Inside and outside information

Daniel Quigley and Ansgar Walther[1]

This draft: January 2018

[1]Quigley: Nuffield College, University of Oxford, Oxford OX1 1NF, UK. Email: daniel.quigley@economics.ox.ac.uk. Walther: Warwick Business School, University of Warwick, Coventry CV4 7AL, UK. Email: ansgar.walther@wbs.ac.uk. We thank Heski Bar-Isaac, James Best, Vince Crawford, Peter Eso, Nicolas Inostroza, Ian Jewitt, Emir Kamenica, Alex Kohlhas, Paul Klemperer, Tim McQuade, Nadya Malenko, Meg Meyer, Ines Moreno de Barreda, David Myatt, Larry Samuelson, Joel Shapiro, Anton Tsoy, Victoria Vanasco, Adrien Vigier, Selma Walther, Lucy White, 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 strategic disclosures of *inside information* when there is also *outside information,* which is beyond insiders' control. A central application is the equilibrium effects of financial stress tests on banks' strategic disclosures about their asset quality. For a range of parameters, the classic 'unraveling' spiral works in reverse, and information becomes fragile: Small changes in the distribution of outside information trigger large reductions in inside disclosures. This implies that optimal stress tests must satisfy a minimum standard of transparency. Moreover, the importance of outside information hinges on the shape of insiders' payoffs, which yields new 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. Stress tests, with associated public disclosures about banks' health, have since become a core tool of financial regulation.[2] The effects of better public information are known to be complex, especially when public signals influence how agents respond to additional private information about economic conditions. A large literature evaluates these trade-offs when private information is dispersed among many agents.[3]

However, less attention has been paid to another important channel through which public information affects economic outcomes: 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 (Milgrom, 2008). 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.

In Section 2, we analyze a standard Sender-Receiver model of verifiable communication with binary responses. Following our leading example, consider a large bank (Sender) who privately observes the quality of its assets, and wants to persuade investors (Receiver) not to withdraw funds (run on the bank).[4] The bank can pay a cost to verifiably disclose its quality to investors,[5] who additionally observe outside information that the bank cannot control.[6]

To understand our main results, it is useful to recall the classic result that verifiable disclosures exhibit a form of strategic complementarity (Grossman and Hart, 1980). If the

[2]See Goldstein and Sapra (2014) and Leitner (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; Angeletos and Pavan (2007) evaluate the impact of public information on equilibria in coordination games.

[4]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).

[5]Lewis (2011) demonstrates the empirical relevance of disclosure costs in online auctions. Leuz and Wysocki (2016) survey a large body of research documenting that disclosures involve both technological and proprietary costs. Our main results continue to hold when disclosure costs are small, in the sense that they would not matter in a model without outside information.

[6]We focus on disclosures that are pre-emptive: When Sender (the bank) decides whether to disclose his quality, he cannot perfectly predict the realization of outside signals. We discuss the foundations of this assumption in Section 2.

best types of bank are expected to disclose, investors rationally assume that no news is bad news, which generates strong incentives for other types to also disclose. In models without outside information, this complementarity leads to *unraveling*: All but the worst types of bank disclose in equilibrium, as long as disclosure costs are small enough. In our model, by contrast, outside information opens the door to *reverse unraveling,* because the strategic complementarity now works in favor of opacity. If outside information is sufficiently precise, the best banks expect an accurate and favorable outside signal, and rationally stay quiet to save on the costs of disclosure. Investors now expect silence from the best, and no news is ambiguous news. Silence then becomes more attractive for the second-best types of bank. If they too stay quiet, no news becomes better news still, the third-best types are tempted to stay quiet, and so forth.

Because of reverse unraveling, outside information strongly crowds out inside information. Indeed, inside information becomes fragile: Small improvements in the quality of outside signals can lead us from equilibria with full disclosure to a discontinuous decline in the amount of inside information that is revealed. We show that discontinuities must arise along *any* continuous path of gradually improving outside signals under natural regularity conditions. Around equilibria involving limited disclosures, the relationship between inside and outside information is more nuanced. Here, better outside information tends to crowd out inside disclosures at the margin if prior beliefs about θ are pessimistic, but can crowd in disclosures if they are optimistic.Thus, there is a clear informational externality associated with outside information as it affects inside disclosures in equilibrium. The optimal design of outside information must take this into account.

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.[7] Therefore, both inside disclosures and outside stress tests appear to matter for resolving uncertainty in financial crises, and the evidence suggests that they interact. In Section 3, we concentrate on optimal stress testing policy in an application to financial panics. Financial crises create externalities between investors, providing a second-best rationale for imperfectly informative stress tests. We show that optimal policy in this context must exploit the informational externality by crowding out disclosures from the strongest banks. This is implemented by releasing stress test results that meet a *minimum standard of transparency*. Intuitively, this standard ensures that the strongest banks are confident enough to stay quiet. This gives weak but solvent banks an opportunity to pool

[7]Bank of England (Quarterly Bulletin Q4, 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) argues that stress-tested banks made fewer loan charge-offs and more frequently changed the classification of loan losses.

with the best, protects them from inefficient bank runs, and enhances welfare. Moreover, we show that if crowding out is sufficiently persistent, then optimal stress tests are always more transparent than they would be in the absence of inside disclosures.

This result is relevant to the growing literature on stress test design and information disclosure during financial crises, which we review in detail below. In short, we show that the Lucas critique has bite for informational policy. Stress test design must account for the endogeneity of inside information to minimize financial distortions. Beyond the application to financial crises, optimal policy is context-specific. For example, in the market for used cars studied by Akerlof (1970), the first-best informational outcome is full transparency. In this setting, policy-makers should avoid crowding out inside disclosures.

In Section 4, we consider a more general Sender-Receiver model where responses need not be binary. While a full characterization of equilibria is elusive in general, 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. Moreover, for any equilibrium where Sender discloses with positive probability, more informative outside signals can leave Receiver worse informed overall.

An additional insight provided by this more general setting 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 on corporate disclosures emphasizes managers' desires to keep stock prices high (Verrecchia, 1983; Acharya et al., 2011) and to enhance market liquidity (Diamond and Verrecchia, 1991). Our model implies, in addition to these factors, that capital structure and executive compensation play a key role in determining disclosure strategies.

Related literature

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

Grossman and Hart (1980), Grossman (1981), Milgrom (1981) and Milgrom and Roberts (1986) study disclosure of verifiable information without outside information or disclosure costs, and establish the classic unraveling result. Another strand of work shows that equilibria with limited disclosures arise when disclosure costs are significant (Verrecchia, 1983) or when it is uncertain whether Sender has any private information (Dye, 1985; Shin, 2003). We contribute by focusing on situations where little or no information is disclosed by insiders,[8] and by establishing that strategic complementarities can work in favor of non-disclosure. 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 in a different, dynamic model.

Feltovich et al. (2002) and Daley and Green (2014) study signaling games with outside information and two or three types of Sender.[9] 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. We go further by characterizing strategic complementarities and strong crowding-out effects, as well as highlighting the key role played by Sender's payoffs, in a setting with many types.[10]

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 credible signal available to investors. 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.[11] Another strand of research (e.g. Leitner and Yilmaz, 2016; Leitner and Williams, 2017) evaluates banks' incentives to misrepresent information by gaming regulatory models. We contribute to this

[8] Jin and Leslie (2003) provide empirical evidence of incomplete disclosure.

[9] A branch of the Accounting literature studies the interaction between verifiable financial reports and non-verifiable (cheap talk) communication by firms, see for example Gigler and Hemmer (1998). Einhorn (2017) considers the case with verifiable but imperfect communication by firms in the presence of outside information. The fragility of information and reverse unraveling, however, are unique to our setting.

[10] On a technical note, verifiable disclosure is a special but more tractable case of signaling, which allows us to derive new insights with many types.

