better news than the truth. However, similar arguments to our main results go through: The best types of Sender have a relatively weak incentive to prepare a verifiable report, because they anticipate that the outside signal s will be good enough to secure a favorable action. Therefore, the effects we emphasize will continue to arise.

[15] For example, in the 'truth plus noise' case where $s = \theta + k\epsilon$, with k small enough, the distribution $H(s|\theta)$ is close to one for types $\theta < s$ and close to zero for types $\theta > s$. Since the disclosure cost satisfies $0 < c(\theta) < 1$, the difference between this probability can only have one crossing with zero.

only if $\theta \geq p$, as she would under full information. Throughout this Section, we will refer to an equilibrium where all types $\theta \geq p$ disclose as an *unraveling equilibrium.*

To illustrate our main point on the fragility of disclosures in the starkest manner, we first focus on outside signal distributions for which full support fails – in this case, there exist signals which distinguish high and low types of Sender beyond doubt. Of course, if outside signals have full support, so that $S(\theta) = S$ for all θ, an unraveling equilibrium can always be sustained. Indeed, if disclosure by all $\theta \geq p$ is expected, then Receiver interprets silence as evidence that $\theta < p$, and no realization of s can convince her otherwise. We nonetheless show below that our results on the fragility of information extend to the full support case in a natural sense.

Fragile information: Reverse unraveling with bounded support

Suppose outside signals do not have full support: that is, $S(\theta)$ is not the same for all types. Let $\hat{s} = \sup \cup_{\theta<p} S(\theta)$ denote the largest outside signal that any type $\theta < p$ can draw. Now any outside signal $s > \hat{s}$ reveals without doubt that $\theta > p$, and therefore guarantees that Receiver chooses the high action $a = 1$. Hence, even if Receiver expects transparency, the best type $\bar{\theta}$ has an incentive to deviate to silence if

$$H(\hat{s}|\bar{\theta}) < c(\bar{\theta}), \tag{2}$$

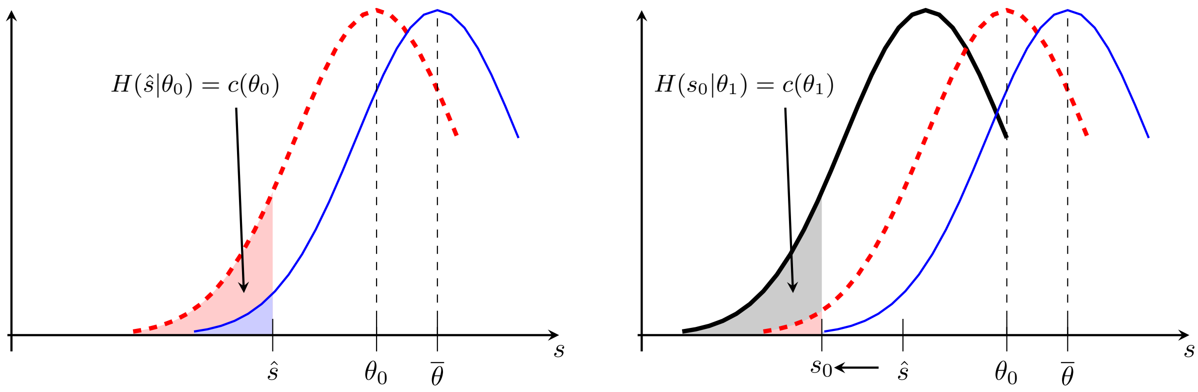
Here, the potential cost to type $\bar{\theta}$ of drawing an unimpressive signal $s \leq \hat{s}$, and therefore triggering the low action $a = 0$ in the absence of inside information, is smaller than the cost of disclosure.

If condition (2) holds, then there must be some interval of highest types, $\theta \in (\theta_0, \bar{\theta}]$, who have a dominant strategy to stay quiet in equilibrium. Figure 1a illustrates this effect. Crucially however, when types in $[0, p) \cup (\theta_0, \bar{\theta}]$ are expected to stay quiet, the left-hand side of (2) actually overestimates the marginal benefit of disclosure: If Receiver believes that types in $[0, p) \cup (\theta_0, \bar{\theta}]$ stay quiet, the critical outside signal that guarantees the high action falls to some $s_0 < \hat{s}$, which solves

$$E[\theta|s_0, \theta \notin [p, \theta_0)] = p.$$

Staying quiet now becomes more attractive. As a result, a wider set of high quality types $\theta \in (\theta_1, \theta_0]$ now have an (iterated) dominant strategy to stay quiet – see Figure 1b.

Indeed, in contrast to classic unraveling, there are now *strategic complementarities in non-disclosures.* Since Receiver's posterior expectations at critical signal s_0 are exactly p, the



(a) Deviations from full disclosure

(b) Amplification

Figure 1: **Reverse unraveling.** The blue (solid) curve in panel (a) is the density of outside signals drawn by the best type of Sender $\bar{\theta}$. The blue shaded area is the left-hand side of (2). We draw the case where (2) holds, so that type $\bar{\theta}$ would deviate from full disclosure. The red (dashed) curve is the density of signals for the critical type θ_0 for whom (2) holds with equality. All types above θ_0 have a dominant strategy to stay quiet. When it is common knowledge that types $\theta > \theta_0$ stay quiet, the critical signal that ensures a high action falls to s_0. The probability that type θ_0 draws a signal below s_0 (the red shaded area) is now less than the cost disclosure. The thick black curve in panel (b) is the scaled density of signals for the new critical type $\theta_1 < \theta_0$ who is indifferent between disclosure and staying quiet.

decision by more good types $\theta \in (\theta_1, \theta_0] > p$ to stay quiet further improves the interpretation of outside signals. This strategic complementarity continues to amplify silence. In response to types above θ_1 staying quiet the critical signal falls further, additional types $\theta \in (\theta_2, \theta_1]$ prefer to stay quiet, silence becomes better news still, and so forth. We call this process *reverse unraveling*. Letting θ_n be the highest type who still discloses at the n^{th} iteration, we can see that no type above $\tilde{\theta} = \lim_{n\to\infty} \theta_n$ discloses in any equilibrium. Below, we prove formally that $\tilde{\theta}$ is bounded away from the best type $\bar{\theta}$: a *discrete* set of high-quality types must stay quiet in equilibrium whenever the best type stays quiet.

This logic leads to fragile information, as captured by a discontinuity in equilibrium outcomes. Intuitively, consider a situation where the quality of outside signals gradually increases, starting from pure noise. Then, when the quality of outside signals crosses a critical threshold, (2) is guaranteed to hold. As we cross this threshold, we move discontinuously from a situation where full transparency is an equilibrium, to a situation where no type above $\tilde{\theta}$ makes any disclosure. We now show this more rigorously, and also consider the case where outside signals have full support.

Fragile information: The general case

Of course, when signals have full support, then we have $H(\hat{s}|\theta) = 1$ for all θ, and condition (2) cannot hold unless outside signals are perfectly revealing. At first glance, it therefore appears that the fragility of information in our model relies on a violation of full support (or indeed on the refinement that Receiver does not attach positive weight to types who cannot draw the outside signal she observes). However, the general effects of outside information are subtle. In particular, due to the strategic complementarity we have highlighted, our model admits multiple equilibria. We now establish that both the *unraveling* equilibrium that we have studied so far, and *less informative* equilibria that exist alongside it, exhibit fragility. The latter case remains relevant even in the case of full support.

Before proceeding, we establish a useful property of equilibria, namely, that Sender and Receiver must both follow forms of threshold strategy. Indeed, by the MLRP property on outside signals s, if Receiver observes non-disclosure by Sender she will play $a = 1$ only if the outside signal exceeds some threshold $s^\star$, where $s^\star$ is endogenously determined by Receiver's conjecture about Sender's strategy (whatever form this may take). This property makes the search for Sender's best response much simpler. Of course, if $\theta < p$, then Sender has a dominant strategy to stay quiet. But if $\theta \geq p$, then Sender prefers to stay quiet if and only if $H(s^\star|\theta) \leq c(\theta)$ – when the costs of disclosure exceed the probability of drawing a bad enough outside signal to warrant the low response ($a = 0$) when Sender remains quiet. Under our regularity condition, we can then show the following:

Lemma 1. *In any equilibrium, there exists a threshold, $\theta^\star$, which summarizes equilibrium play as follows:*

- *Sender discloses if $\theta \in (p, \theta^\star)$ and stays quiet if $\theta < p$ or $\theta > \theta^\star$.*
- *Receiver chooses $a = 1$ if Sender discloses, or if Sender stays quiet and the outside signal is $s > s^\star(\theta^\star)$, defined as the lowest outside signal s satisfying $\mathbb{E}[\theta|\theta \notin (p, \theta^\star), s] \geq p$. If no such s exists, then $s^\star(\theta^\star) = \infty$.*

We can now describe equilibria of our game with a single parameter $\theta^\star$, which denotes the highest good type $\theta > p$ that chooses to disclose. Figure 2 shows a simple diagram with which we can illustrate equilibria. Let the 'best response function' $BR(\theta^\star)$ denote the highest type of Sender who prefers to disclose when Receiver expects disclosures from types $\theta \in [p, \theta^\star]$.[16] Formally, we define

$$BR(\theta^\star) = \sup\{\theta \geq p : H(s^\star(\theta^\star)|\theta) \geq c(\theta)\} \tag{3}$$

[16] Without loss of generality, we focus on equilibria where Sender chooses $m = \theta$ if indifferent, and Receiver takes $a = 1$ if indifferent.

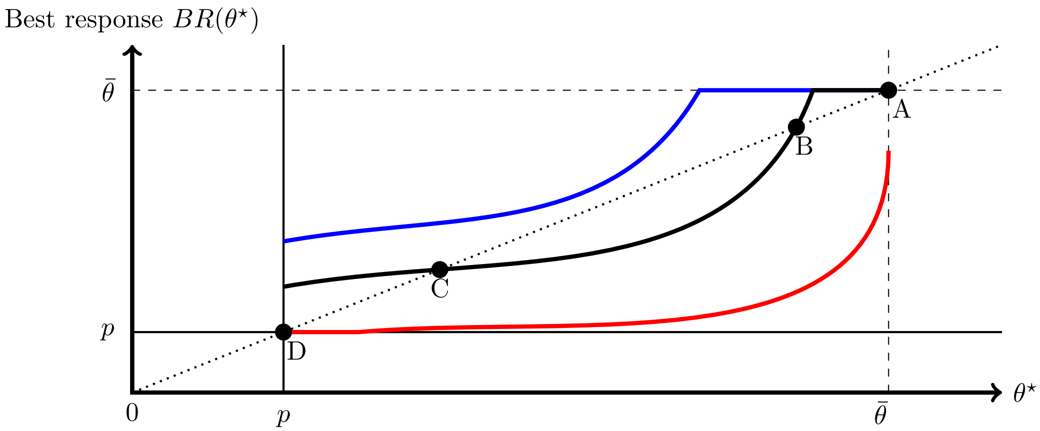


Figure 2: **Best response functions and equilibria.** The blue (upper) curve shows the best response when outside signals are imprecise; the unique equilibrium is unraveling at point A. The black (middle) curve is drawn for intermediate signal precision; there are multiple equilibria at A, B and C due to strategic complements. The red (lower) curve is drawn for low signal precision; the only equilibrium in this case is full opacity at D.

and we set $BR(\theta^\star) = c$ if $H(s^\star(\theta^\star)|\theta) < c(\theta)$ for all θ. The cutoff $\theta^\star$ constitutes an equilibrium if and only if it solves the fixed point equation $BR(\theta^\star) = \theta^\star$.

The best-response mapping is upward-sloping due to the strategic complementarities we have discussed. As $\theta^\star$ rises, fewer high-quality Senders are expected to stay quiet, silence becomes better news, and as a result high-quality senders have a stronger incentive to disclose. A key feature of our model is that this complementarity becomes very strong when only a small set of high-quality types stay quiet:

Lemma 2. *If the best response function $BR(\theta^\star)$ satisfies $BR(\theta^\star) \in (p, \bar{\theta})$ in a neighborhood of the best type $\bar{\theta}$, then it is continuously differentiable in this neighborhood, and moreover, it becomes infinitely steep:*

$$\lim_{\theta^\star \uparrow \bar{\theta}} BR(\theta^\star) = \infty. \tag{4}$$

Lemma 2 represents the analytical equivalent of the intuitive 'reverse unraveling' argument that we presented above. It shows that strategic complementarities can be infinitely strong in a neighborhood of an unraveling equilibrium where $\theta^\star \simeq \overline{\theta}$. Indeed, suppose we begin in an unraveling equilibrium, and reduce transparency at the margin so that types $\theta \in (\bar{\theta} - \epsilon, \bar{\theta}]$ stay quiet. Equation (4) implies that, if $BR(\theta) < \overline{\theta}$ near $\overline{\theta}$, then $BR(\overline{\theta} - \epsilon) < \overline{\theta} - \epsilon$. Thus, a conjecture by Receivers that some small set of high types do not disclose can incentivize an even larger set of types not to disclose, thus becoming a self-fulfilling prophecy. Intuitively, this captures the essence of reverse unraveling and implies that any other equilibrium threshold must be strictly bounded away from $\overline{\theta}$.

Given the strategic complementarities we identify in non-disclosures, our model can admit multiple equilibria (see Figure 2). Intuitively, disclosures by high-quality types of Sender can be self-fulfilling because they imply that no news is bad news. We focus in particular on the *most transparent* equilibrium, associated with the highest equilibrium disclosure threshold $\theta^\star_{max} = \sup\{\theta \geq p : BR(\theta) = \theta\}$, and on the *least transparent* equilibrium, associated with $\theta^\star_{min} = \inf\{\theta \geq p : BR(\theta) = \theta\}$; both exist by Tarski's fixed point theorem We have loosely argued that the most transparent equilibrium is fragile: Full transparency is an equilibrium ($\theta^\star_{max} = \bar{\theta}$) unless the best type has an incentive to deviate from it in the sense of Equation (2); as soon as he does, BR is interior and there is a discrete shift in incentives due to reverse unraveling. This does not apply when outside signals have full support, since in this case full transparency is always an equilibrium.

We find that the least transparent equilibrium is also fragile, and that this fragility survives even when signals have full support. To make these statements rigorous, we refer to a *revealing path* as a smooth sequence of outside signals s_t, indexed by a parameter $t \in [0, 1]$, such that s_0 is pure noise and s_1 perfectly reveals Sender's type θ.[17]

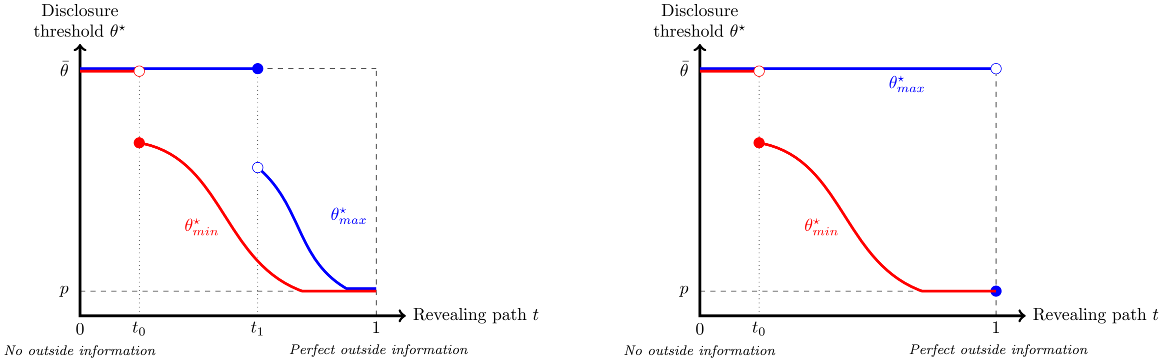
Proposition 1. *For any revealing path s_t, there exist two thresholds $t_0 \in (0, 1)$ and $t_1 \in (0, 1]$, such that $t_1 > t_0$, and:*

- *The least transparent equilibrium is discontinuous around t_0: $\theta^\star_{min} = \bar{\theta}$ for all $t < t_0$, and $\theta^\star_{min} < \bar{\theta}$ for $t = t_0$.*
- *The most transparent equilibrium is discontinuous around t_1: $\theta^\star_{max} = \bar{\theta}$ for all $t \leq t_1$; and if $t_1 < 1$, then $\theta^\star_{max} \leq \theta_1$ for $t \in (t_1, t_1 + \delta)$, where $\theta_1 < \bar{\theta}$ and $\delta > 0$.*

Figure 3 illustrates the result. Along a revealing path, when $t \simeq 0$, outside signals are almost pure noise, and as in the classical case without outside information, the unique equilibrium is unraveling so that $\theta^\star_{min} = \theta^\star_{max} = \bar{\theta}$. First, as outside signals improve,[18] we arrive at a threshold t_0 at which a less transparent equilibrium exists. This transition is not gradual: The Proposition shows that $\theta^\star_{min}$ jumps strictly below $\bar{\theta}$. Second, as outside signals improve further, we may arrive at a second threshold t_1 beyond which the unraveling equilibrium no longer exist. This is precisely the point beyond which (2) holds and the best type would deviate from an unraveling outcome. Again, the transition around t_1 is discontinuous, and

[17] Formally, we assume that the densities $h(s|\theta; t)$ satisfy our assumptions above, are continuous in t, and that (i) $h(s|\theta; 0) = h_0(s)$ for all θ; and (ii) $\lim_{t\to1} h(s|\theta; t) = \delta(\theta)$ for all θ, where δ is the Dirac delta. Smoothness here means that h is continuously differentiable in t.

[18] While thinking of increasing t as an improvement in outside signals helps to guide the intuition, we do not formally require that signals continuously improve along a revealing path. Since Proposition 1 holds for *any* revealing path, it is trivially also true for paths along which $s_{t'}$ is more informative (e.g. in a Blackwell sense) than s_t whenever $t' > t$.



(a) Outside signals without full support

(b) Outside signals with full support

Figure 3: **Fragility of information.** Panel (a) shows a revealing path along which outside signals do not always have full support. Both the most transparent equilibrium, associated with disclosures by types $\theta \in [p, \theta^\star_{max}]$, and the least transparent equilibrium associated with disclosures by $\theta \in [p, \theta^\star_{min}]$, are fragile. Panel (b) shows the case of full support. Here, the most transparent equilibrium is always unraveling, with disclosures by all $\theta \geq p$. The least transparent equilibrium remains fragile around $t = t_0$.

when the unraveling equilibrium breaks down, the most informative equilibrium $\theta^\star_{max}$ has a downward jump. As we have argued above, the former case (t_0) is more general than the latter (t_1) since it arises for all possible revealing paths, even those where s_t has full support throughout. By contrast, an interior t_1 exists only if full support fails. The two panels in Figure 3a and 3b highlight the difference between the two cases.[19]

Intuitively, the second result in the Proposition, where unraveling breaks down at t_1, mirrors exactly the reverse unraveling mechanism we have discussed. The first result is more nuanced, but the economics are similar. As the quality of outside information improves to the point t_0, a less transparent equilibrium than unraveling becomes sustainable. One might expect this transition to be smooth, with the new equilibrium involving non-disclosure only by a small set of the best types. However, this leads to a contradiction: If only a small set of the best types prefer to stay quiet, then by Lemma 2, a discrete set of worse types must wish to join in. Following this logic, we find that any new equilibrium must involve a discrete set of types that stay quiet, which establishes the discontinuity around t_0.

[19] Our exposition is based on the case where Sender's type is drawn from a bounded interval $[\underline{\theta}, \bar{\theta}]$. Similar intuitions apply in the case where $\theta \in \mathbb{R}$ is unbounded, as we discuss in Online Appendix D.1. In this case, an unraveling scenario with $\theta^\star = \infty$ is also the unique equilibrium for all $t \leq t_0$ along a revealing path, but at $t = t_0$ there exists an alternative equilibrium in which only a finite interval of types $\theta \in [p, \theta^\star_{min}]$ discloses.

2.2 Selecting an equilibrium

We have established the fragility of information, in the sense that the least transparent and most transparent equilibria of our game both become discontinuously less informative along any revealing path of outside signals. However, Proposition 1 does not provide unique predictions. We now show that standard equilibrium refinements tend to select the least transparent equilibrium. They therefore yield the lower path in Figure 3, which involves fragility of information regardless of the structure of outside signals, as the unique prediction.