[11] Philippon and Skreta (2012) and Tirole (2012) study auxiliary policy responses to 'lemons' problems in finance, such as asset purchase programs.

literature by analyzing the constraints that endogenous inside disclosures place on stress test design. Accounting for these constraints is important because, as we establish, a 'relaxed' optimal policy that ignores them can lead to discretely lower welfare and lower financial stability.

In Section 2, we study a Sender-Receiver game with binary actions, and apply it to financial crises and stress test design in Section 3. In Section 4, we analyze a more general class of games with many actions. In Section 5, we provide a further application to corporate disclosures when issuing debt and equity. Section 6 concludes.

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 the setup in terms of our leading example: The action a captures the aggregate decision of investors to run on their bank ($a = 0$) or not ($a = 1$). The bank can disclose verifiable information about the quality of its assets, and outside information is made available by policy-makers, for example, in the form of stress tests. In Section 3, we provide a micro-foundation of this interpretation.

We first describe the model and discuss our key assumptions. We then illustrate the reverse unraveling mechanism heuristically, before proving formally that information is fragile and underlining the importance of this mechanism by discussing equilibrium selection. Finally, we discuss whether outside information crowds out or crowds in disclosures.

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 $c(\theta) > 0$; $m = \emptyset$, which we refer to as *staying quiet*, 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 ideas. We return to more general message spaces in Section 4 and the associated Online Appendix F.1.

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): For every non-degenerate prior distribution of θ, the conditional expectation of any increasing function of θ is increasing in s.

Preferences Sender's utility is $a - c(\theta) \times \mathbf{1}_{m=\theta}$; he enjoys high actions but pays the cost of disclosure. Receiver's utility is $a(\theta - p)$, so that she chooses $a = 1$ if and only if, given any information set $\mathcal{I}$, she believes that $E[\theta \mid \mathcal{I}] \geq p$. Here, p parameterizes Receiver's propensity to take the low action.[12] This simple setup nests a much wider class of games with binary responses, subject only to the standard restrictions that Sender prefers high to low actions, and that Receiver prefers high actions only if θ is high.[13]

We focus on the non-trivial case, in which *(i)* $E[\theta] < p$, so that the prior mean is low enough that Receiver would choose $a = 0$ without further information, and *(ii)* $c(\theta) < 1$, so that disclosure costs are smaller than the benefits to Sender of inducing a high action.

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 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 $s \notin S(\theta')$. The latter refinement is natural, and common in the applied literature. Moreover, it is inconsequential when outside signals have full support (i.e. when $S(\theta) = S$ 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.

[12] 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.

[13] Suppose that Receiver's utility is $u(a, \theta)$, assuming only that the net benefit $\Delta(\theta) = u(1, \theta) - u(0, \theta)$ of the high action is increasing in θ. Without loss of generality, we can re-define Receiver's type as $\tilde{\theta} = \Delta(\theta) - p$, yielding a game that is equivalent to our setup. Furthermore, suppose that Sender's utility is $v(a, \theta)$, assuming only that $B(\theta) = v(1, \theta) - v(0, \theta) > 0$. 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 $\tilde{c}(\theta) = c(\theta)/B(\theta)$ consistently with our setup.

Discussion of pre-emptive disclosures In the model, Sender commits to make a disclosure m before he knows the realization s of outside information. This is important: One of our key intuitions is that the best types 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.

We focus on the case of pre-emptive disclosures because we believe that it captures empirically relevant frictions. In many applications, Receivers react very quickly to outside news, and irreparable damage to Sender's prospects may be done if he waits until after this event to prove his quality. For example, financial investors respond quickly to bad news or credit downgrades and managers may lose their job or reputation before they have a chance to respond.

This friction is especially relevant in situations where verifiable information takes time to prepare and circulate. In financial markets, reports need to be prepared and externally audited in advance of their release.[14] There can also be considerable delays in circulating inside information, for example via advertising campaigns, to reach a dispersed audience. More generally, if economic agents have limited capacity for processing information, 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.

Alternatively, pre-emptive disclosures can be motivated in a formally equivalent game, where outside information is privately observed by Receiver. For instance, investors might trade based on their own research, as well as disclosures by firms, and proprietary research does not always become public information. In this environment, Sender (e.g. the firm) must again decide on a communication strategy before he knows the realization of outside information.

Regularity condition In this and the next Section, we assume that the function

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

crosses zero at most once, and if so, it crosses from above.

[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 are likely to 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.

To interpret this condition, note that the function $J(\theta)$ compares two terms. The first term is the probability of receiving a public signal s in the left tail, given that the true state is θ. This is strictly decreasing in θ (by MLRP). The second term reflects the costs of disclosure. 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, or when outside information s is precise enough.[15]

2.1 The benchmark without outside information

Suppose Receiver had access to no outside information, and therefore had to rely exclusively on Sender's disclosures. Given our assumption that $E[\theta] < p$, it is easy to see that in any equilibrium the best type $\bar{\theta}$ of Sender must disclose ($m = \bar{\theta}$). 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 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*.

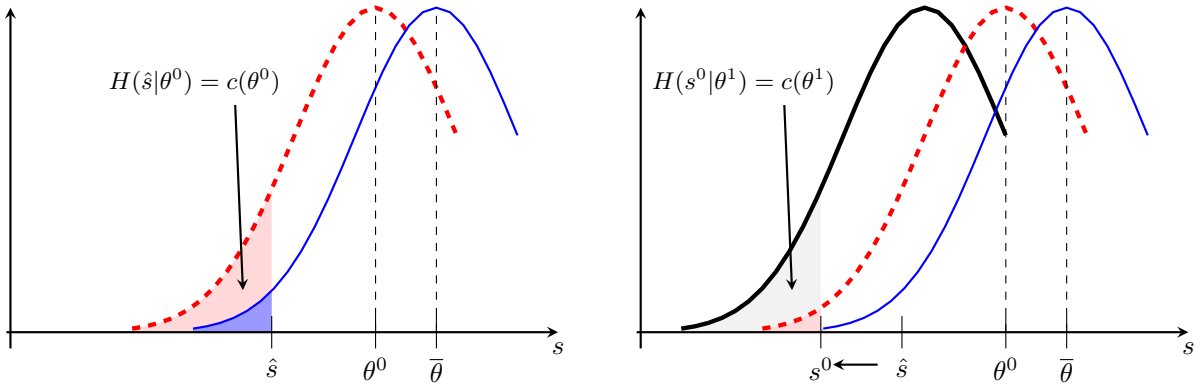
2.2 Fragile information: Reverse unraveling with bounded support

We begin with a heuristic illustration of reverse unraveling. For this Subsection, we focus on the case where 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 disclosures by all types $\theta \geq p$, the best type $\bar{\theta}$ of Sender 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.

[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.



(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 (dark) 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. Panel (b) shows that, as discussed in the text, types in $\left[\theta^1, \theta^0\right]$ have an iterated best response to stay quiet.

If condition (2) holds, then there must be an 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) best response to stay quiet – see Figure 1b.

Indeed, there are now *strategic complementarities in non-disclosures*. Since Receiver's posterior expectations at critical signal s^0 are exactly p, the decision by more good types $\theta \in (\theta^1, \theta^0] > p$ to stay quiet implies that signal s^0 makes $a = 1$ more attractive to Receiver. 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.

2.3 Fragile information: The general case

We return to the general case where outside signals may or may not have full support. Our first result establishes that, under our regularity condition (1), both Sender and Receiver optimally choose threshold strategies in any equilibrium:

Proposition 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 $\theta > p$, or if Sender stays quiet and the outside signal is $s > s^\star(\theta^\star)$, defined as the lowest outside signal s satisfying $E[\theta | \theta \notin (p, \theta^\star), s] \geq p$. If no such s exists, then $s^\star(\theta^\star) = \infty$.*

Thus, we can describe any equilibrium with a single parameter $\theta^\star$, which categorizes types of Sender into three regions. First, weak types $\theta < p$ always prefer to stay quiet, because disclosure would guarantee the low action. Second, strong types $\theta > \theta^\star$ are confident that they will draw a high enough s to trigger the high action, and stay quiet to save the costs of disclosure. Third, mediocre types $\theta \in (p, \theta^\star)$ can ensure the high action by disclosing, and are anxious to do so because they are not sufficiently confident about s.

Let $BR(\theta^\star)$ denote the highest type of Sender who prefers to disclose when Receiver expects disclosures from types $\theta \in (p, \theta^\star)$ – i.e. Sender's best response:[16]

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

with the convention that $BR(\theta^\star) = p$ if $H(s^\star(\theta^\star)|\theta) < c(\theta)$ for all θ. A cutoff $\theta^\star$ induces 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. Thus,

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

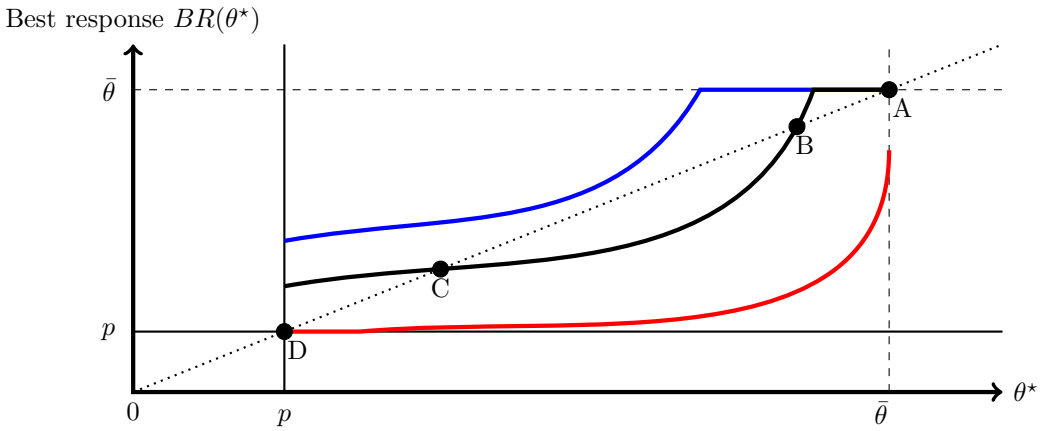


Figure 2: **Best response functions and equilibria.** The blue (upper) curve shows the best response when outside signals are imprecise; the unique equilibrium is unraveling at point A. The black (middle) curve is drawn for intermediate signal precision; there are multiple equilibria at A, B and C due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

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.

A key feature of our model is that strategic complementarity becomes very strong when only a small set of high-quality types stay quiet:

Lemma 1. *The best response function $BR(\theta^\star)$ is increasing in $\theta^\star$. Moreover, if $BR(\theta^\star) \in (p, \bar{\theta})$ in a neighborhood of the best type $\bar{\theta}$, then it is continuously differentiable in this neighborhood, and moreover, it becomes infinitely steep:*

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

This is the analytical equivalent of reverse unraveling. Strategic complementarities, as measured by the slope of $BR(\theta^\star)$, are infinitely strong at the top of the type distribution. When BR is interior near $\overline{\theta}$, then if a small set of the best types $\theta \in (\bar{\theta} - \epsilon, \bar{\theta}]$ stay quiet in equilibrium, Lemma 1 tells us that yet more types will become incentivized not to disclose, or $BR\left(\overline{\theta} - \epsilon\right) < \overline{\theta} - \epsilon$. We show that these strong complementarities follow because draws of s sent with positive probability by these high types are very unlikely to be shared with low types. As a result, even a small interval of non-disclosure at the top can make Receivers much more confident after high values of s, encouraging further silence at the top.

From this logic, we derive the fragility of information. We analyze the *most transparent* and *least transparent* equilibria, associated with the highest and lowest equilibrium disclosure

thresholds:

$$\theta^\star_{max} = \sup\{\theta \geq p : BR(\theta) = \theta\},$$
$$\theta^\star_{min} = \inf\{\theta \geq p : BR(\theta) = \theta\}.$$

This focus is natural. First, both $\theta^\star_{max}$ and $\theta^\star_{min}$ exist, by Tarski's fixed point theorem.[17] Second, we will argue below that $\theta^\star_{min}$ is a focal prediction of several natural selection criteria. 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.[18]

Theorem 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: $lim_{t\uparrow t_0}\theta^\star_{min} = \bar{\theta}$, but $\theta^\star_{min} < \bar{\theta}$ for $t = t_0$.*
- *The most transparent equilibrium is discontinuous around t_1: $\lim_{t\uparrow t_1}\theta^\star_{max} = \bar{\theta}$; 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 $\delta > 0$.*

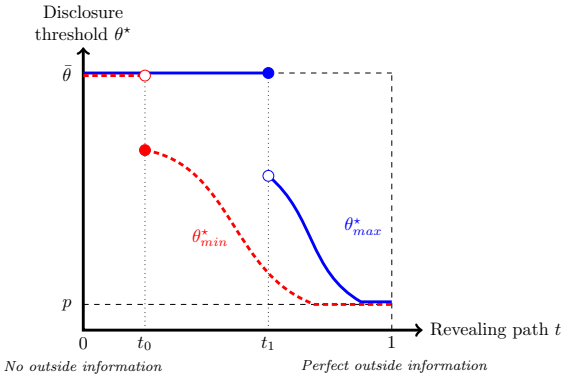
Figure 3 illustrates the result. Along a revealing path, when $t \simeq 0$, outside signals are almost pure noise, and the unique equilibrium is unraveling with $\theta^\star_{min} = \theta^\star_{max} = \bar{\theta}$. As outside signals improve,[19] we arrive at a threshold t_0 at which a less transparent equilibrium exists. This transition is not gradual: The Theorem 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 exists, and at this point the most informative equilibrium $\theta^\star_{max}$ has a downward jump. The discontinuity at t_0 is general and occurs for all revealing paths. By contrast, the second threshold t_1 is interior only if signals do not have full support – see Figure 3b.

Intuitively, the discontinuity 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,

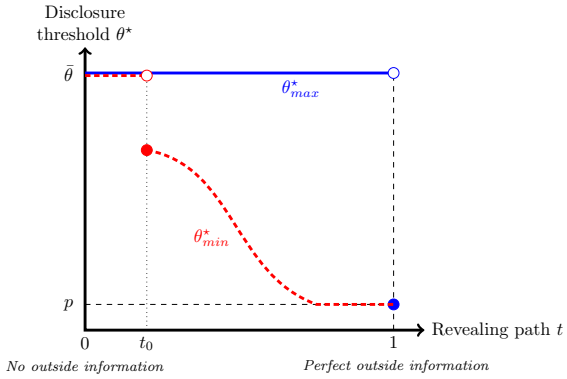
[17] The type space is a compact subset of $\mathbb{R}$ (and hence a lattice), and BR is monotone in θ.

[18] 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) $\lim_{t\to 1} h(s|\theta;t) = \delta(\theta)$ for all θ, where δ is the Dirac delta. Smoothness here means that h is continuously differentiable in t.

[19] We do not formally require that signals monotonically improve everywhere along a revealing path. Since Theorem 1 holds for *any* revealing path, it is trivially also true for paths along which s_t becomes more informative (e.g. in a Blackwell sense) as t increases.



(a) Outside signals without full support



(b) Outside signals with full support

Figure 3: **Fragility of information.** Panel (a) shows a revealing path of outside signals s_t which do not always have full support. Both the most transparent equilibrium $\theta^\star_{max}$ (the blue / solid line), associated with disclosures by types $\theta \in [p, \theta^\star_{max}]$, and the least transparent equilibrium $\theta^\star_{min}$ (the red / dashed line) 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 $t = t_0$. Of course, the equilibrium *correspondence* is nonetheless upper hemi-continuous.

this leads to a contradiction: If only a small set of the best types prefer to stay quiet, then by Lemma 1, 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.