First, we note that the least transparent equilibrium is Pareto-dominant from Sender's perspective. This follows from simple revealed preference and the nature of externalities across Sender types. In the least transparent equilibrium, an interval of good types $(\theta^\star_{min}, \bar{\theta}]$ prefer to stay quiet, thereby earning at least as much as they could by disclosing. Moreover, disclosures by types above p unambiguously imposes negative externalities on other *non-disclosing* types of Sender.[20] Thus, the least transparent equilibrium improves payoffs for all Sender types. The Sender-preferred equilibrium would be selected, for example, if Sender were able to act as a mechanism designer and commit to an (incentive-compatible) disclosure strategy before the game begins.[21]

Second, we show that the least transparent equilibrium is the unique *neologism-proof* equilibrium, introduced in the context of cheap-talk games by Farrell (1993). We adapt this criterion to verifiable disclosure games, following Bertomeu and Cianciaruso (2015). Since the full support case is the one with which we are most concerned and allows the cleanest application of the refinement, we assume here outside signals s have full support. Given an equilibrium disclosure strategy described by a threshold $\theta^\star$, we refer to a *self-signaling set* as a set of Sender's types $\mathcal{S} \subset \left[\underline{\theta}, \bar{\theta}\right]$ for which:

- All types $\theta \in \mathcal{S}$ have strictly lower expected utility in equilibrium than a situation in which Receiver *(i)* believes that $\theta \in \mathcal{S}$, *(ii)* observes outside information s, drawn from $h(s|\theta)$, and *(iii)* acts according to her best response given this belief and information; and

- All types $\theta \notin \mathcal{S}$ have weakly higher expected utility in equilibrium than in the above situation.

An equilibrium $\theta^\star$ is *neologism-proof* if there are no self-signaling sets. Intuitively, if there is a self-signaling set, then all types of Sender $\theta \in \mathcal{S}$ can gain, relative to the equilibrium

[20] By continuity of Sender's utility, all types in the interior of the interval strictly prefer their current equilibrium outcome.

[21] For related reasons, the literature on Bayesian Persuasion (Kamenica and Gentzkow, 2011) focuses on Sender-preferred equilibria.

outcome, by using a 'new' message (a neologism) to announce that $\theta \in \mathcal{S}$, and moreover, no other types would have an incentive to mimic this announcement, thus making it credible information.

The least transparent equilibrium is the unique survivor of this refinement:

Proposition 2. *The unique neologism-proof equilibrium is the least transparent one, with* $\theta^\star = \theta^\star_{min}$*.*

This is intuitive. The fact that more transparent equilibrium $\theta^\star > \theta^\star_{min}$ cannot survive the refinement follows immediately from our revealed preference argument above. Indeed, if $\theta^\star$ were played, then all types outside the set $[p, \theta^\star_{min}]$ would be strictly better off if equilibrium switched to $\theta^\star_{min}$. All types inside $[p, \theta^\star_{min}]$, by contrast, are indifferent to this change because they fully disclose in both situations. Thus, the types $\theta \notin [p, \theta^\star_{min}]$ form a self-signaling set. The proof of Proposition 2 further establishes existence, i.e., that the least transparent equilibrium allows no self-signaling.

To complement these refinements, we show in Online Appendix D.2 that unraveling equilibria are often unstable, in a sense that has a natural definition, which again suggests the selection of less transparent outcomes even when neologism-proofness is not required.

To summarize, several natural refinements to our equilibrium context predict that the least transparent equilibrium, with the lowest disclosure threshold $\theta^\star_{min}$, is selected.[22] In this case, the relevant fragility of information is the discontinuity around t_0 in Proposition 1 and Figure 3. Moreover, when we apply these refinements, information is fragile regardless of whether outside signals have full support.

2.3 Crowding in or crowding out?

Our results so far emphasize a strong 'crowding out' effect of outside information on inside information: As outside signals improve along a revealing path, equilibrium play jumps from a very informative outcome (unraveling) to a discretely less informative one. We now evaluate whether improvements in outside information also crowd out inside disclosures locally, around equilibria with intermediate levels of disclosure $\theta^\star \in (p, \bar{\theta})$. Generally, it turns out that local effects are ambiguous; better outside information can crowd out or crowd in inside disclosures at the margin.

[22] Note that the Cho and Kreps (1987) intuitive criterion, or indeed the D1 criterion due to the same authors, do not provide such discipline in our environment, since there are no messages off the equilibrium path that might dominate the equilibrium outcome from Sender's perspective. Bertomeu and Cianciaruso (2015) give a more rigorous exposition of this point.

The intuition is cleanest if we step slightly outside our baseline framework and assume that types and outside signals are normally distributed.[23] Let $\theta \sim \mathcal{N}(\mu, \sigma^2)$, and suppose that outside signals take the form $s = \theta + k\epsilon$, where $\epsilon \sim \mathcal{N}(0,1)$ and the parameter $k \geq 0$ captures noise in outside information. For ease of exposition, we make disclosure costs constant in the Normal case: $c(\theta) = c$. We characterize the response of an equilibrium cutoff $\theta^\star$ to a change in the noise parameter k. We restrict attention to local changes around a stable equilibrium $\theta^\star$, where the best response function $BR(\theta)$ crosses the 45-degree line from above.[24]

Lemma 3. *Suppose θ, s are jointly Normal and consider a stable interior equilibrium $\theta^\star \in (p, \bar{\theta})$. More precise outside information crowds out inside disclosures ($\frac{d\theta^\star}{dk} > 0$) if $\mu < p$ and $c \leq \frac{1}{2}$. Conversely, there exists a function $\overline{c}\,(\mu)$ and parameter $\overline{\mu}$ such that more precise outside information crowds in disclosures ($\frac{d\theta^\star}{dk} > 0$) whenever $\mu > \overline{\mu}$ and $c = \overline{c}\,(\mu)$.*

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 μ of θ is low) while crowding in is more likely to be a feature when μ is high.

To see why, we first consider the low-μ case. Under Normality, Receiver's posterior expectations given only the public signal can be expressed $E[\theta|s] = \alpha\mu + (1-\alpha)s$, where the weight α on the prior is increasing in k. When public signals become more informative ($\downarrow k$), Receiver places more weight on the outside signal. When μ is low, this shift increases expectations and the marginal Receiver must become more willing to take the high action when Sender stays quiet. Therefore, the critical signal $s^\star(\theta^\star)$ falls. Turning to Sender's incentives, if $c \leq \frac{1}{2}$, type $\theta^\star$ believes that the risk of drawing an outside signal $s < s^\star(\theta^\star)$, which would trigger the low action, is a left-tail event (otherwise, he would strictly prefer to disclose given that the cost is low). In this scenario, an increase in signal precision additionally makes type $\theta^\star$ more confident that $s > s^\star(\theta^\star)$. Receiver now interprets silence more favorably and Sender becomes unambiguously more confident in silence. Both effects imply that equilibrium disclosures must decrease and we obtain crowding out.

The converse intuition applies when the prior μ is high relative to the cost of disclosure; Receiver shifts weight away from her optimistic prior and interprets silence less favorably, while Sender becomes less confident. Strictly, this heuristic argument does not take into account that in such situations, $s^*(\theta^\star)$ is likely to fall below $\theta^\star$. In this case, more precise public information induces an offsetting benefit to staying quiet by reducing the probability of the

[23] This is outside the baseline model because types and signals have unbounded support; this case is analyzed formally in Online Appendix D.1.

[24] For a formal definition of stability, see Online Appendix D.2.

tail event $s < s^*(\theta^\star)$. The Proof of Lemma 3 deals with this complication by simultaneously reducing the cost of disclosure to keep $\theta^\star$ unchanged.

3 Financial crises and stress test design

In Section **2**, we motivated our analysis as a model of bank disclosures during financial crises. We now make this link more explicit by micro-founding investors' incentives, and address the policy question that motivates our main application: How much information should a policymaker such as the Fed release into a potentially panicked banking system?

Consider a bank (Sender) interacting with a continuum of short-term risk-neutral investors (Receivers) and a policymaker. 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 the bank. The bank invests this cash in a long-term project, which yields a stochastic gross return r at date 2.

However, projects are illiquid: if a proportion $l \in [0, 1]$ of the long-term investment is withdrawn at time 1, then the return on the remaining projects is reduced to $r - 2pl$ (when they eventually mature at time 2). The parameter $p > 0$ measures the degree of asset illiquidity. Investors who withdraw at date 1 are each entitled to an immediate payment of one unit of cash. Investors who wait to withdraw at date 2 are residual claimants on the bank's assets at time 2. Thus, each investor who withdraws receives a certain payoff of 1, while each investor who rolls over receives a stochastic payoff $r - 2pl$. We have followed the approach of Morris and Shin (2000), who make the analysis of investor's incentives particularly tractable by assuming that the liquidation technology is linear in the return r and the proportion of withdrawals l. Moreover the structure of assets and contracts with investors are taken as given. Both assumptions can be relaxed using the techniques of Goldstein and Pauzner (2005) without affecting the qualitative insights below.

The gross return on assets at date 2 is

$$r = 1 + \theta + \eta.$$

At date 1, the bank privately observes the first return component $\theta \in [0, \bar{\theta}]$, which measures asset quality and is drawn from a smooth distribution with density $f(\theta) > 0$, while investors observe only an outside signal $s = \theta + k\epsilon$, where ϵ is a random variable with smooth, log-concave density $h(\epsilon)$. For concreteness, we will interpret s as the publicly observable result of a regulatory stress test. The parameter $k \geq 0$ measures noise contained in this test; when $k = 0$ the stress test perfectly reveals θ, while $k = \infty$ corresponds to pure noise. At date 1, the bank can further send investors inside information denoted $m \in \{\emptyset, \theta\}$. As before, $m = \theta$ credibly reveals asset quality and costs $c(\theta)$, while $m = \emptyset$ reveals no verifiable information but is costless.

The second return component $\eta \sim N(0, \sigma_\eta^2)$ is independent of θ and observed by neither the bank nor investors directly. However, each investor $i \in [0,1]$ receives a private signal $t^i = \eta + \zeta^i$, where $\zeta^i \sim N(0, \sigma_\zeta^2)$ is independent of θ, η and ζ^j, $j \neq i$. Introducing η allows us to simultaneously model *(i)* aggregate uncertainty over returns (and hence value to communication), and *(ii)* small deviations from common knowledge, which induce a unique equilibrium in the coordination game among investors.[25] By assuming that η and θ are independent, we retain tractability, but abstract from issues of multiplicity in the coordination 'subgame' among investors.[26] We will consider the standard noiseless limit where both η and t^i collapse to zero.

We emphasize that our results are not about global games; rather, we choose this canonical tool in order to speak to the existing literature. The same qualitative insights obtain in a model where there is no coordination motive, as long as investors have a reason (e.g. concerns about moral hazard in a badly capitalized bank) to withdraw their funds when they are pessimistic about asset values θ.

Note that in the limiting case $\eta \to 0$, investors' joint utility if nobody withdraws is $1 + \theta$, while in a bank run scenario where everybody withdraws it is 1. By assuming that $\theta \geq 0$, we therefore restrict attention to banks that are solvent – continuation dominates a run in terms of aggregate utility – but potentially illiquid when individual investors decide to withdraw. In Online Appendix E.2, we extend the model to allow for insolvent banks.

We analyze strategic communication between bank and investors as before. In addition, to address our policy question, we consider the information design decision of a policy-maker, who chooses the noise parameter k. A low value of k can be interpreted as a more revealing stress test scenario. For parsimony, we assume that reducing the noise in outside signals is not costly for the policymaker, but this is not crucial; our main results below remain qualitatively similar when we allow stress testing to be costly. The timing of the game is then as follows:

- **Date 0**: The policymaker chooses k and commits to this choice.

[25] The additive-independent return specification follows Bouvard et al. (2015).

[26] For instance, if agents' private information were noisy signals of θ, perverse equilibria can emerge involving runs against high types, and no runs against low types. Since we are focused on the informational effects of disclosure, we abstract from these kind of multiple responses. Interestingly, we nevertheless find multiple equilibria in our game. In a global game setting without outside information, Angeletos and Pavan (2013) similarly find that signaling can generate multiplicity. Inostroza and Pavan (2017) analyze the optimal design of information in global games in detail.

- **Date 1:**
  1. The bank privately observes θ, and chooses a message m. Each investor i observes m, the outside signal s, and her private signals t^i;
  2. Each investor chooses whether to withdraw her investment from the bank.
- **Date 2:** Returns are realized and claims are settled.

Each investor acts to maximize her expected utility, taking as given the fraction l of other investors that are withdrawing. The bank acts to maximize the joint utility of all investors, net of disclosure costs. In the baseline model, we therefore rule out explicit conflicts of interest between bank managers and investors, to which we return in Online Appendix E.1. However, the coordination problem among investors introduces an implicit conflict of interest, because as in Diamond and Dybvig (1983), individual investors will not necessarily choose efficient actions.

As in Section 2, we impose a mild regularity condition on the costs of disclosure. For a given realization s of the outside signal, the function

$$J(\theta) = H(s|\theta) - \frac{c(\theta)}{\theta}$$

crosses zero at most once, and if it crosses once, then it must cross from above. The first term measures the probability of receiving an outside signal lower than s, which is strictly decreasing in θ by MLRP. The second term is the ratio of the cost of disclosure $c(\theta)$ to the benefit θ of avoiding a bank run. We require that the cost-benefit ratio does not decrease too quickly with θ. In the context of banking, this restriction is reasonable because practitioners commonly think of the costs of disclosure for financial institutions as *proprietary*, such as the costs of revealing one's investment portfolio to competitors. These costs are likely increasing in portfolio quality θ.

3.1 Equilibrium: Investors' choices and bank runs

At time 1, all investors observe the same signal s and message from the bank m. Therefore, they share a common posterior expectation over the first return component, $\mathbb{E}_\mu[\theta|m, s]$. We solve for investors' equilibrium decision, as a function of these beliefs, using a standard global games argument (see Morris and Shin, 2000). Each investor then uses her private signal t^i to learn about the second component η, and makes her withdrawal decision based on that assessment and her beliefs about other investors' behavior. 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 (given $\mathbb{E}_\mu[\theta|m,s]$) in which investor t^i withdraws if and only if $\mathbb{E}[\eta|t^i] < \eta^\star$, where $\eta^\star$ is a critical value. The argument is a simple application of Morris and Shin (2000). Indeed, as shown in Bouvard et al. (2015), in the unique limiting equilibrium a bank run occurs if and only if

$$\mathbb{E}_\mu[\theta|m,s] < p. \tag{5}$$

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 p is high.

However, notice that in our setting the quality of information available to investors responds to the Bank's equilibrium reporting behavior. Indeed, as we now describe, banks' endogenous disclosure decisions are both a meaningful driver of bank runs and affect the incentives of policymakers to design their stress tests.

3.2 Banks' inside disclosure strategies

Equation (5) implies that investors collectively behave as the binary-action Receiver in Section 2. Thus, the analysis of the communication game between the bank and investors proceeds exactly as in Section 2. In particular, for any given choice k of noise in the stress test, the bank discloses its type if and only if $\theta \in [p, \theta^\star]$, where the threshold $\theta^\star$ must be a fixed point of the best response function $BR(\theta^\star; k)$, defined as in (3) for any given k. Weak banks with $\theta < p$ have a dominant strategy to stay quiet because disclosure would lead to a certain bank run, while strong banks with $\theta > \theta^\star$ stay quiet because they expect a favorable stress test result.

Of course, the banking model inherits all the conclusions of Section 2. In particular, bank disclosures exhibit fragility (Proposition 1): Along any revealing path, moving from pure noise ($k = \infty$) to perfect stress tests ($k = 0$), both the least and most transparent equilibrium exhibit discontinuous drops in disclosure due to reverse unraveling. We now explore the implications of this mechanism for welfare and stress test design.

3.3 Stress test design

We study the policymaker's optimal choice of outside information, as captured by k. When there are multiple equilibria to the disclosure game for a given k, we focus our attention on the least transparent one, that is on the lowest equilibrium disclosure threshold $\theta_k^\star \equiv \inf\{\theta^\star | BR(\theta^\star; k) = \theta^\star\}$. This choice is natural for three reasons. First, as we describe below, it is the policymaker's preferred equilibrium. Second, it is the unique neologism-proof equilibrium and thus a somewhat focal prediction of behavior. Third, it always exists, making for well-defined comparative statics.

Given a choice of noise k, ex-ante expected welfare (as measured by aggregate investor utility) is:

$$W(\theta_k^\star; k) = 1 + \int\limits_{\theta \in [p, \theta_k^\star]} (\theta - c(\theta))\, dF(\theta) + \int\limits_{\theta \notin [p, \theta_k^\star]} Pr\left[s \geq s^\star(\theta_k^\star; k) | \theta; k\right] \theta dF(\theta) \tag{6}$$

where $s^\star(\theta_k^\star; k)$ denotes the critical outside signal below which investors run on a quiet bank in equilibrium, given that they expect banks $\theta \in [p, \theta_k^\star]$ to disclose. To understand (6), consider that if a bank chooses to disclose ($\theta \in [p, \theta_k^\star]$), it avoids a run for certain. The social value of such banks is $\theta - c(\theta)$, the value of avoiding the run less the cost of disclosure. The first integral in (6) captures this value. However, if the bank stays quiet ($\theta \notin [p, \theta_k^\star]$), a run is avoided if and only if the outside signal satisfies $s \geq s^\star$, in which case it provides investors with value θ. The second integral represents this contribution.

We have made explicit that welfare depends on k both directly and through the behavioral impact of k on the equilibrium threshold $\theta_k^\star$. An optimal policy must take both effects into account.

The direct effect alone makes for a rich analysis, and is the subject of a growing literature on stress test design which focuses on the case without inside information (e.g. Faria-e-Castro et al., 2016; Goldstein and Leitner, 2017). We will focus instead on what is new in our model, namely the behavioral effect, but first we review the key intuitions relating to the direct effect. We refer the reader to the excellent analysis in existing work (see also Bouvard et al., 2015) for a formal discussion.

In our model the case without outside information corresponds to fixing $\theta_k^\star \equiv p$, $\forall k$, and taking partial derivatives with respect to k in Equation (6) to derive optimal policy. The key intuitive trade-off in this case is between transparency and insurance: On one hand, too much noise ($k = \infty$) would lead to runs on all banks when $E[\theta] < p$, which cannot be optimal. On the other hand, too much transparency ($k \simeq 0$) implies that weak banks with $\theta < p$ face a run with probability close to 1. Since depositors have a coordination problem and run on banks too frequently, this scenario also generates a deadweight loss. This is reminiscent of an informational Hirshleifer (1971) effect. An optimal (intermediate) level of noise prevents a systemic crises, but allows weak banks to survive with positive probability, since they can get lucky and draw a high enough s to prevent a run. In general, the resulting optimal policy is quite fine-tuned to the parameters of the model, especially to prior beliefs.

In the remainder of this Section, we characterize the additional effects that a policymaker must take into account when inside information is endogenous, and potentially responds to stress test design. In other words, we seek a characterization of stress tests that is robust to the Lucas critique.

3.4 A minimum standard of transparency

Based on the logic of Proposition 1, a first simple observation is that information becomes fragile as stress tests improve beyond a critical threshold k_0:

Corollary 1. *There exists a threshold $k_0 \in (0, \infty)$ such that the least transparent equilibrium is discontinuous at k_0: $\theta^\star_k = \bar{\theta}$ for all $k > k_0$, while $\theta^\star_{k_0} < \bar{\theta}$.*

To translate this into an implication for optimal policy, note that when outside information is noisy ($k > k_0$), all strong banks $\theta \geq p$ disclose, and all weak banks $\theta < p$ suffer a run. When $k = k_0$, by contrast, a discrete set of strong banks stays quiet in equilibrium. By revealed preference, these strong, quiet banks are weakly better off than under transparency. Moreover, the silence of the strong strictly improves the utility of the weak, who can now avoid a run with positive probability (if they obtain a sufficiently impressive s). Therefore it can never be optimal to introduce noise above the critical level k_0 into stress tests:

Proposition 3. *Any optimal policy sets $k \leq k_0$. Moreover, welfare increases discontinuously when k crosses k_0 from above.*

The intuition is clear. As a consequence of the Hirshleifer effect, opacity by strong banks provides implicit insurance for weak ones. Here, some silence at the top of the quality distribution, which is achieved when $k = k_0$, must improve welfare relative to an unraveling scenario where all good banks disclose. The discontinuity of welfare around k_0 is due to reverse unraveling, and follows from Proposition 1. Moreover, it is easy to show that under certain circumstances, for example if the prior belief $E[\theta]$ is relatively optimistic and the costs $c(\theta)$ of inside disclosure are low, a policymaker who ignores inside disclosures would optimally set $k > k_0$. The minimum standard of transparency is therefore new relative to the literature that treats inside information as exogenous. Once inside information responds, a 'naive' choice $k > k_0$ would trigger unraveling, and disclosures by all strong banks $\theta \geq p$, which completely cancel out the insurance benefits of imperfect transparency.

3.5 Optimal transparency with inside information

Despite Proposition 3, it is nonetheless possible that even a policymaker who considers inside information to be exogenous (or entirely absent) would choose to satisfy the minimum standard of transparency, $k \leq k_0$ (for example, if prior beliefs are sufficiently pessimistic). In that case, Proposition 3 does not tell us whether stress tests should become more or less transparent once we acknowledge that inside information responds endogenously.

To study this formally, consider a naive policymaker who takes the bank's disclosure threshold $\theta_n^\star$ as exogenously given, and maximizes welfare $W(\theta_n^\star; k)$, considering only the partial derivative with respect to k. A sophisticated, fully optimizing policymaker instead maximizes $W(\theta_k^\star; k)$, considering the total derivative. For the most direct comparison, we endow the naive policymaker with consistent beliefs. That is, at its optimal policy choice $k^\star$, the naive policymaker's conjectured disclosures are exactly the equilibrium, $\theta_n^\star = \theta_{k^\star}^\star$.[27] When beliefs are correct, we emphasize the naive objective function by writing $W_n(k) \equiv W(\theta_k^\star; k)$. To aid local arguments, we focus on the (interesting) interior case where $\theta_k^\star \in (p, \bar{\theta})$.

A naive policymaker takes into account the trade-off between transparency and insurance discussed above. In addition, a sophisticated policymaker realizes that changes in k will affect the threshold $\theta_k^\star$ and thus trigger a further indirect change in $s_k^\star$. Our analysis in Section 2 suggests that this effect is ambiguous: An improvement in the quality of stress tests ($\downarrow k$) can either crowd in or crowd out disclosures by strong banks. We now show that this distinction is crucial for the optimal transparency of stress tests:

Proposition 4. *At the naive policymaker's optimal choice $k = k^\star$, the marginal effect on welfare of lowering k is positive if there is crowding out ($\frac{\partial \theta_k^\star}{\partial k} > 0$) and negative if there is crowding in ($\frac{\partial \theta_k^\star}{\partial k} < 0$).*

If crowding out is 'persistent' ($\theta_k^\star > \theta_{k^\star}^\star$, $k > k^\star$), and

$$W_n(k^\star) \geq W_n(k) + \left(F(\theta_k^\star) - F(\theta_{k^\star}^\star)\right) \mathbb{E}\left[\theta\left(\Pr\left[s \leq s^\star(\theta_k^\star, k) \mid \theta; k\right] - c\right)\right], \forall k > k^\star \tag{7}$$

then a sophisticated policymaker chooses $k < k^\star$. Conversely, if crowding in is persistent, in the sense that $\theta_{k^\star}^\star < \theta_k^\star$ for all $k > k^\star$ and (7) holds for $k < k^\star$, then a sophisticated policymaker chooses $k > k^\star$.

Proposition 4 first considers the marginal effect on welfare of changing the noise k contained in stress tests relative to the naive policymaker's optimum $k^\star$. It states that welfare can be improved by making stress tests marginally more precise ($\downarrow k$) if and only if this change crowds out disclosures by strong banks ($\downarrow \theta_k^\star$). To understand this result, suppose that there is crowding out. While this change has only a second-order effect on the payoffs

[27] This can be formalized by changing the game to have the policymaker and banks choose simultaneously. Then, the naive policymaker's optimum is simply a Bayes Nash equilibrium of the reduced-form game between policymaker and banks, in which investor's behavior is described by the threshold function $s^\star(\theta_k^\star; k)$.

of those banks at the margin (they are indifferent), it has important consequences for the signal extraction problem of investors. As more strong banks with $\theta > p$ switch to staying quiet, silence becomes better news. This lowers the critical signal $s^\star$ required to avoid a run on a quiet bank. Therefore, the behavioral effect of lowering k in this case is to reduce the probability with which quiet banks experience a run, thereby increasing welfare. Conversely, if there is crowding in, welfare can be improved by making stress tests marginally less precise ($\uparrow k$).

In general even the naive policymaker's problem is not typically concave in k, so that local arguments are insufficient for comparing the *global* solutions to the naive and sophisticated problems.[28] The second part of Proposition 4 provides simple sufficient conditions under which this marginal analysis carries through to determine the optimal choice of stress test. Suppose that crowding out effects are persistent, and consider a sophisticated policymaker who sets $k > k^\star$. The welfare effects of such a policy can be decomposed as follows: First, the policymaker would understand the direct effects of the policy change (i.e. in the absence of any response in equilibrium disclosures). But this is just $W_n(k) - W(k^\star)$, the naive cost. In addition, the sophisticated policymaker understands there will be two further welfare effects of such a policy. First, types in $[\theta^\star_{k^\star}, \theta^\star_k]$ switch to disclosure, thereby minimizing the welfare cost to which they would otherwise be exposed. The second right-hand side term in (7) captures this effect. Second, the crowding out effect imposes a strict negative externality on all non-disclosing banks. Since this final externality is unambiguously negative, a sufficient condition for $k > k^\star$ to be suboptimal is that the naive costs are large enough even when intermediate types protect themselves via disclosure. Reversing the argument, persistent enough crowding in effects can push a policymaker towards noisier stress tests.

As indicated by Lemma 3, persistent crowding out is common when in periods of crisis. For example, when θ, s are jointly normally distributed, persistent crowding out is guaranteed whenever $\mathbb{E}[\theta] < p$ and disclosure costs are not too large. Moreover, under these conditions, (7) is likely to hold: by the Envelope Theorem, it is easy to show that (7) is satisfied locally around $k^\star$. Moreover, it continues to hold at large k. For instance, if k is large enough to ensure full disclosure is the unique equilibrium, the right-hand side of (7) would simply be the full disclosure payoff. But this is strictly worse for the naive policymaker than setting $k = 0$, which is in turn revealed worse than setting $k = k^\star$.

Therefore, in severe financial crises it is likely that a sophisticated policymaker will choose a more transparent stress test than a naive one who ignores the endogenous response of inside

[28] For example, since W_n limits to a constant (full disclosure) payoff as $k \to \infty$, and $W_n(0) > \lim_{k\to\infty} W_n(k)$, W_n cannot be concave. At best W_n may still be quasiconcave, though this depends highly on parameters.

information. The Lucas critique implies a case for greater transparency in financial policy.

4 A general Sender-Receiver model

We now show that the main insights of Section 2 continue to be present in more general Sender-Receiver games with outside information. In particular, we show how our results on fragility of disclosures to outside information can be extended. Additionally, we provide new insights regarding the role of payoffs in predicting when equilibria will feature reverse unraveling, and when they will feature traditional unraveling.

Consider a game between a *Sender (he),* who has the opportunity to disclose verifiable inside information, and a *Receiver (she),* who chooses an action $a \in A \subset \mathbb{R}$ based on these disclosures and outside information. Payoffs depend on this action and on the state of the world $\theta \in \Theta = \{\theta_1, ..., \theta_N\} \subset \mathbb{R}$, where $\theta_N > \theta_{N-1} > ... > \theta_1$. We focus on finite types in this Section to ensure the existence of equilibrium.

Sender's payoff, $v(a)$, depends only on the action taken and is strictly increasing in a. Receiver's payoff $u(a, \theta)$ is log-supermodular in a and θ, so that she optimally chooses higher actions when optimistic about θ.[29] We assume that, when Receiver knows θ with certainty, she has a unique best response denoted $a^\star(\theta) = \arg\max_{a \in A} u(a, \theta)$.

Sender privately observes the state θ, which we refer to as his type, and sends a message $m \in \{\theta, \emptyset\}$. The message $m = \theta$ is available only to type θ, and therefore amounts to full and verifiable disclosure of θ, but it reduces Sender's utility by $c > 0$. The null message $m = \emptyset$, which we refer to as staying quiet, is costless but reveals no verifiable information.

Again, we focus on this simple message space for clarity of exposition. In Online Appendix F.3, we show that similar results obtain in more general message spaces for m, or when costs depend on θ, as long as verifiable messages are more costly than cheap talk, and the cost of sending such messages is not too steep as a function of their informativeness.

In addition to m, Receiver observes an outside signal s drawn from a finite set $S \subset \mathbb{R}$. We write $\mu_0(\theta)$ for the prior distribution of θ and $\pi(s|\theta)$ for the conditional distribution of s given θ, which are common knowledge. We assume that $\mu_0(\theta) > 0$ for all θ.

The timing is as before:

1. Sender privately observes θ, and chooses send a message m.

2. Receiver observes m, as well as the outside signal s.

[29] More precisely, Receiver's optimal action increases whenever her beliefs about θ become more optimistic in the sense of the monotone likelihood ratios (Milgrom, 1981; Athey, 2002).

3. Receiver chooses an action $a \in \{0, 1\}$.[30]

We consider Perfect Bayesian Equilibria in which off the equilibrium path, Receiver places zero probability on type θ' if she observes a signal such that $\pi(s|\theta') = 0$. We call an equilibrium *monotone (increasing)* if the probability of disclosure $Pr[m = \theta|\theta]$ is increasing in the type θ, and strictly increasing for some pair of types. We call an equilibrium *opaque* if nobody discloses and $Pr[m = \theta|\theta] = 0$. Finally, we call an equilibrium *non-monotone* if $Pr[m = \theta|\theta]$ is strictly increasing for some pair of types and strictly decreasing for another. The fact that the worst type θ_1 has a dominant strategy to stay quiet guarantees that these are the only possibilities. Finally, an *unraveling equilibrium* is a special case of monotone equilibrium in which all $\theta > \theta_1$ disclose with probability 1.

4.1 Fragile information

An important feature of the reverse unraveling mechanism is amplification: Due to strategic complementarities, second-order changes in the environment can trigger first-order responses in Sender's equilibrium communication strategy. As before refer to a *revealing path* as a continuous sequence of outside signals s_t, indexed by a parameter $t \in [0, 1]$ and with associated conditional distributions $\pi(s|\theta; t)$, such that s_0 is pure noise and s_1 perfectly reveals Sender's type θ. We show that, regardless of the primitives of the model, there is always a revealing path that induces fragility of information:

Proposition 5. *For any payoffs $\{u, v\}$ and any prior μ_0, assume that c is sufficiently small to ensure that there is an unraveling equilibrium when public signals are pure noise. Then there exists a revealing path and a critical point t_0 such that there is an unraveling equilibrium when Receiver observes s_t for $t \leq t_1$, while full opacity is the unique equilibrium when she observes s_t for $t > t_1$.*

In contrast to our full characterization in the binary response case, this is an existence result, and we have not shown that discontinuities arise along *every* revealing path. However, we have significant degrees of freedom when choosing the path of signal structures $\Pi(t)$. In particular, we show in Online Appendix F.1 that the results continue to go through on an appropriately defined open set of revealing paths. One caveat is that the revealing path must have a violation of full support around $t = t_1$. With full support, there is always an unraveling equilibrium, as discussed in Section 2. In other words, we have no equivalent of the first part of Proposition 1, which states that less transparent equilibria are also fragile

[30] This action is not contractible: Sender and Receiver cannot pre-specify a as a function of m and s. Hart et al. (2017) derive a class of verifiable message games in which equilibria with and without commitment are identical.

with full support, in the general case, although we expect that similar results could be recovered by imposing more structure on preferences and distributions.

The basic intuition of the result is that of reverse unraveling. The path we identify has the property that, at time $t^\star + \epsilon$ and beyond, the highest type of Sender, θ_N, prefers not to disclose in any equilibrium. While the other types would prefer disclosure if all but the lowest type were expected to do so, their marginal preference for disclosure at $t^\star + \epsilon$ is small. Indeed, once θ_N prefers not to disclose in any equilibrium, we show that this infects the optimal disclosure decision of type θ_{N-1}, and subsequently type θ_{N-2}, and so on, until iterated elimination of nonequilibrium strategies yields full opacity as the unique equilibrium at all times beyond $t^\star$.

4.2 Outside information and equilibrium informativeness

In addition to establishing the fragility of information, we consider how access to better outside information affects the quality of the information that Receiver observes in equilibrium. We use the Blackwell (1953) order to rank information structures. A general signal $\tau' \in T'$ is said to be more informative about θ than another signal $\tau \in T$ if τ a 'garbled' version of τ'.[31] In the context of our model, we can use Blackwell's criterion to rank the informativeness of two outside signals s and s'. We can also rank the informativeness of Receiver's equilibrium information set $\{m, s\}$ across different scenarios.[32]

We show that more informative signals can always leave Receiver worse informed in equilibrium:

Proposition 6. *Suppose that, when outside information is s, there is an equilibrium $\mathcal{E}$ in which Sender makes a disclosure $m = \theta$ with strictly positive probability. Then there exists an outside signal s' such that*

- *s' is more informative than s in the sense of Blackwell, and*

[31] Formally, Nature first draws the clean signal τ' and then randomly converts it to the garbled signal τ, so that we can write

$$Pr[\tau|\theta] = \sum_{\tau' \in T'} Pr[\tau'|\theta] g(\tau|\tau')$$

for some conditional distribution $g(\tau|\tau')$. Blackwell's theorem shows that this notion of informativeness is equivalent to requiring that every Bayesian decision-maker weakly prefers to observe realizations of τ' instead of τ.

[32] In any equilibrium $\mathcal{E}$, Receiver observes the signal $\tau = \{s, m\}$, which contains both outside and inside information and has conditional distribution $Pr[\tau|\theta] = \pi(s|\theta) \times Pr[m|\theta]$, where the second factor is determined endogenously by Sender's equilibrium strategy. We consider situations where outside information changes to s' and Sender changes his equilibrium disclosure strategy. The resulting new equilibrium $\mathcal{E}'$ induces an appropriately defined signal $\tau' = \{s', m'\}$. Receiver is less informed in the new equilibrium in the sense of Blackwell if we can write τ' as a garbling of τ

- *In the game where outside information is s', there is an equilibrium $\mathcal{E}'$ in which Receiver is less informed than in $\mathcal{E}$ in the sense of Blackwell.*

Proposition 6 follows from the interaction between outside signals and insiders' incentives to disclose. When outside signals become more informative, they crowd out incentives for voluntary disclosures by Sender. Since Sender is better informed than Receiver, this crowding-out can unambiguously reduce information sharing in the economy. Intuitively, considering a type θ_i of Sender that makes a disclosure with positive probability, we can allow Receiver to observe a compound lottery between s and a signal that reveals θ_i perfectly. If the probability of revelation is large enough, type θ_i strictly stays quiet in a new equilibrium of the game. Receiver now observes garbled information about θ_i in equilibrium, while she observed θ_i perfectly (via m) in the old equilibrium. By carefully choosing the distribution of auxiliary signals, the proof ensures that Receiver's information about other types θ_j, $j \neq i$, also becomes (weakly) worse, so that we obtain an overall Blackwell-deterioration in Receiver's information. For this Proposition, we do not require full support of signals, as discussed in Online Appendix F.2.

4.3 Disclosures and the shape of payoffs

So far, we have shown that specific distributions of outside information have a strong tendency to crowd out inside disclosures. Key to these effects was the fact that incentives to disclose are weakened by precise outside information, particularly for strong types of Sender. We now conduct a complementary exercise. For a *given* distribution of outside information, we study whether strong types of Sender have an incentive to stay quiet in equilibrium, thus giving rise to the effects we have emphasized. As we noted above, the best types of Sender always have the strongest incentive to disclose in the absence of outside information. In the presence of outside information, we show that the shape of Sender's payoff plays a crucial role.

Concavity and convexity in a model with 'virtual types'

Consider the common special case where Receiver's optimal action takes the form

$$\arg\max E_\mu[u(a,\theta)] = E_\mu\left[X\left(\theta\right)\right],$$

for some increasing function $X(\theta)$, and where Sender's utility is simply $v\left(a\right) = a$. Such preferences are natural in standard seller-buyer interactions where a denotes the willingness-to-pay of a buyer (or indeed of a mass of buyers in a competitive market) for an indivisible item that gives her utility $X(\theta)$, or in settings with quadratic Receiver utility. This case permits a useful re-interpretation of payoffs in terms of virtual types. If Sender stays quiet and has true type $\theta = \theta_i$, his expected payoff is

$$\sum_{s \in S} \pi(s|\theta_i) E[X(\theta)|s, m = \emptyset] = \sum_{j=1}^{N} q_{ij} X(\theta_j),$$

where $q_{ij} = E[Pr[\theta_j|s, m = \emptyset]|\theta_i]$ is the *expected* probability mass that Receiver places on type θ_j. Payoffs in the absence of disclosure are therefore equivalent to a game in which Sender draws a virtual type θ_j according to the conditional distribution q_{ij}. Note that q_{ij} is indeed a probability distribution since $\sum_j q_{ij} = 1$ for all i. We write $Q_{ij} = \sum_{k \leq j} q_{ik}$ for the cumulative distribution of virtual types.

We assume that outside signals satisfy the strict Monotone Likelihood Ratio Property (MLRP; defined as in Milgrom, 1981): For $\theta' > \theta$ and $s' > s$, we impose that

$$\pi(s'|\theta')\pi(s|\theta) > \pi(s'|\theta)\pi(s|\theta').$$

We further assume that neighboring types share signals: For each i, there is an s such that $\pi(s|\theta_i) > 0$ and $\pi(s|\theta_{i-1}) > 0$ (clearly, any signal distribution with full support satisfies this restriction).

We write $\Delta X_i = X(\theta_{i+1}) - X(\theta_i)$ for the increment in Receiver's action if she learns that Sender's type increases from θ_i to the next-best type θ_{i+1}, and let

$$\mathcal{N}(\theta) \equiv X(\theta) - \sum_{s \in S} \pi(s|\theta_i) E[X(\theta)|s, m = \emptyset]$$

be type θ's marginal payoff from disclosure (net of costs). Integrating by parts, we can re-write this net payoff as:

$$\mathcal{N}(\theta_i) = \sum_{j=1}^{i-1} \Delta X_j Q_{ij} - \sum_{j=i}^{N-1} \Delta X_j (1 - Q_{ij}). \tag{8}$$

Equation (8) expresses the net payoff from disclosure as the sum of two components. The first term is the downside risk that Sender takes by staying quiet; with probability Q_{ij} his virtual type is below θ_j for $j < i$, and the associated incremental loss is ΔX_j. The second term is the upside risk; with probability $1 - Q_{ij}$, Sender's virtual type is above θ_j for $j > i$, and the associated incremental gain is again ΔX_j. If the downside exceeds the upside by more than c, Sender prefers to disclose. Virtual types are useful because they inherit the ordering of signals: In the Proof of Proposition 7 below, we show that type θ_{i+1} draws better virtual types – in the sense of first-order stochastic dominance – than type θ_i. Thus, the probability weights Q_{ij} on downside risk decrease as Sender's true type i improves, while the weights $1 - Q_{ij}$ on upside risk increase.

To assess the relevance of the shape of the payoff function $X(\theta)$ to disclosures, we define the following measures of concavity and convexity:

$$\text{Concavity} \equiv \min_i \frac{\Delta X_i}{\Delta X_{i+1}}$$
$$\text{Convexity} \equiv \min_i \frac{\Delta X_{i+1}}{\Delta X_i}$$

When Concavity > 1, payoffs are concave in the sense that the marginal value of being perceived as a better type diminishes as Sender's type improves. Similarly, when Convexity > 1, the marginal value of being perceived as a better type increases as Sender's type improves, and payoffs are convex. We can relate these parameters to disclosure strategies in equilibrium:

Proposition 7. *If the Concavity of payoffs is sufficiently large, then if disclosure costs are not too small, all equilibria are non-monotone or fully opaque. Conversely, if Convexity is sufficiently large, then there are no non-monotone equilibria.*

In other words, when payoffs are concave enough, the strongest types of Sender must stay quiet in any equilibrium, whenever disclosure costs c exceed a threshold c_0 (chosen to rule out an equilibrium in which only type θ_1 stays quiet).[33] Conversely, when payoffs are convex enough, the strongest types of Sender always disclose in equilibrium (unless costs are prohibitive so that nobody wants to disclose). The result uses our characterization (8) of the net benefit of disclosure: When payoffs are sufficiently concave, incentives to disclose come mainly from the downside increments $\Delta X_1, ..., \Delta X_{i-1}$. Since the probability weights on these increments fall as Sender's true type improves, strong types have weak incentives to disclose. Then, the logic of reverse unraveling leads to a non-monotone or fully opaque equilibrium, as in the binary example of Section 2. When payoffs are sufficiently convex, by contrast, we can rule out non-monotone equilibria as follows: Let θ_n be the highest quiet type in a non-monotone equilibrium, and $\theta_d < \theta_n$ a disclosing type below him. We show that type θ_n has strictly stronger incentives to disclose than θ_d because of the large, convex, utility he gains by raising his virtual type to θ_n. The proof constructs uniform bounds on Concavity and Convexity that ensure these properties for all possible (pure or mixed) strategy profiles; the bounds depend on the prior signal distribution, but not on equilibrium play.