2.4 Selecting an equilibrium

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. Moreover, when we apply these refinements, information is fragile regardless of whether outside signals have full support.

First, the least transparent equilibrium is Pareto-dominant from Sender's perspective. In the least transparent equilibrium, an interval of good types $(\theta^\star_{min}, \bar{\theta}]$ prefer to stay quiet, and by revealed preference they are better off than in any equilibrium in which they disclose. Moreover, disclosures by types above p unambiguously impose negative externalities on other *non-disclosing* types of Sender. Thus, the Sender-preferred equilibrium is $\theta^\star_{min}$. A focus on Sender-preferred equilibrium is natural when Sender acts as a mechanism designer and can suggest equilibrium strategies subject to incentive compatibility constraints.

Second, the least transparent equilibrium is the unique *neologism-proof* equilibrium, in-

troduced in the context of cheap-talk games by Farrell (1993).[20] 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 $\mathcal{S} \subset \left[\underline{\theta}, \overline{\theta}\right]$ 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 outcome, by using a 'new' cheap talk 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. A more transparent equilibrium $\theta^\star > \theta^\star_{min}$ cannot survive the refinement. If it did, then by the revealed preference argument above, types $\theta \notin (p, \theta^\star_{min})$ would prefer $\theta^\star_{min}$ to be played, and therefore form a self-signaling set. The proof of Proposition 2 further establishes that the least transparent equilibrium allows no self-signaling.

Finally, we show in Online Appendix D that unraveling equilibria are often unstable under best response dynamics following small deviations from equilibrium play, which again suggests the selection of less transparent outcomes even when neologism-proofness is not required.

[20] Note that the Cho-Kreps 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.

2.5 Crowding in or crowding out?

We now evaluate the local interaction between outside information and inside disclosures, around equilibria with intermediate levels of disclosure $\theta^\star \in (p, \bar{\theta})$.

Comparative statics are particularly clean if we assume that types and outside signals are normally distributed.[21] 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: $c(\theta) = c$. 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.[22]

Proposition 3. *Suppose θ and s are normally distributed, as defined above, 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.

When public signals become more informative, Receiver places more weight on s relative to the prior μ. When μ is low, this makes her more optimistic at the margin. As long as $c \leq \frac{1}{2}$, Sender-type $\theta^\star$ also believes that drawing $s < s^\star(\theta^\star)$ is a left-tail event, and therefore becomes less likely as the signal distribution contracts. Both effects discourage disclosures, and we obtain a crowding out effect. Conversely, when μ is high, the re-weighting makes Receiver more pessimistic. Sender then becomes less confident and keener to disclose, so that we obtain crowding in. The statement of Proposition 3 deals with additional complications that arise from the endogeneity of $s^\star(\theta^\star)$ in this case, by simultaneously reducing the cost of disclosure in order to consider changes in prior beliefs while holding $\theta^\star$ constant.

3 Financial crises and stress test design

We now analyze a micro-founded model of banks and investors, which maps directly into the Sender-Receiver model of Section 2, but better highlights the interaction between bank runs, stress tests, and disclosures of banks' inside information in equilibrium. We then address

[21] This steps slightly outside the baseline model because types and signals have unbounded support.

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

the policy question that motivates our main application: How much information should a policy-maker such as the Fed release into a potentially panicked banking system?

Agents and timing A bank (Sender) interacts with a continuum of short-term risk-neutral investors (Receivers) and a policy-maker. Everybody is risk-neutral, and there are three dates $t \in \{0, 1, 2\}$.

Bank deposits and investments Our model of bank deposits and investments is taken from Morris and Shin (2000), who take as given the structure of bank assets and contracts with investors. 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.

At date 1, each investor can withdraw his investment early, in which case she is entitled to an immediate payment of one unit of cash, or roll over, in which case she is a residual claimant on the bank's assets at time 2. If a proportion $l \in [0, 1]$ of investors withdraw at time 1, then the value of the bank's assets at time 2 is reduced to $r - 2pl$ (when they eventually mature at time 2). The parameter $p > 0$ measures the degree of asset illiquidity. The linear liquidation technology, and the exogeneity of deposit contracts, make for a particularly clean analysis of bank runs in equilibrium. These assumptions can be relaxed, in principle, using the techniques of Goldstein and Pauzner (2005).

Returns and information The return on assets at date 2 is

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

The first return component θ is drawn from a smooth distribution with density $f(\theta) > 0$ on $[\underline{\theta}, \bar{\theta}]$. 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, $j \neq i$. We will consider the standard noiseless limit where both η and t^i collapse to zero.

This small deviation from common knowledge allows us to select a unique equilibrium in the coordination game among investors using global games. The additive-independent return specification, which follows Bouvard et al. (2015), avoids re-introducing multiplicity when there are additional public signals about θ. Note, moreover, that our analysis does not depend on the coordination motive. The same qualitative insights obtain in any model where investors have a reason (e.g. concerns about moral hazard in badly capitalized banks) to withdraw their funds when they are pessimistic about asset values θ.

Inside and outside information At date 1, the bank privately observes θ, and can send a message $m \in \{\theta, \emptyset\}$ that is observed by all investors. As before $m = \theta$ verifiably reveals θ but reduces the bank's utility by $c(\theta)$, while $m = \emptyset$ contains no verifiable information but is free to send. In addition, investors observe a noisy 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.

Preferences 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 abstract from 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 as in Diamond and Dybvig (1983).

Policy 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 at date 0 in order to maximize expected investor welfare. A low value of k can be interpreted as a more revealing stress test scenario. For ease of exposition, we assume that reducing the noise in outside signals is not costly for the policy-maker, but this is not crucial; our main results below remain qualitatively similar when we allow stress testing to be costly. In our model, the policy-maker has only one degree of freedom in determining the distribution of outside information. This parsimoniously represents a world in which policy-makers can influence, but not fully determine, the conditional distribution of s. Goldstein and Leitner (2017) and Inostroza and Pavan (2017) analyze the case of full information control using tools from Bayesian Persuasion.

Game timing At date 0, the policy-maker chooses k and commits to this choice. At date 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; then she decides whether to withdraw early. At date 2, returns are realized and claims settled.

Parametric assumptions Note that early liquidation of the bank's assets is socially costly (in the limiting case where the second return component $\eta \to 0$) whenever the bank's type $\theta \geq 0$. In this Section, we will focus on the case where all banks have positive continuation

value:

$$\underline{\theta} \geq 0$$

This assumption allows us to focus on the Diamond and Dybvig (1983) case, where banks are solvent but potentially illiquid. Allowing for insolvent banks is straightforward and generates some additional insights; we analyze this case in detail in Online Appendix E.2.

We impose the mild regularity condition from Section 2. 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, from above. In the banking context, this restriction is reasonable – 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, in which case we might think of $c(\theta)/\theta$ as approximately constant. This guarantees the single crossing condition since $H(s|\theta)$ is strictly decreasing by MLRP.

3.1 Equilibrium: Investors' choices and bank runs

We begin by analyzing investors' incentives at time 1. All investors observe the same outside signal s and inside message from the bank m. Therefore, they share a common posterior expectation over the first return component, $E_\mu[\theta|m, s]$. We solve for investors' equilibrium decision, as a function of these beliefs, using a standard global games argument (see Bouvard et al., 2015). In the noiseless limit a bank run occurs if and only if

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

This condition is intuitive: Investors run whenever they have pessimistic beliefs about fundamentals, and when bank assets are highly illiquid, i.e. when p is high.

The key innovation in our setting is that the quality of information available to investors responds to the bank's equilibrium disclosure m. Indeed, as we now describe, banks' endogenous disclosure decisions are both a meaningful driver of bank runs and affect the incentives of policy-makers to design their stress tests.

3.2 Banks' inside disclosure strategies and optimal stress tests

Equation (5) shows that investors collectively behave as the binary-action Receiver in Section 2. Thus, the analysis of equilibrium disclosures proceeds exactly as in Proposition 1, and

the bank discloses whenever $\theta \in (p, \theta^\star)$ for an endogenous equilibrium cutoff $\theta^\star$.

The policy-maker chooses k in anticipation of these equilibrium disclosure strategies. When there are multiple equilibria to the disclosure game for a given k, we focus our attention on the least transparent one. This choice is natural for three reasons. First, it is the policy-maker'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) has the obvious definition:

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

where $s^\star(\theta_k^\star; k)$ denotes the critical outside signal below which investors run on a quiet bank in equilibrium. The first integral measures continuation value of disclosing banks $\theta \in (p, \theta_k^\star)$, net of the costs of disclosure. The second integral measures the expected continuation value of quiet banks $\theta \notin (p, \theta_k^\star)$, who survive only if they draw an outside signal $s \geq s^\star(\theta_k^\star; k)$.

An optimal policy must take into account the *direct* effect of k on welfare, but also its indirect impact on the equilibrium threshold $\theta_k^\star$. 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). The key trade-off in these papers 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, which is inefficient. Some noise is optimal because it provides implicit insurance for weak banks; this is reminiscent of the Hirshleifer (1971) effect.

Instead of repeating the excellent existing analyses of this trade-off, we focus on what is new in our model, namely inside information. In other words, we seek a characterization of optimal stress tests that is robust to the Lucas critique when inside disclosures respond to policy changes.

3.2.1 A minimum standard of transparency

Based on the logic of Theorem 1, we show that it can never be optimal to introduce noise above a critical level k_0 into stress tests:

Corollary 1. *There exists a threshold $k_0 \in (0, \infty)$ such that the least transparent equilibrium is discontinuous at k_0: $\theta_k^\star = \bar{\theta}$ for all $k > k_0$, while $\theta_{k_0}^\star < \bar{\theta}$. Any optimal policy sets $k \leq k_0$. Moreover, welfare increases discontinuously when k crosses k_0 from above.*

When the noise $k > k_0$, 'unraveling' is the unique equilibrium, and weak banks $\theta < p$ fail with probability 1. When we set $k \leq k_0$, by contrast, a discrete mass of strong banks stays quiet due to reverse unraveling, which generates valuable implicit insurance for small banks. Strong banks are also (weakly) better off, by revealed preference. Therefore, $k \leq k_0$ is always optimal.

The threshold k_0 represents a minimum standard of transparency that any optimal stress tests must satisfy. Moreover, this minimum standard can represent a binding constraint for the regulator. It is easy to show circumstances (e.g. if disclosure costs $c(\theta)$ are low enough) under which a policy-maker who ignores inside disclosures would optimally set $k > k_0$. This insight is 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.

3.2.2 Optimal transparency with inside information

We now ask whether more generally, optimal stress tests are more transparent once the endogeneity of inside information is taken into account. This also allows for the case where the minimum standard of transparency is not binding.

To study this formally, consider a naive policy-maker who takes the bank's disclosure threshold $\theta^\star = \theta_n^\star$ as exogenously given, and maximizes welfare $W(\theta_n^\star; k)$, considering only the partial derivative with respect to k. A sophisticated, fully optimizing policy-maker instead maximizes $W(\theta_k^\star; k)$, considering the total derivative. For the most direct comparison, we endow the naive policy-maker with consistent beliefs. That is, at its optimal policy choice $k^\star$, the naive policy-maker's conjectured disclosures are exactly the equilibrium, $\theta_n^\star = \theta_{k^\star}^\star$.[23] When beliefs are correct, we emphasize the naive objective function by writing $W_n(k) \equiv W(\theta_{k^\star}^\star; k)$. To aid local arguments, we focus on the (interesting) interior case where $\theta_k^\star \in (p, \bar{\theta})$.

A naive policy-maker takes into account the direct trade-off between transparency and insurance discussed above. In addition, a sophisticated policy-maker 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

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

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 policy-maker'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' in the sense that $\theta_k^\star > \theta_{k^\star}^\star$ for all $k > k^\star$, and if the naive policy-maker's objective function satisfies the condition

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

then a sophisticated policy-maker 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 policy-maker chooses $k > k^\star$.

The first part of Proposition 4 considers the marginal effect on welfare of changing the noise k contained in stress tests relative to the naive policy-maker's optimum $k^\star$. It states that welfare can be improved locally by making stress tests marginally more precise ($\downarrow k$) if and only if this change crowds out disclosures by strong banks ($\downarrow \theta_k^\star$). This is true because crowding out generates more implicit insurance for weak banks, a first-order gain, while the direct effect of changing k has only a second-order welfare effect at the naive optimum. Conversely, with crowding in, reducing k locally harms welfare. The second part provides simple sufficient conditions under which this marginal analysis carries through to determine the optimal choice of stress test, even though the welfare function is not generally concave. The condition requires that crowding out effects are persistent, and that the 'naive cost' of selecting $k > k^\star$ (the left-hand side) exceeds the offsetting benefit to types in $[\theta_{k^\star}^\star, \theta_k^\star]$ from voluntarily switching to disclosure, which reduces their own exposure to the welfare cost of setting k too high. If (7) holds, the 'true' welfare cost of setting k high is even larger, due to the negative externalities imposed on quiet banks by the disclosures of types in $[\theta_{k^\star}^\star, \theta_k^\star]$.

A natural question is under which conditions on model primitives we can expect persistent crowding out. As indicated by Proposition 3, persistent crowding out is common in periods of crisis. For example, when θ, s are jointly normally distributed, persistent crowding out is guaranteed whenever $E[\theta] < p$ and $c < \frac{1}{2}$. Moreover, under these conditions, (7) is likely to hold: For instance, it is easy to show (Envelope Theorem) that (7) is satisfied locally around $k^\star$. Moreover, it continues to hold at large k. 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 policy-maker 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 policy-maker will choose a more transparent stress test than a naive one who ignores the endogenous response of inside 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 the 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. The proofs for this Section are in Online Appendix C.

We extend the model of Section 2. The timing and disclosure choices of Sender are as before. However, we now allow Receiver to take actions a, from a compact set $A \subset \mathbb{R}$. Players' payoffs depend on this action and Sender's type, θ. To avoid technicalities, we assume here that the type space is finite: $\theta \in \Theta = \{\theta_1, ..., \theta_N\} \subset \mathbb{R}$, where $\theta_N > \theta_{N-1} > ... > \theta_1$. Similarly, outside signals are drawn from a finite set, $S \subset \mathbb{R}$. We write $\mu_0(\theta) \equiv \Pr[\theta]$ for the prior distribution over 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.

Sender's payoff, $v(a)$, depends only on the action taken and is strictly increasing in a. If Sender chooses to disclose evidence, he incurs a cost c (independent of type). Receiver's payoff $u(a, \theta)$ is log-supermodular in a and θ, so that she optimally chooses higher actions when optimistic about θ.[24] We assume that, when Receiver knows θ with certainty, she has a unique best response denoted $a^\star(\theta) = \arg\max_{a \in A} u(a, \theta)$. We consider Perfect Bayesian Equilibria. Again, we require that off the equilibrium path, Receiver places zero probability on type θ' if she observes a signal such that $\pi(s|\theta') = 0$.

We focus on the simple message space $\{\theta, \emptyset\}$ with fixed disclosure costs for clarity of exposition. In Online Appendix F.1, 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.

[24] 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).

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

As before, we 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 $t > t_1$.*

The basic intuition behind Proposition 5 is that of reverse unraveling. The path we identify has the property that, at point $t^\star + \epsilon$ 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 non-equilibrium strategies yields full opacity as the unique equilibrium at all points beyond $t^\star$.