[33] Note, however, that this threshold c_0 is larger than it would be in the absence of outside information. Hence, we can focus on the case where $c > c_0$, but still assume that costs are 'small' in the sense unraveling is an equilibrium outcome in the absence of outside information.

A simple example clarifies the logic of Proposition 7.

Example. Consider the 'virtual type' case with three types $\theta \in \{\theta_1, \theta_2, \theta_3\}$, five outside signals $s \in \{s_0, ..., s_4\}$, and a uniform prior $\mu_0(\theta_i) = 1/3$. Each type draws the outside signal to the left of his type with probability $\pi(s_{i-1}|\theta_i) = p$, that to the right with probability $\pi(s_{i+1}|\theta_i) = r$, and the signal matching his type with the remaining probability $\pi(s_i|\theta_i) = q = 1 - p - r$.

We have Concavity $= \frac{\Delta X_1}{\Delta X_2}$. Define the maximal punishment for silence $\mathcal{M}(\theta)$ as Sender's net payoff from disclosure when Receiver adopts the most pessimistic feasible beliefs following $m = \emptyset$ (see Appendix F). We have $\mathcal{M}(\theta_2) = \Delta X_1(p + q)$ for the middle type and $\mathcal{M}(\theta_3) = \Delta X_1 p + \Delta X_2(p + q)$ for the top type. We have $\mathcal{M}(\theta_3) < \mathcal{M}(\theta_2)$ when Concavity $> 1 + \frac{p}{q}$, that is, when the concavity of payoffs is large relative to the likelihood ratio of left-tail outside signals to intermediate ones. Under this condition, the high type θ_3 has the strongest incentives to deviate from an unraveling equilibrium, and it is easy to see that such an equilibrium exists if and only if $c \leq \mathcal{M}(\theta_3) \equiv c_0$. Whenever $c > c_0$, the top type must therefore stay quiet, and since the bottom type also has a dominant strategy to stay quiet, resulting equilibria must be non-monotonic of fully opaque, in line with the first part of Proposition 7.

Moreover, we have Convexity $= \frac{1}{\text{Concavity}} = \frac{\Delta X_2}{\Delta X_1}$. Consider a non-monotone equilibrium in pure strategies, where only the middle type θ_2 discloses. In this equilibrium, Receiver is certain that $\theta = \theta_1$ when the outside signal is $s \leq s_1$, and equally certain that $\theta = \theta_3$ when $s \geq s_3$. When $s = s_2$, she places probability $\frac{r}{p+r}$ on type θ_1 and complementary probability $\frac{p}{p+r}$ on type θ_3. The implied distribution of virtual types has $Q_{21} = p + q\frac{r}{p+r} = Q_{22}$ and $Q_{31} = p\frac{r}{p+r} = Q_{32}$. For optimality, the middle type must prefer to disclose and the top type must prefer to stay quiet: $\mathcal{N}(\theta_2) \geq c \geq \mathcal{N}(\theta_3)$. Thus a non-monotone equilibrium exists for some c if and only if $\mathcal{N}(\theta_2) \geq \mathcal{N}(\theta_3)$. Substituting into (8) and rearranging, this is equivalent to Convexity $\leq \frac{\lambda}{1-\lambda}$, where $\lambda = Q_{21} - Q_{31}$ is the downside risk perceived by the top type relative to the middle type. Conversely, when Convexity $> \frac{\lambda}{1-\lambda}$, there is no non-monotone equilibrium.■

The example further helps us to understand disclosures when parameters fall between the bounds we have established in Proposition 7, that is, when payoffs are neither very convex nor very concave. It turns out that the *skewness* of outside information becomes critical in this case. For instance, when payoffs are linear (Concavity = 1), non-monotone equilibria exist in the example if and only if $\lambda = p(1 - \frac{r}{p+r}) + q\frac{r}{p+r} \geq \frac{1}{2}$. With a symmetric outside signal distribution ($p = r$) this is impossible unless signals are perfectly revealing. When outside signals are precise with $q > \frac{1}{2}$, we have a non-monotone equilibrium if and only if outside information is right-skewed, with $\frac{r}{p}$ sufficiently large. Intuitively, a right skew increases the advantage of top types over mediocre types, since the outside signals drawn by mediocre types are now interpreted chiefly as having come from low types. As a result, payoffs must now be strictly convex to rule out non-monotone disclosures.

4.4 An Application: Disclosures when issuing debt and equity

We now model corporate disclosures by a firm wishing to raise funds from investors. We consider two modes of finance: issuing bonds and issuing shares. Since bonds give investors a concave claim and shares give a convex one, Proposition 7 suggests that incentives to disclose will differ between the two scenarios.

Consider a firm whose profits are $\theta \sim U[0,1]$. The firm wishes to maximize the amount of money it raises by selling a given security to risk-neutral financial investors. As usual, the firm privately observes θ and can verifiably disclose it ($m = \theta$) at a cost c. Investors subsequently observe an outside signal $s = \theta + k\epsilon$, where $\epsilon \sim U[-1,1]$. For simplicity, we assume that the noise parameter $k > 1/2$, so that any two types have some signals in common.

If the firm sells shares, then the payoff to buyers of shares is the convex claim $\max\{\theta - d, 0\}$, where d denotes the face value of any existing debt. The firm's payoff is the market price of shares, which is given by investors' willingness to pay given inside and outside information:

$$p(m,s) = E[\max\{\theta - d, 0\}|m,s].$$

Low-quality firms with $\theta \leq d$ have a dominant strategy to stay quiet, since disclosing θ would yield $p = 0$. For firms with $\theta > d$, the net payoff from disclosure is the expected gain in share prices

$$\mathcal{N}(\theta) = (\theta - d) - \frac{1}{2k}\int\limits_{\theta-k}^{\theta+k} p(\emptyset, s)ds.$$

This net payoff is strictly increasing in θ: The first term (the payoff from full disclosure) increases with θ at rate 1. The second term increases at rate

$$\frac{1}{2k}\left[p(\emptyset, \theta + k) - p(\emptyset, \theta - k)\right]$$

since increasing θ shifts probability mass from low signals around $\theta - k$ to high signals around $\theta + k$. Under the assumption that signals are not too precise ($k > 1/2$), this rate is always less than one. Since the net payoff from disclosure is increasing in θ, all equilibria must have a cutoff property, where only firms with high quality $\theta \geq \theta^\star$ make disclosures, with $\theta^\star > d$.

If the firm sells bonds, by contrast, the payoff to buyers is the concave claim $\min\{d, \theta\}$. The firm's payoff is the market price of bonds

$$q(m, s) = E[\min\{d, \theta\}|m, s].$$

The net payoff from disclosure is now

$$\mathcal{N}(\theta) = \min\{d, \theta\} - \frac{1}{2k} \int\limits_{\theta-k}^{\theta+k} q(\emptyset, s) ds.$$

It is easy to see that incentives to disclose are strongest at the kink of the payoff function where $\theta = d$. For lower-quality firms with $\theta < d$, the first term increases at rate 1, while the second term increases at rate $\frac{1}{2k} [q(\emptyset, \theta + k) - q(\emptyset, \theta - k)] < 1$. For high-quality firms with $\theta > d$, the first term is fixed, while the second term is still increasing. Therefore, the net payoff from disclosure has a peak at $\theta = d$. It follows that all equilibria must have interval strategies, where only firms with intermediate quality $\theta \in (\theta_L^\star, \theta_H^\star]$ make disclosures, with $\theta_L^\star \leq d \leq \theta_H^\star$.[34]

The empirical predictions of this model are that disclosures come mainly from high-quality firms if they are selling shares, and mainly from intermediate-quality firms if they are selling bonds. Moreover, since firm quality and outside information are positively correlated, we predict that disclosures come mainly from firms with favorable subsequent realizations of outside information (e.g. optimistic analyst opinions) if selling shares, and mainly from firms with intermediate signals (e.g. mediocre credit ratings) if selling bonds. These predictions need to be qualified by allowing for sample selection: We have assumed that the security sold by the firm is exogenously determined and independent of its quality θ. In the classic 'pecking order' theory of Myers and Majluf (1984), debt is selected by high-quality firms as a signal. Daley et al. (2017) study a related model to ours, where debt issuance serves as an (inside) signal of quality in a model with (outside) credit ratings and two possible realizations of firm quality. Our model complements this literature by showing that – conditional on selecting debt, or in situations where debt is unambiguously more attractive (e.g. due to tax advantages) – disclosures tend to be made by firms of intermediate quality.

[34] For both bonds and shares, it is straightforward to show that the relevant thresholds exist, although they may not be unique, and are interior for a range of parameters.

5 Conclusion

In this paper, we have studied the determinants of a privately informed Sender's incentives to produce and disclose verifiable evidence to a decision-maker, who also has access to other outside sources of information. Motivated by our leading example of bank disclosures in financial markets, we also address the policy question of how much information (in the form of a stress test) a policymaker should provide to markets, when banks can also voluntarily disclose information about their asset quality.

Our main finding is that the presence of outside information can drastically alter insiders' incentives to disclose. Indeed, we show that the presence of outside information generates a stark contrast in predictions compared with the classic literature on verifiable disclosures. In contrast to the usual unraveling results, we show that the presence of outside information can generate *reverse unraveling*, where silence by 'high quality' Senders becomes contagious and incentivizes yet more silence by those with lower quality. Indeed, in a binary action setting, when outside information is sufficiently precise all equilibria must take this form. An important consequence is that, as outside information becomes precise enough to admit reverse unraveling, there is a discontinuous contraction in equilibrium disclosures. We identify this as a *fragility of disclosures* to outside information. Away from this threshold, the interaction between inside and outside information is more subtle – they may be local substitutes or complements.

In a model of financial crises, we consider the implications of our results for information policy. We identify an informational externality for which policymakers should account when conducting stress testing policy: Stress tests affect banks' equilibrium disclosures. Our results make a new case for more informative stress-testing during periods of financial crisis. In the context of the recent literature on stress-testing, which identifies the benefits of pooling information about banks in a crisis, this finding may seem counter-intuitive. However, when bank disclosures are voluntary, we argue that this same reasoning implies a need for greater transparency: by increasing the informativeness of stress tests, policymakers minimizes the costs of separation by disincentivizing high quality banks from disclosing.

An analysis of a more general model highlights that the effect of outside on inside information depends on the shape of Sender's payoffs. Reverse unraveling can gain traction, and disclosure tend to come only from mediocre types, when Sender's utility is concave in his perceived quality. By contrast, when utility is convex, only a set of the best types make disclosures, and the impact of outside on inside information is muted. An interesting application is to corporate disclosures: If a firm finances itself with debt, which is a concave claim, disclosure strategies are non-monotone in firm profitability; if it is equity-financed, they are monotone and disclosures come from the most profitable firms.

References

Acharya, V. V., P. DeMarzo, and I. Kremer (2011). Endogenous information flows and the clustering of announcements. *American Economic Review 101*(7), 2955–2979.

Akerlof, G. E. (1970). The market for 'lemons': Quality uncertainty and the market mechanism. *Quarterly Journal of Economics 84*(3), 488–500.

Amador, M. and P.-O. Weill (2010). Learning from prices: Public communication and welfare. *Journal of Political Economy 118*(5), 866–907.

Angeletos, G. and A. Pavan (2007). Efficient use of information and social value of information. *Econometrica 75*(4), 1103–1142.

Angeletos, G.-M., C. Hellwig, and A. Pavan (2006). Signaling in a global game: Coordination and policy traps. *Journal of Political Economy 114*(3), 452–484.

Angeletos, G.-M. and A. Pavan (2013). Selection-free predictions in global games with endogenous information and multiple equilibria. *Theoretical Economics 8*(3), 883–938.

Athey, S. (2002). Monotone comparative statics under uncertainty. *Quarterly Journal of Economics 117*(1), 187–223.

Bank of England (2013). Banks' disclosure and financial stability. *Bank of England Quarterly Bulletin 2013 Q4*, 326–335.

Bertomeu, J. and D. Cianciaruso (2015). Verifiable disclosure. *Economic Theory*, 1–34.

Blackwell, D. (1953). Equivalent comparison of experiments. *Annals of Mathematical Statistics 24*(2), 265–272.

Boissay, F., F. Collard, and F. Smets (2015). Booms and banking crises. *Journal of Political Economy (forthcoming).*

Bouvard, M., P. Chaigneau, and A. de Motta (2015). Transparency in the financial system: Rollover risk and crises. *Journal of Finance 70*(4), 1805–1837.

Cho, I.-K. and D. M. Kreps (1987). Signaling games and stable equilibria. *Quarterly Journal of Economics 102*(2), 179–221.

Daley, B. and B. Green (2014). Market signaling with grades. *Journal of Economic Theory 151*, 114–145.

Daley, B., B. Green, and V. Vanasco (2017). Securitization, ratings, and credit supply. *Mimeo.*

Diamond, D. W. and P. H. Dybvig (1983). Bank runs, deposit insurance, and liquidity. *Journal of Political Economy 91*(3), 401–19.

Diamond, D. W. and R. E. Verrecchia (1991). Disclosure, liquidity, and the cost of capital. *Journal of Finance 46*(4), 1325–1359.

Dye, R. A. (1983). Communication and post-decision information. *Journal of Accounting Research*, 514–533.

Dye, R. A. (1985). Disclosure of nonproprietary information. *Journal of Accounting Research 23*(1), 123–145.

Faria-e-Castro, M., J. Martinez, and T. Philippon (2016). Runs versus lemons: information disclosure and fiscal capacity. *The Review of Economic Studies*.

Farrell, J. (1993). Meaning and credibility in cheap-talk games. *Games and Economic Behavior 5*(4), 514–531.

Federal Reserve (2017). Dodd-frank act stress test 2017: Supervisory stress test methodology and results. Technical report, Board of Governors of the Federal Reserve System.

Feltovich, N., R. Harbaugh, and T. To (2002). Too cool for school? Signalling and countersignalling. *RAND Journal of Economics 33*(4), 630–649.

Gigler, F. and T. Hemmer (1998). On the frequency, quality, and informational role of mandatory financial reports. *Journal of Accounting Research* (36), 117–147.

Giovannoni, F. and D. J. Seidmann (2007). Secrecy, two-sided bias and the value of evidence. *Games and Economic Behavior 59*(2), 296–315.

Goldstein, I. and Y. Leitner (2017). Stress tests and information disclosure. *Mimeo.*

Goldstein, I. and A. Pauzner (2005). Demand–deposit contracts and the probability of bank runs. *Journal of Finance 60*(3), 1293–1327.

Goldstein, I. and H. Sapra (2014). Should banks' stress test results be disclosed? an analysis of the costs and benefits. *Foundations and Trends in Finance 8*(1), 1–54.

Gorton, G. (2008). The panic of 2007. NBER Working Paper 14358, National Bureau of Economic Research.

Grossman, S. J. (1981). The informational role of warranties and private disclosure about product quality. *Journal of Law and Economics 24*(3), 461–483.

Grossman, S. J. and O. D. Hart (1980). Disclosure laws and takeover bids. *Journal of Finance 35*(2), 323–334.

Hagenbach, J., F. Koessler, and E. Perez-Richet (2014). Certifiable pre-play communication: Full disclosure. *Econometrica 83*(3), 1093–1131.

Hart, S., I. Kremer, and M. Perry (2017). Evidence games: Truth and commitment. *American Economic Review 107*(3), 690–713.

Heider, F., M. Hoerova, and C. Holthausen (2015). Liquidity hoarding and interbank market rates: The role of counterparty risk. *Journal of Financial Economics (forthcoming).*

Hellwig, C. (2002). Public information, private information, and the multiplicity of equilibria in coordination games. *Journal of Economic Theory 107*(2), 191–222.

Hirshleifer, J. (1971). The private and social value of information and the reward to inventive activity. *American Economic Review 61*(4), 561–574.

Inostroza, N. and A. Pavan (2017). Persuasion in global games with application to stress testing. *Mimeo.*

Jin, G. Z. and P. Leslie (2003). The effect of information on product quality: Evidence from restaurant hygiene grade cards. *The Quarterly Journal of Economics 118*(2), 409–451.

Jovanovic, B. (1982). Truthful disclosure of information. *Bell Journal of Economics 13*(1), 36–44.

Kamenica, E. and M. Gentzkow (2011). Bayesian persuasion. *American Economic Review 101*(6), 2590–2615.

Leuz, C. and P. Wysocki (2016). The economics of disclosure and financial reporting regulation: Evidence and suggestions for future research. *Journal of Accounting Research 54*(2), 525–622.

Mankiw, G. (1986). The allocation of credit and financial collapse. *Quarterly Journal of Economics 101*(3), 455–470.

Mathios, A. D. (2000). The impact of mandatory disclosure laws on product choices: An analysis of the salad dressing market. *Journal of Law and Economics 43*(2), 651–678.

Milgrom, P. and J. Roberts (1986). Relying on the information of interested parties. *The RAND Journal of Economics 17*(1), 18–32.

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

Mishkin, F. S. (1990). Asymmetric information and financial crises: A historical perspective. NBER Working Paper 3400, National Bureau of Economic Research.

Morris, S. and H. S. Shin (2000). Rethinking multiple equilibria in macroeconomic modeling. *NBER Macroeconomics Annual 15*, 139–182.

Morris, S. and H. S. Shin (2002). Social value of public information. *American Economic Review 92*(5), 1521–1534.

Myers, S. C. and N. S. Majluf (1984). Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics 13*(2), 187–221.

Orlov, D., P. Zryumov, and A. Skrzypacz (2017). Design of macro-prudential stress tests. *Mimeo.*

Philippon, T. and V. Skreta (2012). Optimal interventions in markets with adverse selection. *American Economic Review 102*(1), 1–28.

Schelling, T. (1978). *Micromotives and Macrobehavior.* WW Norton & Company.

Seidmann, D. J. and E. Winter (1997). Strategic information transmission with verifiable messages. *Econometrica 65*(1), 163–169.

Shahhosseini, M. (2016). The unintended consequences of bank stress tests. *Mimeo.*

Shin, H. S. (1994). News management and the value of firms. *RAND Journal of Economics 25*(1), 58–71.

Shin, H. S. (2003). Disclosures and asset returns. *Econometrica 71*(1), 105–133.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics 50*(3), 665–690.

Tirole, J. (2012). Overcoming adverse selection: How public intervention can restore market functioning. *American Economic Review 102*(1), 29–59.

Verrecchia, R. E. (1983). Discretionary disclosure. *Journal of Accounting and Economics 5*, 179–194.

Vives, X. (1997). Learning from others: A welfare analysis. *Games and Economic Behavior 20*(2), 177–200.

A Proofs for Section 2

Throughout this Appendix, we write $\sigma(\theta) = Pr[m = \theta|\theta]$ for (potentially mixed) disclosure strategies, and $\sigma = \{\sigma(\theta)\}_{\theta \in [\underline{\theta}, \bar{\theta}]}$ for strategy profiles.

Lemma 1

We first show that in any equilibrium, Receiver's best response takes a threshold form so that Receiver plays $a = 1$ if $m = \emptyset$ and $s > s^\star$, and $a = 0$ if $m = \emptyset$ and $s < s^\star$.

Indeed, take any equilibrium with disclosure strategy σ. We can define the intermediate belief

$$F_\emptyset(\theta') = Pr_\sigma[\theta \leq \theta'|m = \emptyset]$$

which denotes the distribution of θ given the null message alone. Consider Receiver's response in the event $\{m = \emptyset, s\}$. Recall that $\hat{s} = \sup \cup_{\theta < p} S(\theta)$. If $s > \hat{s}$ then Receiver strictly prefers $a = 1$. If $s \leq \hat{s}$, then $s \in S(\theta)$ for some $\theta < p$, and the event $\{m = \emptyset, s\}$ is on the equilibrium path because $\theta < p$ have a strictly dominant strategy to play $m = \emptyset$. Since m and s are independent conditional on θ, Receiver's posterior beliefs given $\{m = \emptyset, s\}$ are formed by updating the prior $F_\emptyset$ using the outside signal s and Bayes' rule. By MLRP and Proposition 1 of Milgrom (1981), the expected value $\mathbb{E}[\theta|m = \emptyset, s]$ is strictly increasing in s. Now we can define the desired $s^\star$ as the lowest $s \leq \hat{s}$ satisfying $\mathbb{E}[\theta|m = \emptyset, s] \geq p$, or if this is impossible, as $s^\star = \hat{s}$.