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 show in Online Appendix F.2 that the results continue to go through on an appropriate (relatively) open set of revealing paths. One caveat is that the revealing path must have limited support around $t = t_1$. If the signal support were full, then an unraveling equilibrium always exists, as discussed in Section 2. In this case, the result can still be applied to the study of the least transparent equilibrium.[25]

[25]However, equilibrium selection via Pareto-dominance or neologism proofness is more subtle in the general environment.

While inside disclosures $m \in \{\theta, \emptyset\}$ have a simple all-or-nothing form in our model, this does not drive our Proposition 5. In Online Appendix F.1 we allow for Sender to use partially verifiable messages in addition to full disclosure and cheap talk. This more general setting nests the classical message space in Milgrom and Roberts (1986), where Sender can disclose that θ belongs to any subset of the type space which contains the true θ. Similar to the all-or-nothing case, reverse unraveling implies that there is no type that sends a message $m = \theta$ after t_1 along the signal path identified in Proposition 5. Moreover, so long as the marginal costs of more informative disclosure are sufficiently small relative to the fixed costs of transmitting any verifiable information, there is a collapse to full opacity as the *unique* equilibrium at t_1.

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 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'.[26] 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.[27]

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 (Blackwell) than s, and*

[26]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.

[27]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 (Blackwell) than in $\mathcal{E}$.*

We establish the result in the Appendix in two steps. First, we argue (in the spirit of the Revelation Principle) that for any equilibrium $\mathcal{E}$ in which some type discloses with positive probability, we can construct intermediate outside signals which replicate the 'overall' information that Receivers would observe in $\mathcal{E}$. It is then easy to verify that full opacity is a strict equilibrium (that is, all types of Sender strictly prefer non-disclosure) of the game with these intermediate outside signals. Second, we show that it is possible to garble the intermediate signal in such a way that (i) all types still strictly prefer to play $m = \emptyset$, and (ii) the final signal structure is still a Blackwell improvement on the original outside signal. By construction, this final garbling leaves the Receiver with less information than under the equilibrium, $\mathcal{E}$.

4.3 Disclosures and the shape of payoffs

For this Subsection, we restrict attention to 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 and financial markets, 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 or financial security that gives her utility $X(\theta)$, or in settings with quadratic Receiver utility.

We assume that outside signals satisfy the strict Monotone Likelihood Ratio Property (MLRP; defined as in Milgrom, 1981). 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}.

To assess the relevance of the shape of the payoff function $X\left(\theta\right)$ 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.*

When payoffs are sufficiently concave, incentives to disclose for type θ_i come mainly from the downside increments $\Delta X_1, ..., \Delta X_{i-1}$, which he may lose if he stays quiet. 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.

Before moving on to an application of these insights in Section 5, we provide a simple example that illustrates equilibrium disclosures when parameters fall between the bounds we have established in Proposition 7.

Example Consider the 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 $\pi(s_i|\theta_i) = q = 1 - p - r$. To highlight the case where neither of the bounds in Proposition 7 applies, suppose that payoffs are linear with $X(\theta_i) = i$. It is easy to check that there can be monotone equilibria (for a range of disclosure costs) if and only if $p(1 - \frac{r}{p+r}) + q\frac{r}{p+r} \geq \frac{1}{2}$. 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.

5 Corporate disclosures when issuing debt and equity

We use our model to study corporate disclosures by a firm wishing to raise new funds, as a function of the type of liability it issues, either debt or equity. Since debt gives investors a concave claim and equity gives 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, or stay quiet ($m = \emptyset$). 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 equity, then the payoff to buyers of shares is the convex claim $\max\{\theta - d, 0\}$, where d denotes the face value of any existing legacy debt. The firm's payoff is the market price of equity, 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

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

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 debt, 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

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

It is easy to see that the net payoff is largest, and therefore 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 $\frac{1}{2k}\left[q(\emptyset, \theta + k) - q(\emptyset, \theta - k)\right] < 1$. 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 single 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 $\theta_L^\star \leq d \leq \theta_H^\star$.[28]

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 debt. 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 debt.

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. (2016; 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.

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

6 Conclusion

In this paper, we have studied the determinants strategic inside disclosures in the presence of additional outside information. Our main finding is that outside information strongly crowds out inside disclosures. The classic unraveling result turns into *reverse unraveling* when outside information is precise, and when Sender's payoffs from perceived quality are 'flat' towards the top of the type distribution. It is important to take this informational externality seriously when designing outside information. In a model of financial crises, we make a case for greater transparency in stress tests. An opaque stress test regime, which might well seem optimal in the absence of inside information, could give incentives for the strongest banks to prove themselves via inside disclosures and set off an unraveling loop, which eventually minimizes the insurance benefit to weaker, less liquid banks. In addition, we have derived implications for 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.

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.

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.

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

Daley, B., B. Green, and V. Vanasco (2016). Security design with ratings. *Mimeo*.

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

Einhorn, E. (2017). Competing information sources. *The Accounting Review, forthcoming*.

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.

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.

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.

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.

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

Leitner, Y. (2014). Should regulators reveal information about banks. *Federal Reserve Bank of Philadelphia Business Review, Third Quarter*.

Leitner, Y. and B. Williams (2017). Model secrecy and stress tests.

Leitner, Y. and B. Yilmaz (2016). Regulating a model.

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.

Lewis, G. (2011). Asymmetric information, adverse selection and online disclosure: The case of ebay motors. *American Economic Review 101*(4), 1535–1546.

Milgrom, P. (2008). What the seller won't tell you: Persuasion and disclosure in markets. *Journal of Economic Perspectives 22*(2), 115–132.

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.

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

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.

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

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

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.

Appendix

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.

Proposition 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 $E[\theta | m = \emptyset, s]$ is strictly increasing in s. Now we can define the desired $s^{\star}$ as the lowest $s \leq \hat{s}$ satisfying $E[\theta | m = \emptyset, s] \geq p$, 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 $s^{\star}$. 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 Proposition: 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 1

Suppose that $BR(\theta^{\star}) \in (p, \bar{\theta})$ for $\theta^{\star} \in (\bar{\theta} - \delta, \bar{\theta})$ for some δ. 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 $E[\theta | \theta \notin (p, \theta^{\star}), s^{\star}] = p$, or equivalently

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

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

$$\frac{dBR}{d\theta^\star} = \left.\frac{h(s^\star|\theta)}{c'(\theta) - H_\theta(s^\star|\theta)}\right|_{\theta=BR(\theta^\star)} \times \frac{ds^\star}{d\theta^\star} \tag{8}$$

$$\frac{ds^\star}{d\theta^\star} = \frac{(\theta^\star - p)h(s^\star|\theta^\star)f(\theta^\star)}{\int_{\theta\notin(p,\theta^\star)}(p-\theta)h_s(s^\star|\theta)dF + (p-\underline{\theta}(s^\star))h(s^\star|\underline{\theta}(s^\star))f(\underline{\theta}(s^\star))\frac{d\underline{\theta}(s^\star)}{ds}} \tag{9}$$

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 (8) 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 (9) 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 (9) therefore converges to zero. The integral in the denominator is

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

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

Theorem 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 $\theta_{min}^{t_0} < \bar{\theta}$ by contrapositive. Suppose that $\theta_{min}^{t_0} = \bar{\theta}$, 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 $\theta_{min}^{t_0+\epsilon} = \bar{\theta}$, 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 1, 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 $\theta_{min}^{t_0-\epsilon} < \bar{\theta}$ 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, letting subscripts on L denote partial derivatives, 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 1. By continuity of L, L_x and L_y, we can find ϵ and δ such that for all $|x| < \epsilon, |y| < \delta$,

$$L(x, y) = L(0, 0) + \int_0^x L_x(u, 0)du + \int_0^y L_y(x, u)du < 0$$

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 $\theta_{max}^t \leq \bar{\theta} - \frac{\epsilon}{2} \equiv \theta_1$, as required.

Proposition 2

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

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

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

We now show that $\theta_{min}^\star$ 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 Proposition 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

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

if and only if $\theta \in C$. Therefore, $s^\star\left(C\right) > s^\star\left(\theta^\star_{min}\right)$, and moreover (given that we have assumed full support of s), 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 $B\left(\theta'\right) > \theta'$. Thus, there exists a subset of $C \cap \left(p, \theta^\star_{min}\right)$, $\left[\theta', B\left(\theta'\right)\right]$ for whom

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

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

Proposition 3

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

$$\Phi\left(\frac{s^\star\left(\theta^\star\right) - \theta^\star}{k}\right) = c \tag{10}$$

$$E\left[\theta | s^\star\left(\theta^\star\right), \theta \notin [p, \theta^*]\right] = p \tag{11}$$

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

$$y\left(x\right) = \frac{\phi\left(-y\left(x\right)\right) - \phi\left(x - y\left(x\right)\right)}{1 - \Phi\left(x - y\left(x\right)\right) + \Phi\left(-y\left(x\right)\right)} \tag{12}$$

Rewriting (11), we have

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

Differentiating system (10) - (11), after some algebra we find that in any stable equilibrium:

$$\begin{aligned}\frac{d\theta^{\star}}{dk} \overset{sign}{=} & \frac{2k}{k^{2}+\sigma^{2}}s^{\star}(\theta^{\star})+\frac{1}{1-\alpha}\left[\left(y\left(\frac{\theta^{\star}}{\sigma_{s^{\star}}}\right)-\left(\frac{\theta^{\star}}{\sigma_{s^{\star}}}\right)y'\left(\frac{\theta^{\star}}{\sigma_{s^{\star}}}\right)\right)\frac{\sigma}{2\sqrt{\alpha}}-\mu\right]\frac{d\alpha}{dk}\\ & -\Phi^{-1}\left(c\right). \end{aligned} \tag{13}$$

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 $\Psi\equiv\Phi^{-1}\left(c\right)$. 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 (13), 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 (13) 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 policy-maker's payoff function, given disclosure interval (p, θ'), by

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

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

$$\frac{\partial W}{\partial k} = \frac{\partial W_n}{\partial k} + \frac{\partial W_n}{\partial \theta'} \frac{\partial \theta_k^\star}{\partial k}. \tag{14}$$

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

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

Since $h\left(\frac{s_k^\star - \theta}{k}\right) f(\theta) > 0$, so too is the integral above. Moreover, $\frac{\partial s_k^\star}{\partial \theta'} > 0$ 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 (14) 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.[29]

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

Additional Material: For Online Publication

C Proofs for Section 4

Notation

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, which is determined endogenously in equilibrium. We then define the net payoff from a verifiable disclosure as

$$\mathcal{N}(\theta) \equiv V(\theta) - E[v(\alpha(s))|\theta], \tag{15}$$

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

We write $\Pi(t)$ as shorthand for any path of conditional distributions $\pi(s|\theta;t)$ that define an outside signal. 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

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

$\hat{\Pi}(t)$ 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 $\mathcal{M}(\theta_i; t_k^\star) = \mathcal{M}(\theta_j; t_k^\star)$, $\forall i, j \leq k$. Construct $\Pi_{k+1}'(t)$ as follows: if $t_{k+1}' > t_k^\star$, replace row k of $\Pi_k'(t)$ with the functions $\left(\pi'\left(s \mid \theta_k; \frac{t'}{t_k^\star} t\right)\right)_{s=1}^{n}$. Otherwise, replace each row $i < k$ with $\left(\pi'\left(s \mid \theta_i; \frac{t_k^\star}{t'} t\right)\right)_{s=1}^{n}$. Applying this process to $\hat{\Pi}(t)$ clearly yields the required matrix $\Pi(t)$ after N iterations.[30]

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

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

bound $\mu_s \geq \underline{\kappa}_s^t \mathbf{1}_k + (1 - \underline{\kappa}_s^t)\, \mathbf{1}_s > \mathbf{1}_s$, where $\underline{\kappa}_s^t > 0$ satisfies

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

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

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

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

We write S for a set of outside signal realizations and Π as shorthand for the conditional distribution $\pi(s|\theta)$. 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 $\theta_i \in D$:

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

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 non-disclosure 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 $s \in S$.[31]

Proposition 7

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. Let $q_{ij} = E[Pr[\theta_j | s, m = \emptyset] | \theta_i]$ be the *expected* probability mass that Receiver places on type θ_j for given equilibrium beliefs when the true type is θ_i, and let $Q_{ij} = \sum_{j \leq i} q_{ij}$ be the associated 'cumulative distribution'. Integrating by parts, we can re-write the net payoff from disclosure as:

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

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,

[31] The perturbed signal structure here is particularly easy to follow, but violates full support. We show in Online Appendix F.3 that the result does not hinge on violation of full support. .

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 im-

poses 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. Optimal-

ity 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.

D Online Appendix: Additional results for Section 2

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. To clarify the exposition, we focus on the case where disclosure costs are fixed $c(\theta) = c$; where outside signals have full support, so that an unraveling equilibrium always exists; and where outside signals take the "truth plus noise" shape $s = \theta + k\epsilon$, where ϵ is a random variable with smooth distribution $G(\epsilon)$.

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.

Proposition 8. *Unraveling ($\theta^\star = \bar{\theta}$) 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 $\forall \theta^\star \geq \tilde{\theta}'$:*

$$-\frac{Pr\left(\theta \geq \theta^\star | s = \theta^\star + K\right) . E\left[\theta | \theta \geq \theta^\star,\, s = \theta^\star + K\right]}{Pr\left(\theta \leq p \mid s = \theta^\star + K\right) . E\left[\theta | \theta \leq p,\, s = \theta^\star + K\right]} > 1 \tag{16}$$

A proof is below. We establish Proposition 8 by considering a small deviation from unraveling. Instead of expecting every type $\theta \geq p$ to disclose, 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 Receiver to take the high action, even if Sender stays quiet. If signals are precise enough in the sense of condition (16), then 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 $\theta \sim \mathcal{N}(\mu, \sigma^2)$ and $\epsilon \sim \mathcal{N}(0, 1)$, condition (16) has a particularly natural interpretation. In this case, (16) 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, Receiver puts greater weight on s than on the prior μ. In such circumstances, observing a high signal more than offsets Receiver's concern that the 'quiet' signal gets worse as θ_1 increases. Since 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 8

Proof. We prove sufficiency by arguing the contrapositive: if unraveling is stable, then there must exist a $\tilde{K} \in \mathbb{R}$ that violates (16). Thus, suppose that transparency is a stable equilibrium. Then we can find a $\mathcal{B}$ such that $\forall \theta^\star \geq \mathcal{B}$

$$BR(\theta^\star) - \theta^\star > 0 \tag{17}$$

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

$$BR(\theta^\star) = s^*(\theta^\star) - kG^{-1}(c) \tag{18}$$

Substituting (18) into (17) yields a lower bound on $s^*(\theta^\star)$ as a function of $\theta^\star$ for any unraveling equilibrium:

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

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

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

or

$$\Pr\left(\theta \leq p \mid s^*(\theta^\star)\right) . E\left[\theta | \theta \leq p,\ s^*(\theta^\star)\right] + \Pr\left(\theta \geq \theta^\star | s^*(\theta^\star)\right) . E\left[\theta | \theta \geq \theta^\star,\ s^*(\theta^\star)\right] = p \tag{20}$$

Now consider the left hand side of Equation (16) evaluated at $\tilde{K} = kG^{-1}(c)$. By (19), $s^*(\theta^\star) > \theta^\star + \tilde{K}$. Then, it follows immediately from (20) and the MLRP assumption on signals s that

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

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

$$-\left(\frac{\Pr\left(\theta \geq \theta^\star | s = \theta^\star + \tilde{K}\right).E\left[\theta|\theta \geq \theta^\star,\, s = \theta^\star + \tilde{K}\right]}{\Pr\left(\theta \leq p \mid s = \theta^\star + \tilde{K}\right).E\left[\theta|\theta \leq p,\, s = \theta^\star + \tilde{K}\right]}\right) < 1 \tag{21}$$

This final inequality shows our contrapositive claim that if transparency is a stable equilibrium then for $K \leq \tilde{K} = kG^{-1}(c)$, (16) is violated.[32]

(Necessity) We prove the argument by contradiction, in each of two cases. Suppose then that (16) does not hold, but that the unraveling equilibrium is stable for all choices of $c > 0$. Then there exists some $\tilde{K}$, $\tilde{\theta}' \in \mathbb{R}$ such that[33]

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

$\forall \theta^\star \geq \tilde{\theta}'$. Note that (22) holds everywhere, not just in the limit, since the infimum is a non-decreasing function. Denoting for simplicity,

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

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 (22), $\exists \bar{\rho} > 0$ 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

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

[32] MLRP implies that if (21) holds for $\tilde{K}$, then it also holds for all $K \leq \tilde{K}$.