Finally, we show that Sender plays a threshold strategy. Take any equilibrium with disclosure strategy σ with associated critical signal $s^\star$ defined in the previous paragraph. Let $\theta^\star$ be the lowest $\theta \geq p$ such that $H(s^\star \mid \theta) - c(\theta) \geq 0$, or if this is impossible, let $\theta^\star = p$. Given our single crossing condition (1), we now have three cases, each of which satisfies the claim in the Lemma: First, if $\theta^\star \in (p, \bar{\theta})$ then by continuity, $H(s^\star \mid \theta^\star) - c(\theta^\star) = 0$, and Sender strictly prefers to stay quiet for $\theta \in (p, \theta^\star)$ and strictly prefers to disclose for $\theta > \theta^\star$. Second, if $\theta^\star = p$ then Sender strictly prefers to disclose for all $\theta > p$. Third, if $\theta^\star = \bar{\theta}$ then Sender strictly prefers to stay quiet for all $\theta > p$.

Lemma 2

Suppose that $BR(\theta^\star) \in (p, \bar{\theta})$ for $\theta^\star \in (\bar{\theta} - \delta, \bar{\theta})$. Then by continuity of Sender's expected utility, it satisfies Sender's indifference condition $H(s^\star(\theta^\star)|\theta) - c(\theta)|_{\theta = BR(\theta^\star)} = 0$. Moreover, since $c(\theta) \in (0, 1)$, we know that $s^\star(\theta^\star) \in \text{int}(S(\theta))$ for $\theta = BR(\theta^\star)$, implying that it satisfies

Receiver's indifference condition $\mathbb{E}[\theta|\theta \notin (p, \theta^\star), s^\star] = p$, or equivalently

$$\int\limits_{\theta \notin (p, \theta^\star)} (p - \theta) h(s^\star|\theta) dF = 0$$

We can apply the implicit function theorem to both indifference conditions to get

$$\frac{dBR}{d\theta^\star} = \left.\frac{h(s^\star|\theta)}{c'(\theta) - H_\theta(s^\star|\theta)}\right|_{\theta = BR(\theta^\star)} \times \frac{ds^\star}{d\theta^\star} \tag{9}$$

$$\frac{ds^\star}{d\theta^\star} = \frac{(\theta^\star - p) h(s^\star|\theta^\star) f(\theta^\star)}{\int_{\theta \notin (p, \theta^\star)} (p - \theta) h_s(s^\star|\theta) dF + (p - \underline{\theta}(s^\star)) h(s^\star|\underline{\theta}(s^\star)) f(\underline{\theta}(s^\star)) \frac{d\underline{\theta}(s^\star)}{ds}} \tag{10}$$

where $\underline{\theta}(s) = \inf\{\theta | s \in S(\theta)\}$. Our single crossing condition (1) implies that $c'(\theta) > H_\theta(s|\theta)$ at the crossing point $\theta = BR(\theta^\star)$. Moreover, $h(s^\star|BR(\theta^\star)) > 0$, so that the first term in (9) is a positive constant. We still need to show that $\lim_{\theta^\star \uparrow \bar{\theta}} \frac{ds^\star}{d\theta^\star} = +\infty$. All limits in the rest of the proof are taken as $\theta^\star \uparrow \bar{\theta}$.

Let $\bar{BR} = \lim BR(\theta^\star)$. We know that $c(\bar{BR}) = H(\hat{s}|\bar{BR})$, which implies $1 > H(\hat{s}|\bar{BR}) \geq H(\hat{s}|\bar{\theta})$, where the second inequality follows from first-order stochastic dominance (implied by MLRP). Thus we know that $h(\hat{s}|\bar{\theta}) > 0$, and therefore the numerator in (10) converges to a positive constant. We finish by showing that the denominator converges to zero.

We know that $\lim s^\star = \hat{s} = \sup \cup_{\theta < p} S(\theta)$. To see this, note that if $\lim s^\star < \hat{s}$, then Receiver would place strictly positive probability mass on types $\theta < p$ conditional on observing $s^\star$, but near-zero mass on $\theta \geq p$, thus violating Receiver's indifference condition. Moreover, if $\lim s^\star > \hat{s}$, then Receiver would strictly prefer $a = 1$, again violating indifference. As a result, $\lim s^\star = \hat{s}$, which directly implies that $\lim \underline{\theta}(s^\star) = p$. The second term in the denominator of (10) therefore converges to zero. The integral in the denominator is

$$\int\limits_{\underline{\theta}(s)}^{p} (p - \theta) h_s(s^\star|\theta) dF + \int\limits_{\theta^\star}^{\bar{\theta}} (p - \theta) h_s(s^\star|\theta) dF$$

and also converges to zero given that the derivative h_s is bounded, which completes the proof.

Proposition 1

Take any revealing path, write $BR^t(\theta)$ for the best response function in (3) induced by outside signal s_t, and θ^t_{min}, θ^t_{max} for the least and most transparent equilibria given s_t. For this proof we say that BR^t is *flat at the top* if $\exists B^t < \bar{\theta}$ such that $BR^t(\theta) = \bar{\theta}\ \forall \theta \geq B^t$.

Let $t_0 = \inf\{t : \theta^t_{min} < \bar{\theta}\}$ and $t_1 = \inf\{t : \theta^t_{max} < \bar{\theta}\}$. We know that $\bar{\theta}$ is the unique equilibrium in a neighborhood around $t = 0$, so that $t_0, t_1 > 0$. Moreover, since an equilibrium without any disclosure ($\theta^\star = p$) exists for $t = 1$, we know by continuity of $BR^t(\theta^\star)$ that $t_0 < 1$. (We can have $t_1 = 1$, however, for example in the case with full support.)

To establish the (left-)discontinuity at t_0 we argue that $\theta^{t_0}_{min} < \bar{\theta}$ by contrapositive. Suppose that $\theta^{t_0}_{min} = \bar{\theta}$, so that $BR^{t_0}(\theta) > \theta$ for all $\theta \in [p, \bar{\theta})$. If BR^{t_0} is flat at the top, then $BR^{t_0+\epsilon}$ is also flat at the top for small ϵ. It follows by continuity that $BR^{t_0+\epsilon}(\theta) > \theta$ for all $\theta < \bar{\theta}$, implying $\theta^{t_0+\epsilon}_{min} = \bar{\theta}$, contradicting the definition of t_0 as an infimum. If BR^{t_0} is not flat at the top, then it is interior in a neighborhood of $\bar{\theta}$, and so by Lemma 2, we can find a $b < \bar{\theta}$ such that $BR^{t_0}(\theta) < \theta \ \forall \theta \geq b$. For small ϵ, $BR^{t_0-\epsilon}(\theta) < b$, and since the best response is non-decreasing, $BR^{t_0-\epsilon}(\theta) \in [p, b]$ for all $\theta \in [p, b]$. Then by Brouwer's fixed point theorem, there exists an equilibrium $\theta^\star < \bar{\theta}$ so that $\theta^{t_0-\epsilon}_{min} < \bar{\theta}$ for small ϵ, again contradicting the definition of t_0.

To establish the (right-) discontinuity at t_1 when $t_1 < 1$, note for small ϵ, $BR^{t_1+\epsilon}(\theta) \in (p, \bar{\theta})$ in a neighborhood of $\bar{\theta}$ (otherwise, an unraveling equilibrium exists at $t_1 + \epsilon$, a contradiction). Now defining $L(x, y) = BR^{t_1+y}(\bar{\theta} - x) - (\bar{\theta} - x)$, we know that since BR is interior, $L(x, y)$ is continuously differentiable for small $x, y > 0$. Moreover, we can see that *(i)* $L(0, 0) = 0$; since otherwise an unraveling equilibrium does not exist for $t_1 - \epsilon$, *(ii)* $L_y(0, 0) < 0$; since otherwise an unraveling equilibrium exists for $t_1 + \epsilon$, and *(iii)* $L_x(0, 0) < 0$; by Lemma 2. By continuity of L, L_x and L_y, we can find ϵ and δ such that for all $|x| < \epsilon, |y| < \delta$,

$$L(x, y) = L(0, 0) + \int\limits_0^x L_x(u, 0)du + \int\limits_0^y L_y(x, u)du < 0$$

It follows that for all $t \in (t_1, t_1 + \delta)$, and all $\theta \geq \bar{\theta} - \frac{\epsilon}{2}$, $BR^t(\theta) < \theta$, so that $\theta^t_{max} \leq \bar{\theta} - \frac{\epsilon}{2} \equiv \theta_1$, as required.

Proposition 2

Consider an equilibrium threshold $\theta^\star$ which is strictly larger than the smallest equilibrium threshold, $\theta^\star_{min}$. Then, the set $[\theta, p] \cup \left[\theta^\star_{min}, \overline{\theta}\right]$ is self-signaling. Recalling that $s^\star(\theta)$ is increasing in θ, we have

$$\theta \left(1 - H \left(s^\star \left(\theta^\star_{min}\right) \mid \theta\right)\right) > \theta \left(1 - H \left(s^\star \left(\theta^\star\right) \mid \theta\right)\right).$$

Thus, all types in $[\theta, p] \cup \left[\theta^{\star}_{min}, \overline{\theta}\right]$ prefer to switch to a cheap talk message understood to be sent by members of $[\theta, p] \cup \left[\theta^{\star}_{min}, \overline{\theta}\right]$. Moreover, types in $(p, \theta^{\star}_{min})$ would not wish to switch to this message – by definition of $\theta^{\star}_{min}$ as an equilibrium threshold.

We now show that $\theta^{\star}_{min}$ is a neologism proof equilibrium. From the above argument, it is therefore also unique. Suppose not, and for an arbitrary set C, let $s^{\star}(C)$ be the Receiver threshold identified in Lemma 1. Then there must exist a subset of non-disclosing types $C' \subset [\underline{\theta}, p] \cup \left[\theta^{\star}_{min}, \overline{\theta}\right]$ for whom

$$\theta \left(1 - H\left(s^{\star}(C) \mid \theta\right)\right) > \theta \left(1 - H\left(s^{\star}\left(\theta^{\star}_{min}\right) \mid \theta\right)\right)$$

if and only if $\theta \in C$. Therefore, $s^{\star}(C) > s^{\star}\left(\theta^{\star}_{min}\right)$, and moreover (given full support), we must have $[\theta, p] \cup \left[\theta^{\star}_{min}, \overline{\theta}\right] \subset C$. Given our regularity condition on Sender payoffs, we must have $C = [\theta, p] \cup \left[\theta', \overline{\theta}\right]$ for some $\theta' < \theta^{\star}_{min}$. But since $\theta^{\star}_{min}$ is the smallest equilibrium threshold, it is easy to show that we must have $B(\theta') > \theta'$. Thus, there exists a subset of $C \cap (p, \theta^{\star}_{min})$, $[\theta', B(\theta')]$ for whom

$$\theta \left(1 - H\left(s^{\star}(C) \mid \theta\right)\right) < \theta - c,$$

and therefore prefer their equilibrium message to deviating with members of C – a contradiction to C being a self-signaling set.

Lemma 3

With normal distributions, an interior equilibrium $(\theta^{\star}, s^{\star}(\theta^{\star}))$ solves the system

$$\Phi\left(\frac{s^{\star}(\theta^{\star}) - \theta^{\star}}{k}\right) = c \qquad (11)$$

$$\mathbb{E}\left[\theta | s^{\star}(\theta^{\star}), \theta \notin [p, \theta^{*}]\right] = p \qquad (12)$$

Letting $\mu_s = \alpha\mu + (1-\alpha)s - p$, $\sigma_s = \frac{k^2\sigma^2}{k^2+\sigma^2}$, with $\alpha = \frac{k^2}{k^2+\sigma^2}$, denote the conditional mean and variance of $\theta - p$ on observing s, we can use the standard formula for truncated normal distributions to write (12) as

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

or, defining $x = \frac{\theta^{\star}}{\sigma_{s^{\star}}}$, $y(x) = \frac{\mu_{s^{\star}}}{\sigma_{s^{\star}}}$ (recall that s^{*}, and therefore $\mu_{s^{*}}$ are functions of $\theta^{\star}$), we can write

$$y(x) = \frac{\phi(-y(x)) - \phi(x - y(x))}{1 - \Phi(x - y(x)) + \Phi(-y(x))} \tag{13}$$

Rewriting (12), we have

$$s^*(\theta^\star) = \frac{1}{1-\alpha}(\mu_{s^\star} + p - \alpha\mu) = \frac{1}{1-\alpha}.\,(\sigma_{s^\star} y(x) + p - \alpha\mu),$$

Differentiating system (11) - (12), after some algebra we find that in any stable equilibrium:

$$\begin{aligned} \frac{d\theta^\star}{dk} \overset{sign}{=} & \frac{2k}{k^2+\sigma^2} s^\star(\theta^\star) + \frac{1}{1-\alpha}\left[\left(y\left(\frac{\theta^\star}{\sigma_{s^\star}}\right) - \left(\frac{\theta^\star}{\sigma_{s^\star}}\right) y'\left(\frac{\theta^\star}{\sigma_{s^\star}}\right)\right)\frac{\sigma}{2\sqrt{\alpha}} - \mu\right]\frac{d\alpha}{dk} \\ & -\Phi^{-1}(c). \end{aligned} \tag{14}$$

When $\mu < 0$ then we have $s^\star(\theta^\star) > 0$ so that the first term is positive. Letting $\theta^\star/\sigma_{s^\star} = x$, 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 $y'(x)$. It is tedious but straightforward to show that each ray from the origin crosses the implicit function $y(x)$ once from above, which establishes that it must be steeper than $y(x)$ at the point of crossing (a proof was presented in an earlier working paper and is available on request). Finally, the third term is negative by assumption since $\Phi^{-1}(c) < 0$ for all $c < 1/2$.

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

$$\frac{d\Psi}{d\mu} = \frac{1}{k}\frac{ds^\star(\theta^\star)}{d\mu}.$$

Note further that $\frac{ds^\star(\theta^\star)}{d\mu} = \frac{-\alpha}{1-\alpha}$. We now consider the right-hand side of (14), which changes in proportion to $d\mu$ by

$$\begin{aligned} \frac{d}{d\mu}\left\{\frac{2k}{k^2+\sigma^2}\left[s^\star(\theta^\star) - \mu\right] - \Psi\right\} &= \frac{2k}{k^2+\sigma^2}\left[\frac{ds^\star(\theta^\star)}{d\mu} - 1\right] - \frac{1}{k}\frac{ds^\star(\theta^\star)}{d\mu}. \\ &= \frac{-2k}{k^2+\sigma^2}\frac{1}{1-\alpha} + \frac{1}{k}\frac{\alpha}{1-\alpha} \\ &= \frac{-2k}{\sigma^2} + \frac{1}{k}\frac{k^2}{\sigma^2} = -\frac{k}{\sigma^2} < 0. \end{aligned}$$

Thus the right-hand side of (14) changes linearly with $d\mu$ and is guaranteed to be negative whenever $d\mu$ is large enough, which completes the proof.

B Proofs for Section 3

Proposition 4

Define the naive policymaker's payoff function, given disclosure interval $[p, \theta']$, by

$$W_n(k, \theta') := 1 + \int_{p}^{\theta'} (\theta - c(\theta))\, dF(\theta) + \int_{\theta \in [\underline{\theta}, p] \cup [\theta', \overline{\theta}]} Pr\left[s \geq s_k^\star | \theta; k\right] \theta dF(\theta).$$

Using the equivalence $W(k) \equiv W_n(k, \theta_k^\star)$, we can calculate the derivative of the sophisticated welfare function as

$$\frac{\partial W}{\partial k} = \frac{\partial W_n}{\partial k} + \frac{\partial W_n}{\partial \theta'} \frac{\partial \theta_k^\star}{\partial k}. \tag{15}$$

At an interior optimum of the naive policymaker's problem, we have $\frac{\partial W_n}{\partial k}(k^\star, \theta_{k^\star}^\star) = 0$. Moreover, it is easy to see that

$$\begin{aligned} \frac{\partial W_n}{\partial \theta'} &= - \frac{\partial s_k^\star}{\partial \theta'} \cdot \int_{\theta \notin [p, \theta']} h\left(\frac{s_k^\star - \theta}{k}\right) \theta dF(\theta) \\ &= -\frac{\partial s_k^\star}{\partial \theta'} \cdot \mathbb{E}\left[\theta \mid \theta \notin [p, \theta'], s_k^\star\right] \int_{\theta \notin [p, \theta']} h\left(\frac{s_k^\star - \theta}{k}\right) dF(\theta) \\ &= -\frac{\partial s_k^\star}{\partial \theta'} p \int_{\theta \notin [p, \theta']} h\left(\frac{s_k^\star - \theta}{k}\right) f(\theta) d\theta \end{aligned}$$

Since $h\left(\frac{s_k^\star - \theta}{k}\right) f(\theta) > 0$, so too is the integral above. Moreover, $\frac{\partial s_k^\star}{\partial \theta'}$ follows from the MLRP property on signals. Thus, $\frac{\partial W}{\partial k}$ takes the sign of $-\frac{\partial \theta_k^\star}{\partial k}$.

Let $k^{\star\star} := \arg\max W(k)$. We now show that when (7) and $\theta_k^\star \geq \theta_{k^\star}^\star$ holds, for all $k \geq k^\star$ then $k^{\star\star} < k^\star$. Since the argument is analogous, we omit the proof for the crowding in case.

Calculating $W(k^\star) - W(k)$ for $k < k^\star$ we have

$$\begin{aligned} W(k^\star) - W(k) &= W_n(k^\star, \theta_{k^\star}^\star) - W(k) \\ &= W_n(k^\star, \theta_{k^\star}^\star) - W_n(k, \theta_k^\star) \\ &\geq W_n(k^\star, \theta_{k^\star}^\star) - \left(W_n(k^\star, \theta_{k^\star}^\star) + \int_{\theta_{k^\star}^\star}^{\theta_k^\star} (Pr\left[s \leq s^\star(\theta_k^\star, k) | \theta; k\right] - c) \theta dF(\theta) \right) \end{aligned}$$

where the last inequality follows from equation (15) and $s^\star(\theta_k^\star, k) \geq s^\star(\theta_{k^\star}^\star, k)$, for $\theta_{k^\star}^\star > \theta_k^\star$.

From this, the Proposition follows immediately.[35]

C Proofs for Section 4

In this Appendix, we write Receiver's beliefs interchangeably as functions of realizations of θ, as in $\mu(\theta_i)$, or with superscripts, as in μ^i. We write $V(\theta) = v(a^\star(\theta))$ for Sender's payoff when he is taken to be type θ for certain. When Sender stays quiet and outside information is s, let

$$\alpha(s) \in \arg\max E_\mu[u(a,\theta)|s, m = \emptyset]$$

represent Receiver's (potentially random) best response. We then define the net payoff from a verifiable disclosure as

$$\mathcal{N}(\theta) \equiv V(\theta) - E[v(\alpha(s))|\theta], \tag{16}$$

so that Sender prefers to disclose if $\mathcal{N}(\theta) \geq c$. As pointed out by Milgrom and Roberts (1986) and others, it is often useful to consider 'skeptical' beliefs, where Receiver assumes that Sender is of the worst type $\underline{\theta}(s) = \min\{\theta|\pi(s|\theta) > 0\}$ consistent with her outside information s. We define the *maximal punishment* that Sender can suffer by staying quiet as the difference between the payoff he obtains under full disclosure, and the payoff he obtains by staying quiet and facing a skeptical Receiver:

$$\mathcal{M}(\theta) = V(\theta) - E[V(\underline{\theta}(s))|\theta].$$

Proposition 5

Proof. We construct a simple path of signals $\Pi(t)$ that satisfies the claim of the Proposition. Let $p_i : [0,1] \to [0,1]$ be a C^2, strictly increasing function with $p_i(0) = 0$, $p_i(1) = 1$ and whose derivative is equicontinuous, for $i = 1, \ldots, N$. Iteratively define the following class of outside signals: let $\hat{\Pi}(t)$ be an $N \times N$ matrix whose elements are

$$\hat{\pi}(s \mid \theta_i; t) \;=\; \begin{cases} (1 - p_i(t))\, \hat{\pi}(s \mid \theta_{i-1}; t), & s < i \\ p_i(t), & \text{for } s = i \\ 0, & \text{for } s > i. \end{cases}$$

[35] The right-hand side integral is always weakly positive, since from the regularity condition (1) and the equilibrium condition, we have $Pr\left[s \leq s^\star(\theta_k^\star, k)\,|\theta; k\right] \geq c$, for all $\theta \in [p, \theta_k^\star]$, and moreover $\theta \geq \theta_{k^\star}^\star \geq p > 0$.

$\hat{\Pi}(t)$ satisfies MLRP for all t. We show first that

$$\mathcal{M}(\theta_i;t) = \sum_{s=1}^{i-1} \hat{\pi}(s \mid \theta_i;t)(V(\theta_i) - V(\theta_s))$$

is decreasing in t, with $\mathcal{M}(\theta_i;0) > c$, $\mathcal{M}(\theta_i;1) = 0$, $\forall i > 1$.

$$\mathcal{M}(\theta_i;t) = (1 - p_k(t))\mathcal{M}(\theta_{i-1};t) + p_k(t)(V(\theta_i) - V(\theta_k))$$

We argue inductively: If $\mathcal{M}(\theta_{i-1};t)$ is increasing in t then clearly so too is $\mathcal{M}(\theta_i;t)$, since $p_k(t)$ is increasing in t and $V(\theta_i) - V(\theta_s)$ is strictly decreasing in s. Observing that $\mathcal{M}(\theta_2;t) = (1 - p_2(t))(V(\theta_2) - V(\theta_1))$ is decreasing establishes monotonicity. Thus for each θ_i, there is a unique t'_i at which $\mathcal{M}(\theta_i;t) = c$. Moreover, we can find a $\Pi(t)$ such that $t'_i = t'_j = t^\star$, $\forall i,j$. To do this, we iteratively adjust $\hat{\Pi}(t)$: suppose a matrix $\Pi'_k(t)$ induces $\mathcal{M}(\theta_i;t^\star_k) = \mathcal{M}(\theta_j;t^\star_k)$, $\forall i,j \leq k$. Construct $\Pi'_{k+1}(t)$ as follows: if $t'_{k+1} > t^\star_k$, replace row k of $\Pi'_k(t)$ with the functions $\left(\pi'\left(s \mid \theta_k; \frac{t'}{t^\star_k}t\right)\right)_{s=1}^{n}$. Otherwise, replace each row $i < k$ with $\left(\pi'\left(s \mid \theta_i; \frac{t^\star_k}{t'}t\right)\right)_{s=1}^{n}$. Applying this process to $\hat{\Pi}(t)$ clearly yields the required matrix $\Pi(t)$ after N iterations.[36]

For outside signal $\Pi(t)$, transparency is trivially an equilibrium for $t \leq t^\star$. Finally, we show there exists a $\delta > 0$ such that if at $t^\star$, $c \leq \mathcal{M}(\theta_i;t^\star) \leq c + \delta$, with $\mathcal{M}(\theta_k;t^\star) = c$ for at least some k, then for all $t > t^\star$, full opacity is the unique equilibrium of the disclosure game. At any $t > t^\star$, we have $\mathcal{M}(\theta_k;t) < c$. Thus, for any equilibrium strategy profile, θ_k's net payoff from disclosure is

$$\sum_{s=1}^{k-1} \pi(s \mid \theta_k;t)(V(\theta_k) - v(\alpha(s))) \leq \mathcal{M}(\theta_k;t) < c$$

since for any log-supermodular u and increasing v, $v(\alpha(s)) \geq V(\theta_s)$. Thus, in any equilibrium $m(\theta_k) = \emptyset$, $\forall t > t^\star$. Given any signal $s < k$, define the vector of beliefs $\mu_s = (Pr[\theta|s,\emptyset])_{\theta\in\Theta}$, and define $\nu(\mu_s)$ as Sender's utility when Receiver has these beliefs. We can bound $\mu_s \geq \underline{\kappa}^t_s \mathbf{1}_k + (1 - \underline{\kappa}^t_s)\mathbf{1}_s > \mathbf{1}_s$, where $\underline{\kappa}^t_s > 0$ satisfies

$$\begin{aligned} \underline{\kappa}^t_s &= \min_{\{\sigma|\sigma(\theta_k)=0\}} \Pr[\theta_k \mid s, m = \emptyset] &= \frac{\Pr[s, m = \emptyset \mid \theta_k]\mu_0(\theta_k)}{\Pr[s, m = \emptyset]} \\ & &\geq \frac{\pi(s|\theta_k)\mu_0(\theta_k)}{\Pr[s]} > 0 \end{aligned}$$

for any $t < 1$, which follows since $\Pr[m = \emptyset \mid \theta_k] = 1$ in equilibrium and $\Pr[s, m = \emptyset] \leq$

[36] It is simple to re-parameterize $\Pi(t)$ to ensure that $p_i(t) < 1$ for $t < 1$.

$\Pr(s)$. Since $\alpha(s)$ is strictly increasing in the LR order, $\nu(\mu_s) > \nu(\mathbf{1}_s) = V(\theta_s)$. Define $\delta = \min_i[1 - \pi(s \mid \theta_i; t)][\nu\left(\underline{\mu}_s^{t^\star}\right) - \nu(\mathbf{1}_s)]$, where $\underline{\mu}_s^{t^\star} = \underline{\kappa}_s^{t^\star}\mathbf{1}_k + \left(1 - \underline{\kappa}_s^{t^\star}\right)\mathbf{1}_s$. Then in any equilibrium the net payoff to disclosure for type θ_i satisfies

$$\mathcal{N}(\theta_i; t) \leq \mathcal{M}(\theta_i; t) - \sum_{s<k} \pi(s \mid \theta_i; t)\left[\nu(\underline{\mu}_s^{t^\star}) - \nu(\mathbf{1}_s)\right] \leq \mathcal{M}(\theta_i; t) - \delta$$

When $\mathcal{M}(\theta_i; t) \leq c + \delta$, $\forall i$, all types strictly prefer to set $m(\theta_i) = \emptyset$ in any equilibrium. □

Proposition 6

Proof. Fix (S, Π) and a corresponding equilibrium strategy profile $\sigma^\star$, actions $\{\alpha(s)\}_{s\in S}$ and Receiver posterior beliefs $(\mu_s)_{s\in S}$. Suppose further that $\sigma(\theta_i) > 0$ for some θ_i. Partition Θ as follows: $\theta \in Q \iff \mathcal{N}^\star(\theta) < c$, $\theta \in D$ otherwise. Now consider the following modified signal structure, $(S \cup \Theta, \Pi')$ which satisfies

$$\pi'(s \mid \theta_i) = \begin{cases} \pi(s \mid \theta_i), & \theta_i \in Q, s \in S \\ 0, & \theta_i \in Q, s \in \Theta \\ (1 - z_i)\,\pi(s \mid \theta_i), & \theta_i \in D, s \in S \\ z_i, & \theta_i \in D, s = \theta_i \in \Theta \end{cases}.$$

where $z_i \leq \sigma(\theta_i)$. For each θ_i, $\exists \underline{z}_i < \sigma(\theta_i)$ such that, for all $\underline{z}_i < z_i$ and $\theta_i \in D$:

$$\sum_{s\in S\cup\Theta} \pi'(s \mid \theta_i)\left(V(\theta_i) - v(\alpha(s))\right) = (1 - z_i) \sum_{s\in S\cup\Theta} \pi(s \mid \theta_i)\left(V(\theta_i) - v(\alpha(s))\right) < c.$$

Fix $\underline{z}_i \leq z_i \leq \sigma(\theta_i)$. Let $\sigma'(\theta) = 0$. Given strategy profile ζ', and outside signals $(S \cup \Theta, \Pi')$, Receiver's posterior beliefs $(\hat{\mu}_s^i)_{i=1}^N$ given $s \in S$ can be written

$$\frac{\hat{\mu}_s^i}{\hat{\mu}_s^j} = \frac{\mu_0(\theta_i)\,(1 - z_i)\,\pi(s \mid \theta_i)}{\mu_0(\theta_j)\,(1 - z_j)\,\pi(s \mid \theta_j)}$$

As $z_i \to \sigma(\theta_i)$, $\frac{\hat{\mu}_s^i}{\hat{\mu}_s^j} \to \frac{\mu_s^i}{\mu_s^j}$. Thus, $\hat{\mu}_s \to \mu_s$. Given finiteness of Θ, S, for any $\varepsilon > 0$ there exists bounds $(\overline{z}_i)_{i=1}^N$ such that $|\hat{\mu}_s - \mu_s| < \varepsilon$ whenever $\overline{z}_i < z_i < \sigma(\theta_i)$, $\forall i$. If $\alpha(s)$ is continuous in μ (which holds because Receiver has a unique best response), then given strict preference for nondisclosure of all types under action profile $\{\alpha(s)\}_{s=1}^N$, and outside signals $(S \cup \Theta, \Pi')$, we can therefore find a $\varepsilon > 0$ such that opacity is an equilibrium of this game.

Finally, the opaque equilibrium with outside signals $(S \cup \Theta, \Pi')$ is a Blackwell garbling of equilibrium information structure with $\sigma^\star$ and (S, Π). To see this, note that one can construct the equilibrium signal Receiver observes under the former equilibrium by the garbling the Sender's disclosures in the latter equilibrium as follows: given message $m = \emptyset$ and signal s, use the 'truthful' garbling $\Pr(s \mid s, m = \emptyset) = 1$; given message $m = \theta$, garble to signal $s \in S \cup \Theta$ with probabilities $\Pr(s = \theta \mid m = \theta) = \frac{z_i}{\sigma^\star(\theta_i)}$ for $s = \theta$, $\Pr(s \mid m = \theta) = \left(1 - \frac{z_i}{\sigma^\star(\theta_i)}\right) \pi(s \mid \theta_i)$ for $s \in S$. □

Proposition 7

Proof. We write Concavity $= \chi$ and Convexity $= \xi$. We split the proof into two parts. First, we show that sufficiently concave payoffs imply that all equilibria are non-monotone or opaque. Second, we show that sufficiently convex payoffs imply that there are no non-monotone equilibria.

Part 1: Concave payoffs

Let $\Sigma_m \subset [0,1]^N$ be the space of monotone increasing disclosure strategies. For any $\sigma \in \Sigma_m$, define $d(\sigma) = \min\{i|\sigma_i > 0\}$ as the lowest disclosing type, and $q(\sigma) = d(\sigma) - 1$ the highest type who stays quiet with probability 1. We first derive a bound on the weights that these two types attach to being perceived as type $q(\sigma) - 1$ or worse, assuming that this type exists (i.e., that $q(\sigma) > 1$). For all $j < q(\sigma)$, Receiver's beliefs if Sender stays quiet are interior with $Pr[\theta \leq \theta_j|\emptyset] \in (0,1)$. For any pair of signals (s', s), where $s' > s$ and at least one of them is drawn by type j or worse with positive probability, the strict MLRP of signals implies strict first-order stochastic dominance (see Milgrom 1981, Theorem 1):

$$Pr[\theta \leq \theta_j|\emptyset, s'] < Pr[\theta \leq \theta_j|\emptyset, s] \text{ for all } s' > s.$$

The cumulative distribution of virtual types $Q^\sigma_{ij} = E_s[Pr[\theta \leq \theta_j|s, \emptyset]|\theta_i]$ is therefore the expectation of an decreasing function of s, where the superscript σ is introduced to highlight the dependence of Receiver's beliefs on equilibrium play. Using the MLRP again, we find that Q^σ_{ij} is strictly decreasing in i for $j < q(\sigma)$. Thus for all feasible σ,

$$Q^\sigma_{d(\sigma),q(\sigma)-1} - Q^\sigma_{q(\sigma),q(\sigma)-1} < 0,$$

Define

$$\mathcal{Q}_m = \max_{\{\sigma \in \Sigma_m | q(\sigma) > 1\}} Q^\sigma_{d(\sigma),q(\sigma)-1} - Q^\sigma_{q(\sigma),q(\sigma)-1}.$$

It is easy to see that the constraint set is compact, so that the maximum is achieved and satisfies $\mathcal{Q}_m < 0$. Note that we can repeat the stochastic dominance argument above for

$j \geq q(\sigma)$: In this case we may have $Q^{\sigma}_{ij} = 0$ for a range of i (for example, if only types below j stay quiet with positive probability), but a parallel argument establishes that Q^{σ}_{ij} is non-increasing in i.

Next, suppose that $\sigma \in \Sigma_m$ is a monotone increasing strategy played in equilibrium, and assume that $c > c_0$, where c_0 satisfies

$$c_0 = \min_{\theta > \theta_1} \mathcal{M}(\theta).$$

By definition, when $c > c_0$ there must exist a type $\theta_j = \arg\min_{i \geq 2} \mathcal{M}(\theta_i)$ who has a dominant strategy to stay quiet, so that $\sigma_j = 0$. By monotonicity, we have $\sigma_i = 0$ for all $i \leq j$, and it follows that the highest quiet type $q(\sigma) > 1$. Optimality requires that this type prefers to stay quiet and the lowest discloser $d(\sigma)$ prefers to disclose. We obtain $\mathcal{N}(\theta_{d(\sigma)}) \geq c \geq \mathcal{N}(\theta_{q(\sigma)})$, implying

$$\begin{aligned}
0 &\leq \mathcal{N}(\theta_{d(\sigma)}) - \mathcal{N}(\theta_{d(\sigma)-1}) \\
&= \Delta X_{d(\sigma)-1} + \sum_{i=1}^{N-1} \Delta X_i (Q^{\sigma}_{d(\sigma),i} - Q^{\sigma}_{d(\sigma)-1,i}) \\
&\leq \Delta X_{d(\sigma)-1} + \Delta X_{d(\sigma)-2} (Q^{\sigma}_{d(\sigma),i} - Q^{\sigma}_{d(\sigma)-1,i}) \\
&\leq \Delta X_{d(\sigma)-1} + \Delta X_{d(\sigma)-2} \mathcal{Q}_m.
\end{aligned}$$

where the second inequality follows by first-order stochastic dominance, and the third imposes the bound derived above. Dividing by $\Delta X_{d(\sigma)}$ and using $\mathcal{Q}_m < 0$, we obtain

$$\frac{\Delta X_{d(\sigma)-2}}{\Delta X_{d(\sigma)-1}} \leq \frac{1}{|\mathcal{Q}_m|}.$$

This further implies that the concavity parameter $\chi \leq \frac{1}{|\mathcal{Q}_m|}$. We have now shown that the existence of a monotone increasing equilibrium with $c > c_0$ implies that $\chi$is bounded above. By contrapositive, if χ is sufficiently large, then there is no monotone equilibrium for the range of disclosure costs $c > c_0$, as required.

Part 2: Convex payoffs

Let $\Sigma_{nm} \subset [0, 1]^N$ be the space of non-monotones strategy profiles. For $\sigma \in \Sigma_{nm}$, we can define $q(\sigma) = \max\{i | \sigma_i < 1\}$ as the highest type who stays quiet with positive probability, and $d(\sigma) = \max\{i < q(\sigma) | \sigma_i > 0\}$ as the highest discloser below $q(\sigma)$. Since type θ_1 has a dominant strategy, we have $d(\sigma) > 1$ and $q(\sigma) > 2$. We first derive a bound on the weights that these two types attach to being perceived as type $q(\sigma) - 1$ or worse. Let

$$\mathcal{Q}_{nm} = \inf_{\sigma \in \Sigma_{nm}} \{Q^{\sigma}_{q(\sigma),q(\sigma)-1} - Q^{\sigma}_{d(\sigma),q(\sigma)-1}\}.$$

Since the cumulative probabilities $Q^{\sigma}_{ij} \in [0, 1]$, we have $\mathcal{Q}_{nm} \geq -1$. We show that this inequality is strict. Suppose, for a contradiction, that $\mathcal{Q}_{nm} = -1$. Then for every ϵ we can find strategies $\sigma \in \Sigma_{nm}$ such that $Q^{\sigma}_{q(\sigma),q(\sigma)-1} - Q^{\sigma}_{d(\sigma),q(\sigma)-1} < -1 + \epsilon$. This implies two requirements: $Q^{\sigma}_{q(\sigma),q(\sigma)-1} < \epsilon$ and $Q^{\sigma}_{d(\sigma),q(\sigma)-1} > 1 - \epsilon$, that is, type $q(\sigma)$ almost never draws a virtual type worse than himself, while type $d(\sigma)$ almost never draws a better virtual type than $q(\sigma) - 1$. Since neighboring types share signals, we can find a realization $s = s'$ that is drawn with positive probability by both $q(\sigma)$ and $q(\sigma) - 1$. Our first requirement implies that Receiver's posterior belief, after observing $m = \emptyset$ and $s = s'$, satisfies $Pr[\theta \leq \theta_{q(\sigma)-1} | \emptyset, s'] \leq \delta(\epsilon)$, where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$. For small enough ϵ, this is only possible if type $q(\sigma) - 1$ discloses with positive probability (otherwise Receiver would place a discrete probability mass on this type when she observes $m = \emptyset$). Therefore, we know that the highest discloser below $q(\sigma)$ is his neighbor: $d(\sigma) = q(\sigma) - 1$. Our second requirement now implies that Receiver's posterior belief satisfies $Pr[\theta > \theta_{q(\sigma)-1} | \emptyset, s'] \leq \hat{\delta}(\epsilon)$, where $\hat{\delta}(\epsilon) \to 0$ as $\epsilon \to 0$. We can write

$$1 = Pr[\theta \leq \theta_{q(\sigma)-1} | \emptyset, s'] + Pr[\theta > \theta_{q(\sigma)-1} | \emptyset, s'] \leq \delta(\epsilon) + \hat{\delta}(\epsilon),$$

and taking limits as $\epsilon \to 0$, we get a contradiction. Therefore, $\mathcal{Q}_{nm} > -1$.

Next, suppose that $\sigma \in \Sigma_{nm}$ is a non-monotone strategy played in equilibrium. Optimality requires that $\mathcal{N}(\theta_{d(\sigma)}) \geq c \geq \mathcal{N}(\theta_{q(\sigma)})$, implying

$$\begin{aligned} 0 &\geq \mathcal{N}(\theta_{q(\sigma)}) - \mathcal{N}(\theta_{d(\sigma)}) \\ &= \sum_{i=d(\sigma)}^{q(\sigma)-1} \Delta X_i + \sum_{i=1}^{N-1} \Delta X_i (Q^{\sigma}_{q(\sigma),i} - Q^{\sigma}_{d(\sigma),i}). \end{aligned}$$

Note that $Q_{ij} = 1$ for all i and $j \geq q(\sigma)$, since Receiver attaches probability $Pr[\theta_j | \emptyset] = 0$ to types $j > q(\sigma)$ when Sender stays quiet. Moreover, a parallel argument to Part 1 of this proof establishes that the distribution of virtual types given $\theta = q(\sigma)$ first-order stochastically dominates that given the lower type $\theta = d(\sigma)$, so that $Q^{\sigma}_{q(\sigma),i} - Q^{\sigma}_{d(\sigma),i} \leq 0$. Dividing the above inequality by $\Delta X_{q(\sigma)-1}$ and combining these observations,

$$\begin{aligned}
0 &\geq 1 + Q^{\sigma}_{q(\sigma),q(\sigma)-1} - Q^{\sigma}_{d(\sigma),q(\sigma)-1} + \sum_{i=1}^{q(\sigma)-1} \left(\frac{\Delta X_i}{\Delta X_{q(\sigma)-1}} \right) (\mathbf{1}_{i \geq d(\sigma)} + Q^{\sigma}_{q(\sigma),i} - Q^{\sigma}_{d(\sigma),i}) \\
&\geq 1 + Q^{\sigma}_{q(\sigma),q(\sigma)-1} - Q^{\sigma}_{d(\sigma),q(\sigma)-1} + \sum_{i=1}^{q(\sigma)-2} \xi^{-[q(\sigma)-1-i]} (\mathbf{1}_{i \geq d(\sigma)} + Q^{\sigma}_{q(\sigma),i} - Q^{\sigma}_{d(\sigma),i}) \\
&\geq (1 + \mathcal{Q}_{nm}) - \sup_{\sigma \in \Sigma_{nm}} \sum_{i=1}^{q(\sigma)-2} \xi^{-[q(\sigma)-1-i]},
\end{aligned}$$

where the last line follow noting that $\mathbf{1}_{i \geq d(\sigma)} + Q^{\sigma}_{q(\sigma),i} - Q^{\sigma}_{d(\sigma),i} \geq -1$ and then taking the infimum. We know that the first term $1 + \mathcal{Q}_{nm} > 0$. Thus the second term must be smaller than $-(1 + \mathcal{Q}_{nm})$. However, it is easy to see that the limit of this term as $\xi \to \infty$ is zero, so that the above series of inequalities gives us $\xi \leq \xi_0$ for some finite ξ_0. We have now shown that the existence of a non-monotone equilibrium implies an upper bound on ξ. By contrapositive, if ξ is sufficiently large, then there is no non-monotone equilibrium, as required. □

Online Appendix

D Online Appendix: Additional results for Section 2

D.1 Unbounded types

The results of Section 2 refer to the case where Sender's type θ is drawn from a bounded interval $[\underline{\theta}, \bar{\theta}]$. We now consider the case where the distribution $F(\theta)$ of θ has full support on the real line. It is easy to see that the threshold properties of equilibria (see Lemma 1) go through in the unbounded case. Also, a similar result to Proposition 1 applies:

Proposition 8. *For any revealing path s_t, there exists a threshold $t_0 \in (0, 1)$ such that $\theta^\star_{min} = \infty$ for all $t < t_0$ and $\theta^\star_{min} < \bar{\theta}$ for $t = t_0$.*

Proof. We adopt the same notation (BR^t, θ^t_{min}) as in the Proof of Proposition 1. Let $t_0 = \inf\{t : \theta^t_{min} < \infty\}$ We know that unraveling ($\theta^\star = \infty$) is the unique equilibrium in a neighborhood around $t = 0$, so that $t_0 > 0$. Moreover, since an equilibrium without any disclosure ($\theta^\star = p$) exists for $t = 1$, we know by continuity of $BR^t(\theta^\star)$ that $t_0 < 1$. Furthermore, we must have $\theta^t_{min} < \infty$; otherwise we can find a small ϵ such that $BR^{t_0-\epsilon}(\theta) > \theta$ for all θ, contradicting the definition of t_0. □

D.2 Instability of unraveling equilibrium

Next, we derive an alternative equilibrium selection criterion based on the stability of equilibria in population games as in Schelling (1978). We allow for either bounded or unbounded types ($\theta \in [\underline{\theta}, \bar{\theta}]$, with $\bar{\theta} < \infty$ or $\bar{\theta} = \infty$ respectively).

Definition 1. An equilibrium with disclosure threshold $\theta^\star \in [p, \infty]$ is unstable if $BR(\theta) - \theta$ has the same sign as $\theta - \theta^\star$ for all θ in some neighborhood of $\theta^\star$.

An unstable interior equilibrium is one for which the best response function in Figure 2 crosses the 45-degree line from below. Then, small mistakes in Sender's disclosure strategy lead to divergence of equilibrium play from $\theta^\star$ under best response dynamics. By analogy, an unstable unraveling equilibrium in the case of unbounded types is one where $BR(\infty) = \infty$ but $BR(\theta) - \theta < 0$ for all $\theta \geq B$ for some B, so that the best response function approaches the 45-degree line from below in the limit. In this case, a deviation by any set $\theta \geq B'$, no matter how large B' is, leads to divergence from unraveling under best response dynamics. In an earlier working paper, we provided a formal proof that the above definition is equivalent to a definition in terms of best response dynamics, which is available on request.

With full support (as we have assumed), the unraveling equilibrium $\theta^\star = \bar{\theta}$ always exists. However, we can find a condition under which it is unstable. We focus here on the case where $s = \theta + k\epsilon$, where ϵ is a random variable with smooth distribution $G(\epsilon)$.

Proposition 9. *Unraveling ($\theta^\star = \infty$) is an unstable equilibrium for all disclosure costs* $c > 0$ *if and only if for any $K \in \mathbb{R}$, $\exists \tilde{\theta}'$ such that $\forall \theta^\star \geq \tilde{\theta}'$:*

$$-\frac{Pr\left(\theta \geq \theta^\star | s = \theta^\star + K\right) . E\left[\theta | \theta \geq \theta^\star,\, s = \theta^\star + K\right]}{Pr\left(\theta \leq p \mid s = \theta^\star + K\right) . E\left[\theta | \theta \leq p,\, s = \theta^\star + K\right]} > 1 \tag{17}$$

A proof is below. We establish Proposition 9 by considering a small deviation from unraveling. Instead of expecting every type $\theta \geq p$ to disclose, the Receiver mistakenly expects an small portion of high quality types $[\theta_1, \infty)$ 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 (17), 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.

When θ and ε are jointly Normally distributed with $\theta \sim \mathcal{N}\left(\mu, \sigma^2\right)$ and $\epsilon \sim \mathcal{N}\left(0, 1\right)$, condition (17) has a particularly natural interpretation. In this case, (17) holds if and only if the signal-to-noise ratio is greater than one, $\sigma > k$. 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.

Proof of Proposition 9

Proof. We prove sufficiency by arguing the contrapositive: if unraveling is stable, then there must exist a $\tilde{K} \in \mathbb{R}$ that violates (17). Thus, suppose that transparency is a stable equilibrium. Then we can find a $\mathcal{B}$ such that $\forall \theta^\star \geq \mathcal{B}$

$$BR(\theta^\star) - \theta^\star > 0 \tag{18}$$

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

$$c = G\left(\frac{s^*(\theta^\star) - BR(\theta^\star)}{k}\right)$$

or

$$BR(\theta^\star) = s^*(\theta^\star) - kG^{-1}(c) \tag{19}$$

Substituting (19) into (18) yields a lower bound on $s^*(\theta^\star)$ as a function of $\theta^\star$ for any unraveling equilibrium:

$$s^*(\theta^\star) > \theta^\star + kG^{-1}(c) \tag{20}$$

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

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

or

$$\Pr(\theta \leq p \mid s^*(\theta^\star)) . \mathbb{E}[\theta|\theta \leq p, \, s^*(\theta^\star)] + \Pr(\theta \geq \theta^\star | s^*(\theta^\star)) . \mathbb{E}[\theta|\theta \geq \theta^\star, \, s^*(\theta^\star)] = p \tag{21}$$

Now consider the left hand side of Equation (17) evaluated at $\tilde{K} = kG^{-1}(c)$. By (20), $s^*(\theta^\star) > \theta^\star + \tilde{K}$. Then, it follows immediately from (21) and the MLRP assumption on signals s that

$$\begin{aligned} &\Pr\left(\theta \notin [p, \theta^\star] \mid s = \theta^\star + \tilde{K}\right) . \mathbb{E}\left[\theta | s = \theta^\star + \tilde{K}, \, \theta \notin [p, \theta^\star]\right] \\ &\quad = \Pr\left(\theta \leq p \mid s = \theta^\star + \tilde{K}\right) . \mathbb{E}\left[\theta|\theta \leq p, \, s = \theta^\star + \tilde{K}\right] \\ &\quad + \Pr\left(\theta \geq \theta^\star | s = \theta^\star + \tilde{K}\right) . \mathbb{E}\left[\theta|\theta \geq \theta^\star, \, s = \theta^\star + \tilde{K}\right] \quad < \quad p \end{aligned}$$

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

$$-\left(\frac{\Pr\left(\theta \geq \theta^\star | s = \theta^\star + \tilde{K}\right) . \mathbb{E}\left[\theta|\theta \geq \theta^\star, \, s = \theta^\star + \tilde{K}\right]}{\Pr\left(\theta \leq p \mid s = \theta^\star + \tilde{K}\right) . \mathbb{E}\left[\theta|\theta \leq p, \, s = \theta^\star + \tilde{K}\right]}\right) < 1 \tag{22}$$

This final inequality shows our contrapositive claim that if transparency is a stable equilibrium then for $K \leq \tilde{K} = kG^{-1}(c)$, (17) is violated.[37]

(Necessity) We prove the argument by contradiction, in each of two cases. Suppose then that (17) does not hold, but that the unraveling equilibrium is stable for all choices of $c > 0$.

[37] MLRP implies that if (22) holds for $\tilde{K}$, then it also holds for all $K \leq \tilde{K}$.

Then there exists some $\tilde{K}$, $\tilde{\theta}' \in \mathbb{R}$ such that[38]

$$-\frac{\Pr\left(\theta \geq \theta^{\star}|s = \theta^{\star} + \tilde{K}\right).\mathbb{E}\left[\theta|\theta \geq \theta^{\star},\, s = \theta^{\star} + \tilde{K}\right]}{\Pr\left(\theta \leq p \mid s = \theta^{\star} + \tilde{K}\right).\mathbb{E}\left[\theta|\theta \leq p,\, s = \theta^{\star} + \tilde{K}\right]} \leq 1 \tag{23}$$

$\forall\theta^{\star} \geq \tilde{\theta}'$. Note that (23) holds everywhere, not just in the limit, since the infimum is a non-decreasing function. Denoting for simplicity,

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

we now consider the following two exhaustive cases:

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

Case 1.

We argue that, under condition (23), $\exists\bar{\rho} > 0$ such that transparency is a stable outcome for all $c < \bar{\rho}$. Specifically, for all $\theta^{\star} \geq \tilde{\theta}'$, we know that

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

which can be equivalently expressed as

$$\begin{aligned} p \;>\; & \Pr\left(\theta \leq p \mid s = \theta^{\star} + \tilde{K}\right).\mathbb{E}\left[\theta|\theta \leq p,\, s = \theta^{\star} + \tilde{K}\right] \\ & + \Pr\left(\theta \geq \theta^{\star}|s = \theta^{\star} + \tilde{K}\right).\mathbb{E}\left[\theta|\theta \geq \theta^{\star},\, s = \theta^{\star} + \tilde{K}\right] \end{aligned}$$

or

$$\mathbb{E}\left[\theta|\theta \notin [p, \theta^{\star}]\,,\, s = \theta^{\star} + \tilde{K}\right] < p \tag{24}$$

Therefore, since s satisfies the MLRP condition, (24) implies that $s^{*}(\theta^{\star}) > \theta^{\star} + \tilde{K}$ for all $\theta^{\star} \geq \tilde{\theta}'$ or

$$s^{*}(\theta^{\star}) - \theta^{\star} > \tilde{K} \tag{25}$$

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

$$s^{*}(\theta^{\star}) - BR(\theta^{\star}) = kG^{-1}(c) \geq \tilde{K} \tag{26}$$

[38] Note by MLRP that if (23) holds for $\tilde{K}$, then it also holds for all $K \leq \tilde{K}$.

Comparing (25) and (26) establishes that for all $\theta^\star \geq \tilde{\theta}'$, $BR(\theta^\star) > \theta^\star$ - a contradiction to the assumed instability of the unraveling equilibrium.

Case 2.

We argue that under condition (23), $\exists \bar{\rho} > 0$ (defined as above) such that transparency is a *neutrally* stable outcome for all $c \leq \bar{\rho}$: 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 (23), we can find always find a sequence of values $\left\{\tilde{\theta}'_n\right\}_{n=1}^{\infty} \to \infty$ such that

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

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

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

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

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

defines intervals $\left[\theta'_q, \theta'_{q+1}\right]$. By the assumed continuity of $\mathbb{E}\left[\theta \mid s, \theta \notin [p, \theta']\right]$ 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,

$$\mathbb{E}\left[\theta | \theta^*_q + \tilde{K}, \theta \notin \left[p, \theta^*_q\right]\right] = p$$

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

From the proof of Lemma **??**, we have established that for any $\theta_n \in \left[\theta'_q, \theta'_{q+1}\right]$, best response dynamics imply convergence 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 unraveling equilibrium is at least neutrally stable for all $c < \bar{\rho}$. This contradicts out assumption that the unraveling equilibrium was unstable. □

E Online Appendix: Additional results for Section 3

E.1 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 until Section E.2. In particular, managers have a contract with investors which implies that managers' private benefit of avoiding a run is $B(\theta)$ and their private cost of disclosure is $D(\theta)$, while the true social costs and benefits are $c(\theta)$ and θ respectively, as before.

As before, we consider the effect of increasing k beyond the naive policymaker's optimal choice. The local effects we described in Proposition 4 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 10. *At the naive policymaker's optimal choice $k^\star$, the marginal effect of further increasing k on welfare is the sum of the effect described in Proposition 4, and a term which has the same sign as*

$$\frac{\partial \theta_k^\star}{\partial k} \times \left[\frac{D(\theta_k^\star)}{B(\theta_k^\star)} - \frac{\delta(\theta_k^\star)}{\theta_k^\star} \right] \tag{27}$$

If better stress tests ($\downarrow k$) crowd out disclosures, then $\partial\theta_k^\star/\partial k > 0$. Proposition 10 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 managers 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 (as suggested by Lemma 3), 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.

E.2 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 Section 3, 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 E.1. 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 $\theta \in [p, \theta^\star]$ are anxious and disclose.

Perhaps surprisingly, the basic welfare analysis is also unchanged. Crowding out disclosure has a positive (albeit quantitatively smaller) effect on welfare, as demonstrated in Proposition 4, 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 resolution policy which serves to remove insolvent banks from the market. For example, one could allow policymakers to scrutinize banks' assets at date 1 and force banks with $\theta < 0$ into resolution, which would unambiguously improve welfare.

F Online Appendix: Additional results for Section 4

F.1 Robustness of Proposition 5 to Perturbations

Let $\mathcal{O}$ be the set of all $N \times N$ outside signals $\Pi(t)$ that are continuous in $t \in [0,1]$, lower triangular and obey MLRP.[39] Note that $\mathcal{O}$ is a non-empty set - indeed, the signal constructed in the proof of Proposition 1 is in $\mathcal{O}$.

Here we show that the conclusions of Proposition 1 are robust on open subsets of $\mathcal{O}$ - in particular, the nature of the discontinuity implies that 'small perturbations' of outside signals are still consistent with collapses in equilibrium disclosures – even when full disclosure is a strict equilibrium for most types of Sender at $t^\star$. In particular, Proposition 1 can be generalized to the following:

Proposition. ***(A1)****Suppose $v(\alpha(s))$ is increasing in the MLRP order, and $c \leq V(\theta_2) - V(\theta_1)$. For any $\epsilon > 0$, there exists an open set $\mathcal{O}_\epsilon \subset \mathcal{O}$ the following properties:*

- $\Pi(0)$ *is pure noise, while* $\Pi(1)$ *is fully revealing,*
- *There exists critical points $t_1^\star \leq t_2^\star \in (0,1)$ such that, when Receiver observes the signal induced by* $\Pi(t)$*, full disclosure is an equilibrium for $t \leq t_1^\star$ and is a strict equilibrium at $t_2^\star$, while full opacity is the unique equilibrium for $t > t_2^\star$.*

Moreover, as $\epsilon \to 0$, $t_1^\star \to t_2^\star$.

Proof. We first construct a generalization of the signal path in Proposition 1 which is in the interior of $\mathcal{O}$. Let $p_i : [0,1] \to [0,1]$ be a C^2, strictly increasing function with $p_i(0) = 0$, $p_i(1) = 1$ and whose derivative is equicontinuous, for $i = 1, \ldots, N$. Iteratively define the following class of outside signals: let $\tilde{\Pi}_\omega(t)$ be an $N \times N$ matrix whose elements are

$$\tilde{\pi}_\omega(s \mid \theta_i; t) \;\; = \begin{cases} (1 - p_i(t)) \frac{\tilde{\pi}_\omega(s|\theta_{i-1};t)\omega^{i-s-1}}{\Omega_i(t)}, & s < i \\ p_i(t), & \text{for } s = i \\ 0, & \text{for } s > i \end{cases}$$

for some $\omega < 1$, where $\Omega_i(t)$ is chosen so that $\sum_{s=1}^{i} \tilde{\pi}(s \mid \theta_i; t) = 1$, $\forall i$. In particular, note that $\tilde{\Pi}(t)$ is everywhere lower-triangular and satisfies MLRP with everywhere strict inequality.

[39] The restriction to MLRP signals is not necessary for our results. We impose the restriction only to highlight that the result goes through for this common class of signals.

First, we show that as $\omega \to 1$, $\mathcal{M}_t^\omega(\theta_i)$ converges uniformly to the decreasing function $\mathcal{M}_t(\theta_i)$. Consider

$$\mathbb{E}\left[V(\theta_i) - V(\theta_s) \mid s < k\right] = (1 - \rho_k(t))\left(\mathbb{E}\left[V(\theta_i) - V(\theta_s) \mid s < k-1\right]\right) + \rho_k(t)\left(V(\theta_i) - V(\theta_k)\right)$$

where $\rho_k(t) = \frac{p_k(t)}{p_k(t)+\sum_{i\leq k} y_i(t)}$, $y_i(t) = \omega^{1/2(k-i)(k-i+1)} p_i(t)\prod_{j>i}(1 - p_j(t))$. Notice that, by continuity of $p_k(t)$, $\forall t \in [0,1]$, $\mathbb{E}\left[V(\theta_i) - V(\theta_s) \mid s < k\right]$ is continuous in t on $[0,1]$. Since $\rho_k(t)$ is decreasing in ω, for all t, an inductive argument analogous to the proof of Proposition 1 shows that $\mathbb{E}\left[V(\theta_i) - V(\theta_s) \mid s < k\right]$ is increasing in ω, for all $k \in \{1, \ldots, i\}$, $t \in [0,1]$. But

$$\mathcal{M}_t^\omega(\theta_i) = \mathbb{E}\left[V(\theta_i) - V(\theta_s) \mid s < k\right]$$

Thus, $\mathcal{M}_t^\omega(\theta_i)$ is continuous on the bounded domain $t \in [0,1]$ and everywhere monotone in ω. By Dini's Theorem, $\mathcal{M}_t^\omega(\theta_i)$ converges uniformly to its pointwise limit $\mathcal{M}_t(\theta_i)$ as $\omega \to 1$. Thus, for any $\frac{\epsilon}{2}$, $\exists \omega_\epsilon < 1$ such that $\sup\left|\mathcal{M}_t^\omega(\theta_i) - \mathcal{M}_t(\theta_i)\right|$ for all $\omega_\epsilon \leq \omega \leq 1$.

Given $\tilde{\Pi}_\omega(t)$, consider the set of all continuous, lower-triangular $\Pi(t)$ s.t. $\sup\left|\Pi - \tilde{\Pi}_\omega\right| < \delta$, for some $\delta > 0$. For any $\omega < 1$, $\exists \delta_\omega$, all such $\Pi \in \mathcal{O}$ for all $\left|\Pi - \tilde{\Pi}_\omega\right| < \delta_\omega$. To show this, we need only establish that for such δ_ω, MLRP holds for all Π and $t \in [t_l, t_h]$. For any s, θ_i such that $s > i$, the relation continues to hold trivially. For all $s \leq i$, $\theta_i \in \Theta$, $\pi(s \mid \theta_i; t)$ is continuous in t for any such Π, and bounded away from 0. Thus, the same holds true for any likelihood ratio

$$r_i^t(s', s) = \frac{\pi(s' \mid \theta_i; t)}{\pi(s \mid \theta_i; t)}$$

where s, $s' \leq i$ and in particular for $\Pi(t) = \tilde{\Pi}_\omega$ the minimum

$$\min_{t, i > j, s' > s}\left|\tilde{r}_i^t(s', s) - \tilde{r}_j^t(s', s)\right| = b$$

exists and is bounded strictly above 0. Moreover, there exists δ_ω such that for all $\delta \leq \delta_\omega$, $\overline{r} := \frac{\tilde{\pi}_\omega(s'|\theta_i;t)+\delta}{\tilde{\pi}_\omega(s|\theta_i;t)-\delta}$ and $\underline{r} := \frac{\tilde{\pi}_\omega(s'|\theta_i;t)-\delta}{\tilde{\pi}_\omega(s|\theta_i;t)+\delta}$ can be everywhere bounded such that[40]

$$\sup\left|\overline{r} - \tilde{r}_i\right| \leq \frac{b}{3}$$

For any such $\delta \leq \delta_\omega$, the order of all likelihood ratios must therefore remain the same as under $\tilde{\Pi}_\omega$ for any Π such that $\left|\Pi - \tilde{\Pi}_\omega\right| < \delta_\omega$.

[40] For instance, setting $\delta_\omega = \min k \cdot \tilde{\pi}_\omega(s \mid \theta_i; t)$ for some k. We can make sure that all the fractions differ by no more than $\left|\left(\frac{1+\delta_\omega}{1-\delta_\omega}\right) - 1\right| \max r_i^t(s', s)$, which can be bounded uniformly below b by taking $k \to 0$.

Now consider the maximal punishment for some outside signal Π, $\mathcal{M}_t'(\theta_i)$:

$$\mathcal{M}_t'(\theta_i) = \sum_{s=1}^{i-1} \pi(s \mid \theta, t)(V(\theta_i) - V(\theta_s))$$

Since $\mathcal{M}_t'(\theta_i)$ is an average of bounded values, for any $\epsilon > 0$, there exists a $0 < \delta^\star \leq \delta_\omega$ such that

$$|\mathcal{M}_t'(\theta_i) - \mathcal{M}_t^\omega(\theta_i)| \leq \frac{\epsilon}{2}$$

Thus, $\forall \omega_\epsilon \leq \omega \leq 1$ and Π such that $\left|\Pi - \tilde{\Pi}_\omega\right| < \delta^\star$, we have $\Pi \in \mathcal{O}$ and (by the triangle inequality) $|\mathcal{M}_t'(\theta_i) - \mathcal{M}_t(\theta_i)| \leq \epsilon$.

Given the above, it is easy to verify that for all $\epsilon > 0$ sufficiently small the steps in the proof of Proposition 1 can be applied to establish the claims in Proposition A.F.1 for any corresponding $\mathcal{M}_t'(\theta_i)$, where $t_2^\star$ is the smallest t' at which $\mathcal{M}_t'(\theta_i) \leq 0$ for all $t \geq t'$ for some $\theta_i \in \Theta$. □

F.2 Extending Proposition 6 to Outside Signals with Full Support

Here we explain when the proof of Proposition 2 can be extended to signal distributions that have full support. The key requirement for the proof of Proposition 2 to extend to full support signals is that the Receiver's optimal action be continuous in posterior beliefs (as defined below).

Consider the induced posteriors from outside signal (S, Π) (not conditioned on equilibrium disclosures), which generates posterior $\hat{\mu}_s := \Pr(\theta \mid s) \in \Delta\Theta$ with probability $\hat{\tau}_s \in [0, 1]$ and satisfies

$$\sum_s \hat{\tau}_s \hat{\mu}_s = \mu_0.$$

Similarly, $(S \cup \Theta, \Pi')$ generates a lottery $(\tau_s')_{s \in S \cup \Theta}$ over some posterior distributions $\mu_s' \in \Delta\Theta$ which satisfies $\mu_\theta' = \mathbf{1}_\theta$ for any $s \in \Theta$,

$$\sum_{t \in S \cup \Theta} \tau_s' \mu_t' = \mu_0.$$

Moreover, because of the two-stage signal structure of $(S \cup \Theta, \Pi')$, it induces a mean-preserving spread over beliefs induced by (S, Π): that is each posterior $\hat{\mu}_s$ can be expressed as

$$\tau_s^s \mu_s' + \sum_{t \in \Theta} \tau_t^s \mathbf{1}_t = \hat{\mu}_s,$$

for some $(\tau_t^s)_{t \in \{s\} \cup \Theta}$ satisfying $\sum_{t \in \{s\} \cup \Theta} \tau_t^s = 1$, $\tau_t^s \geq 0, \forall t$. For each $s \in S$, let $\frac{\sum_{t \in \Theta} \tau_t^s \mathbf{1}_t}{1 - \tau_s^s} := \phi_s$

and notice that $\phi_s \in \Delta\Theta$.

Consider now an alternative collection of posterior beliefs

$$\left\{\mu'_s, \{\beta\left(\alpha\mathbf{1}_t + (1-\alpha)\phi_s\right) + (1-\beta)(1-\alpha)\mu'_s\}_{t\in\Theta}\right\}_{s\in S}$$

for $0 \le \alpha, \beta \le 1$. Notice that this collection of posteriors can be written as a MPS of (S, Π) – for each s, letting $\gamma_s + (1-\gamma_s)(1-\beta) = \frac{\tau^s_s + \beta - 1}{\beta}$, we can use the conditional weights

$$\begin{aligned}
\gamma_s\mu'_s + \sum_{t\in\Theta}\frac{(1-\gamma_s)\tau^s_t}{1-\tau^s_s}\left[\beta\left(\alpha\mathbf{1}_t + (1-\alpha)\phi_s\right) + (1-\beta)(1-\alpha)\mu'_s\right] &= \\
\left[\gamma_s + (1-\gamma_s)(1-\beta)\right]\mu'_s + \beta(1-\gamma_s)\sum_{t\in\Theta}\frac{\tau^s_t}{1-\tau^s_s}\left(\alpha\mathbf{1}_t + (1-\alpha)\phi_s\right) &= \\
\left[\gamma_s + (1-\gamma_s)(1-\beta)\right]\mu'_s + \beta(1-\gamma_s)\phi_s &= \hat{\mu}_s
\end{aligned}$$

For β close enough to 1, $\gamma_s \in (0,1)$ such that these weights are indeed feasible for all $s \in S$. Integrating back from $\hat{\mu}_s$ using $\hat{\tau}$ establishes that a lottery $\tau'' \in \Delta\Delta\Theta$ over $\left\{\mu'_s, \{\beta\left(\alpha\mathbf{1}_t + (1-\alpha)\phi_s\right) + (1-\beta)(1-\alpha)\mu'_s\}_{t\in\Theta}\right\}_{s\in S}$ exists when the prior is μ_0.

Therefore, from Proposition 1 in Kamenica and Gentzkow (2011) there exists a signal structure $(S \cup \Theta, \Pi'')$ that generates posterior lottery τ''. Moreover, as we established above $(S \cup \Theta, \Pi'')$ induces beliefs that constitute a MPS of those induced by signal structure (S, Π). Therefore, $(S \cup \Theta, \Pi'')$ is strictly more informative than (S, Π) (Blackwell (1953)). Finally, the new signal structure has full support whenever (S, Π) has full support (since μ'_s must put strictly positive weight on all $\theta \in \Theta$).

Now, if $\alpha(s)$ is continuous in μ, then as $\alpha, \beta \to 1$ the net payoffs to disclosure in an opaque strategy with outside signals $(S \cup \Theta, \Pi'')$ must limit to their value under outside signal $(S \cup \Theta, \Pi')$. Since all agents have a strict incentive to play $m = \emptyset$ in this limit, there must exist $\underline{\alpha}, \underline{\beta} < 1$ such that opacity is an equilibrium with outside signals $(S \cup \Theta, \Pi'')$ for all $\underline{\alpha} \le \alpha \le 1$, $\underline{\beta} \le \beta \le 1$.

Finally, we need to show that the opaque equilibrium outcome with outside signals $(S \cup \Theta, \Pi'')$ is less informative than the initial equilibrium under (S, Π). In fact, it is simpler to compare equilibrium informativeness under $(S \cup \Theta, \Pi'')$ and $(S \cup \Theta, \Pi')$ respectively. Since both signals induce opaque equilibria, we need only compare the informativeness of the signals directly. It is simple to verify from the above that the posterior beliefs following signal $(S \cup \Theta, \Pi')$ form a MPS over those induced by $(S \cup \Theta, \Pi'')$ Reapplying Blackwell (1953), $(S \cup \Theta, \Pi'')$ is strictly less informative than $(S \cup \Theta, \Pi')$. Appealing to the proof of Proposition 2, it is therefore also less informative than the equilibrium $\boldsymbol{\zeta}^\star$ with information structure (S, Π).

F.3 On Robustness to Broader Message Spaces

In this section, we briefly describe how the results in the main text can be extended to broader classes of verifiable message spaces, so long as the marginal costs of finer disclosures are not too large.

Suppose that we adapt the model of Section 3 as follows: given type θ_i, Sender may now choose a message m from a (finite) set $\mathbb{M}(\theta)$, with the properties that for each i, $\emptyset \in \mathbb{M}(\theta)$ and moreover there exists a non-empty subset $\underline{\mathbb{M}}_i \subset \mathbb{M}(\theta_i)$ such that $\mathbb{M}(\theta_j) \cap \underline{\mathbb{M}}_i = \emptyset$, $\forall j < i$. The first assumption ensures the existence of at least one unverifiable message, i.e. one that can be sent by all types. Call this set of messages $\mathbb{M}^c$. We write $\mathbb{M} := \cup_{\theta \in \Theta} \mathbb{M}(\theta)$ The second assumption ensures that verifiable disclosures are possible – in particular, any type θ_i can always prove that his type is at least θ_i. This message structure allows for among others, the all-or-nothing disclosures in the main text, message structures that form nested intervals, $\mathbb{M}_i \subsetneq \mathbb{M}_j$, for all $i < j$, $i, j \in \{1, \ldots, N\}$ as well as the classic *true assertions* disclosure strategies of Milgrom and Roberts (1986) in which types can send any subset $\mathbb{A}_i \in 2^\Theta$ satisfying $\theta_i \in \mathbb{A}_i$. For the sake of brevity, we assume here that all types θ_i, θ_j share at least one outside signal with positive probability.[41]

To each message $m_i \in \mathbb{M}_i$, we assign a disclosure cost $c_i(m_i) \geq 0$, which type θ_i pays if he chooses to send m_i. To capture the idea that finer disclosures are costly at the margin, we assume that the cost function is weakly decreasing in the number of types for whom the signal is available, $|\Theta_{m_i}|$, where $\Theta_{m_i} := \{\theta_j : m_i \in \mathbb{M}_j, j = 1, \ldots, N\}$. Notice that this implies unverifiable messages are 'cheap talk' – $m = \emptyset$ is the cheapest message available to any type. For the sake of notational ease, normalize $c_i(\emptyset) = 0$, $\forall i$.

For any $m \in \mathbb{M}_i$, we can now define an m-dependent maximal punishment (including disclosure costs as) as

$$\mathcal{M}(\theta, m) := \sum \pi(s|\theta) \left[V(\theta) - V(\underline{\theta}(s, m))\right]$$

where $\underline{\theta}(s, m) := \min \left\{\tilde{\theta} : s \in S\left(\tilde{\theta}\right) \cap \mathbb{M}\left(\tilde{\theta}\right)\right\}$. This extends the maximal punishment from

[41] This was true in all the main constructions we made to prove Propositions 1 and 2, so does not come at much incremental cost. In any case, the arguments that follow here continue to go through without this assumption under MLRP, at the cost of additional notation. Essentially, one must redefine $\underline{\mathbb{M}}_i$ to include messages that might be sent any lower type θ_j for which $S(\theta_i) \cap S(\theta_j) = \emptyset$. Type i can reasonably select some such message in equilibrium, saving on costs and still facing the 'maximal punishment' θ_i. Moreover, for such messages the expected posteriors θ_j faces after such messages are as if θ_i could did not choose a message in $\mathbb{M}(\theta_j)$. MLRP ensures higher types' expected payoffs from such messages do not lie above the maximal punishment.

the main text to reflect the worst case inference Receiver can make on observing (m, s) which can involve less skepticism than following the pair $(\emptyset, s)$. Notice that for each θ_i and message $m \in \underline{\mathbb{M}_j} \cap \mathbb{M}(\theta_i)$, $j < i$, the maximal punishment, $\mathcal{M}(\theta_i, m)$, depends only on θ_i and θ_j. Thus, with some abuse of notation we can simply write maximal punishments as $\mathcal{M}(\theta_i, \theta_j)$ a function of the Sender's type and the minimal θ_j consistent with message m. Similarly, across all messages $m \in \underline{\mathbb{M}_j} \cap \mathbb{M}(\theta_i)$, we can define the least costly one to θ_i as $\underline{c}_i(\theta_j) := \min_{m \in \underline{\mathbb{M}_j} \cap \mathbb{M}(\theta_i)} c_i(m)$.

More broadly, we now explicitly extend Receiver's best response given equilibrium disclosure m, and signal s, as function $\alpha(s, m)$, which is understood to depend implicitly on equilibrium disclosure strategies, where given a strategy profile $\sigma \in \times_i \Delta\mathbb{M}(\theta_i)$, $\alpha(s, m) \in \arg\max_{a \in A} \mathbb{E}^{\sigma}[u(a, \theta) \mid m, s]$.

Finally, since several of the results in the main text refer to (non)-monotonicity of disclosure strategies, we need an appropriate definition of monotonicity in this broader setting that captures the tendency of types to produce some evidence:

Definition. (A1) Sender's message strategy is: *(i) monotone (increasing)* if $\Pr(m \notin \mathbb{M}^c \mid \theta)$ is an increasing function of θ; *(iii) non-monotone* if $\Pr(m \notin \mathbb{M}^c \mid \theta)$ is non-monotone in θ; and *(iii) opaque* if $\Pr(m \notin \mathbb{M}^c \mid \theta)$ otherwise.

With this structure in hand, the analysis of Section 3 and 4 extends straightforwardly so long as disclosure costs are not 'too steep' across verifiable disclosures. For any θ_i, and $\theta_h > \theta_l > \theta_1$ suppose the cost function satisfies

$$\underline{c}_i(\theta_h) - \underline{c}_i(\theta_l) < \mathcal{M}(\theta_i, \theta_l) - \mathcal{M}(\theta_i, \theta_h). \tag{28}$$

Equation (28) states that the incremental cost of a disclosure that identifies Sender as at least θ_h (rather than θ_l) is always smaller than the associated reduction in maximal punishment. Notice that, under this condition, any equilibrium of the game in Section 3 remains an equilibrium with more general messages. Indeed it is easy to verify that for any equilibrium in a type plays $m = \theta_i$ with positive probability in the model of Section 3, there is an equilibrium of the broader game in which θ_i sends some message $m \in \underline{\mathbb{M}_i}$ with corresponding probability (and $m = \emptyset$ otherwise) and all off-path messages are sustained by skeptical beliefs. Therefore:

Corollary. ***(A1)*** *Propositions 1 and 2 extend immediately under (28) and Definition A1.*[42]

[42] Equation (28) is also necessary for the equilibria in the text to go through unchanged. For instance, if it does not hold for two types θ_i, θ_j, $i > j$, there is no equilibrium in which i ever plays a message in $\mathbb{M}_i$. Instead, he would always prefer to take an action in $\mathbb{M}_j/\mathbb{M}_i$. In this case, equilibria will be semi-pooling, in the sense that some disclosing types will be happy to 'pool' on verifiable messages available to them both.

The statement of Proposition 1 in the text makes the stronger claim that there is a signal path $\Pi(t)$ such that after some time $t^{\star} < 1$, the opaque strategy is the *unique* equilibrium of the game. However, with the wider message space available above we introduce the possibility of new equilibria. For example, type θ_N might be happy to send some $m \in \mathbb{M}_{N-1} \cap \mathbb{M}(\theta_N)$ and be 'pooled' with type θ_{N-1}. As t increases, we might therefore find that types prefer to move from making full disclosures to cheaper, semi-pooling verifiable messages. Because these intermediate disclosures cannot be copied by all types, in general the discontinuity result of Proposition 1 may be less stark.

However, so long as verifiable disclosures involve high fixed costs and low marginal costs, it turns out that the same strong discontinuity result of Proposition 1 continues to hold in the more general setting:

Lemma. ***(A1)*** *There exists $\varepsilon > 0$ such that if $\underline{c}_i(\theta_h) - \underline{c}_i(\theta_{h-1}) < \varepsilon$, $\forall i \in \{3, \ldots, N\}$, $3 \leq h \leq i$, then the conclusions of Proposition 1 hold in the extended game with message spaces, $\mathbb{M}(\theta)$, $\theta \in \Theta$.*

Proof. We argue here that the signal $\Pi(t)$ we constructed in Proposition 1 uniquely induces an opaque equilibrium at $t^{\star} + dt$, for all dt sufficiently small. Suppose not. Then at time $t^{\star}$ there is an equilibrium in which some type θ_i, $i \in \{1, 2, \ldots, N\}$ optimally chooses a message $m' \in \underline{\mathbb{M}}_k$, for some $k \in \{2, \ldots, N\}$. For ε sufficiently small, it is easy to see that in any such equilibrium there is some such i and some type θ_h, $k \leq h \leq i$, who prefers to choose m' over any alternative in $\mathbb{M}(\theta_h)$. Otherwise, net payoffs would satisfy

$$\begin{aligned}
\mathbb{E}\left[V\left(\alpha\left(s, m'\right)\right) \mid \theta_h\right] - \underline{c}_i(\theta_k) &= V(\theta_h) + \sum_{s \in S(\theta_h) \cap S(\theta_l)} \pi(s \mid \theta_h)\left(V(\theta_i) - V(\theta_h)\right) - \underline{c}_i(\theta_k) \\
&\geq V(\theta_h) - \underline{c}_i(\theta_k) - N\varepsilon \\
&\geq \mathbb{E}\left[V\left(\alpha(s, m)\right) \mid \theta_h\right] - \underline{c}_i(\theta_k)
\end{aligned}$$

for all $m \in \mathbb{M}(\theta_h)$. Recalling that any two types share signals with strictly positive probability at $t^{\star}$, we can clearly find such an ε (S, Θ are finite).

But since α is strictly increasing in the MLR order, θ_h's payoffs in such an equilibrium strictly exceed $V(\theta_h) - \underline{c}_i(\theta_h)$. Therefore, type θ_h never plays a separating message in equilibrium. That is, $\forall m \in supp\ \sigma(\theta_h)$, there exists at least one θ_j, $j \in \{1, \ldots, N\}$ such that $m \in supp\ \sigma(\theta_j)$. For each $m \in supp\ \sigma$, denote the lowest type who sends such a message in equilibrium by $\theta(m)$. Now, for all $\forall m \in supp\ \sigma(\theta_h)$, $\theta(m) \leq \theta_h$. If $\theta(m) = \theta_k < \theta_h$ for any such m, then by the same argument as above, θ_k must never play a separating message in equilibrium. Iterating the process, we find some θ_l, $l \geq 2$, for which either *(i)* all $m \in supp\ \sigma(\theta_l)$ are pooling with other types and $\theta_l = \theta(m)$, for all $m \in supp\ \sigma(\theta_l)$, with *supp* $\sigma(\theta_l) \cap \mathbb{M}^c = \emptyset$, or *(ii)* $\theta_l = \theta(m)$, for all $m \in$ *supp* $\sigma(\theta_l) / \mathbb{M}^c$ and *supp* $\sigma(\theta_l) \cap \mathbb{M}^c \neq \emptyset$. In case *(i)*, there must exist some message $\underline{m} \in$ *supp* $\sigma(\theta_l)$ for which $\Pr(m = \underline{m} \mid \theta_l) \geq \frac{1}{|\mathbb{M}|}$ and $\Pr(m = \underline{m} \mid \theta_p)$ for some $\overline{\theta}(\underline{m}) \geq \theta_l$, where $\overline{\theta}(m) = \max\{\theta : m \in \textit{supp } \sigma(\theta), \theta \in \Theta\}$. For such a message and a signal $s \in S(\theta_l) \cap S\left(\overline{\theta}(m)\right)$, we must have

$$\frac{\mu_s^l}{\overline{\mu}_s} \geq \frac{\pi(s \mid \theta_l, t^\star)}{\pi\left(s \mid \overline{\theta}(m), t^\star\right)} \frac{\mu_0^l}{\overline{\mu}_0} \frac{1}{|\mathbb{M}|} > 0.$$

Since S, Θ are finite, any two types share a signal with strictly positive probability at $t^\star$ and the prior takes full support on Θ, the above inequality can be uniformly bounded away from 0 by

$$\min_{i,j,s \in S(\theta_i) \cap S(\theta_j)} \frac{\pi(s \mid \theta_i)}{\pi\left(s \mid \overline{\theta}(m)\right)} \min_{i,j} \frac{\mu_0^i}{\mu_0^j} \frac{1}{|\mathbb{M}|} > 0$$

Thus, since V is strictly increasing in the MLR order, the (direct) expected pooling cost to type $\overline{\theta}(m)$ is bounded away from 0 by some $\eta > 0$:

$$V\left(\overline{\theta}(\underline{m})\right) - \mathbb{E}\left[V(\alpha(s, \underline{m})) \mid \theta_h\right] > \eta.$$

Therefore, for $N\varepsilon < \eta$, there can be no such equilibrium, since $\overline{\theta}(\underline{m}) = \theta_i$ would always prefer to deviate from playing $\underline{m}$ to some $m \in \mathbb{M}_i$.

Alternatively, in case *(ii)*, a similar argument establishes that θ_l either plays some $m \in$ *supp* $\sigma(\theta_l) / \mathbb{M}^c$ or some $m' \in \mathbb{M}^c$ with probability at least $\frac{1}{|\mathbb{M}|}$. If this is true for m, then the same argument above rules out any other equilibrium. If on the other hand, type θ_l plays some $m' \in \mathbb{M}^c$, then one can apply the same argument made in the proof of Proposition to show that the payoff to cheap talk messages strictly increases for all players. With the appropriate choice of Π, we can find δ small enough that this change induces a dominant strategy for type θ_N to play messages in $\mathbb{M}^c$, for ε small enough. All types can then be shown to have an iterated dominant strategy to play $m \in \mathbb{M}^c$. □