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

or

$$E\left[\theta|\theta \notin [p,\theta^\star],\, s=\theta^\star+\tilde{K}\right]<p \tag{23}$$

Therefore, since s satisfies the MLRP condition, (23) implies that $s^*(\theta^\star)>\theta^\star+\tilde{K}$ for all $\theta^\star \geq \tilde{\theta}'$ or

$$s^*(\theta^\star)-\theta^\star>\tilde{K} \tag{24}$$

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

$$s^*(\theta^\star)-BR(\theta^\star)=kG^{-1}(c)\geq\tilde{K} \tag{25}$$

Comparing (24) and (25) 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 (22), $\exists\bar{\rho}>0$ (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 (22), 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 $E[\theta\mid s,\theta\notin[p,\theta']]$ in s, θ', there must exist at least one $\theta^*_q\in\left[\theta'_q,\theta'_{q+1}\right]$ such that $P_{wer}\left(\theta^*_q,\tilde{K}\right)=1$, $\forall q$. In other words,

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

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

We can show (a proof is available on request) 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

In the context of our application to financial crises in Section 3, 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 policy-maker'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 9. *At the naive policy-maker'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*

$$\frac{\partial \theta^\star_k}{\partial k} \times \left[\frac{D(\theta^\star_k)}{B(\theta^\star_k)} - \frac{\delta(\theta^\star_k)}{\theta^\star_k}\right] \tag{26}$$

If better stress tests ($\downarrow k$) crowd out disclosures, then $\partial \theta^\star_k / \partial k > 0$. Proposition 9 shows that in this case, the additional welfare effect has the sign of the difference between the private relative cost of disclosure and the social relative cost. Intuitively, when 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 Proposition

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

We now 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. Relative to a naive regulator's optimal choice $k = k^\star$, 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 $c(\theta)$. Thus, 'marginal banks'

are worth approximately $c(\theta) > 0$ from an *ex ante* perspective. Insuring them always yields an *average* welfare improvement which is proportional to $c(\theta)$, 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 policy-makers 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

Throughout this Appendix, we employ the notational conventions introduced in Appendix C.

F.1 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 4 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.[34]

[34]This was true in all the main constructions we made to prove Propositions 5 and 6, 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

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 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 $\underline{c}_i(\theta_j) := \min_{m \in \underline{\mathbb{M}_j} \cap \mathbb{M}(\theta_i)} c_i(m)$. Intuitively, $\underline{c}_i(\theta_j)$ represents the minimized cost of θ_i of masquerading as type θ_j.

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 $\sigma \in \times_i \Delta\mathbb{M}(\theta_i)$, $\alpha(s, m) \in \arg\max_{a \in A} E^{\sigma}[u(a, \theta) \mid m, s]$.

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. 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 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

$$\underline{c}_i(\theta_h) - \underline{c}_i(\theta_l) < \mathcal{M}(\theta_i, \theta_l) - \mathcal{M}(\theta_i, \theta_h). \tag{27}$$

maximal punishment.

Equation (27) 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 4 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 4, 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. *Propositions 6 and 7 extend immediately under (27) and the above definition of monotonicity.*[35]

The statement of Proposition 5 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 5 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 5 continues to hold in the more general setting:

Lemma. *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 5 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 5 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

[35] Equation (27) 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.

any alternative in $\mathbb{M}(\theta_h)$. Otherwise, net payoffs would satisfy

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

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 $\overline{\theta}(m)=\max\{\theta: m\in supp\ \sigma(\theta),\theta\in\Theta\}$. For such a message and a signal $s\in S(\theta_l)\cap S\left(\overline{\theta}(m)\right)$, we must have

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

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

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

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 $\eta>0$:

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

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 5 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 $m \in \mathbb{M}^c$. □

F.2 Robustness of Proposition 5 to Perturbations

We write $\Pi(t)$ as shorthand for the conditional distribution $\pi(s|\theta;t)$ along a path of outside signals. We work in the case where the space of signals $S = \Theta = \{1, ..., N\}$, so that $\Pi(t)$ is an $N \times N$ matrix. We say that $\Pi(t)$ is *lower triangular* if $\pi(s|\theta_i;t) = 0$ for all $s < i$.

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.[36] Note that $\mathcal{O}$ is a non-empty set - indeed, the signal constructed in the proof of Proposition 5 is in $\mathcal{O}$.

Here we show that the conclusions of Proposition 5 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 5 can be generalized to the following:

Proposition 10. *Suppose that $c \leq V(\theta_2) - V(\theta_1)$. For any $\epsilon > 0$, there exists an open set $\mathcal{O}_\epsilon \subset \mathcal{O}$ with the following properties for all $\Pi(t) \in \mathcal{O}_\epsilon$:*

- *$\Pi(0)$ 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 5 which is in the interior of $\mathcal{O}$. Let $p_i : [0,1] \to [0,1]$ be a C^2, strictly increasing function with $p_i(0) = 0$,

[36] 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.

$p_i(1) = 1$ 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.

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

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

where $\rho_k(t) = \frac{p_k(t)}{p_k(t) + \sum_{i<k} y_i(t)}$, $y_i(t) = \omega^{1/2(k-i)(k-i+1)} p_i(t) \prod_{j>i}(1 - p_j(t))$. Notice that, by continuity of $p_k(t)$, $\forall t \in [0,1]$, $E[V(\theta_i) - V(\theta_s) \mid s < k]$ 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 5 shows that $E[V(\theta_i) - V(\theta_s) \mid s < k]$ is increasing in ω, for all $k \in \{1, \ldots, i\}$, $t \in [0,1]$. But

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

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 ϵ, $\exists \omega_\epsilon < 1$ such that $\sup |\mathcal{M}_t^\omega(\theta_i) - \mathcal{M}_t(\theta_i)| \leq \epsilon$ 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[37]

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

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 $\left|\Pi - \tilde{\Pi}_\omega\right| < \delta_\omega$.

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

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

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

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

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 5 can be applied to establish the claims in Proposition 10 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.3 Extending Proposition 6 to Outside Signals with Full Support

Here we explain when the proof of Proposition 6 can be extended to signal distributions that have full support. The key requirement for the proof of Proposition 6 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 $\mu'_s \in \Delta\Theta$

[37] 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 $k \to 0$.

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

$$\tau^s_s \mu'_s + \sum_{t \in \Theta} \tau^s_t \mathbf{1}_t = \hat{\mu}_s,$$

for some $(\tau^s_t)_{t \in \{s\} \cup \Theta}$ satisfying $\sum_{t \in \{s\} \cup \Theta} \tau^s_t = 1$, $\tau^s_t \geq 0$, $\forall t$. For each $s \in S$, let $\frac{\sum_{t \in \Theta} \tau^s_t \mathbf{1}_t}{1 - \tau^s_s} := \phi_s$ and notice that $\phi_s \in \Delta\Theta$.

Consider now an alternative collection of posterior beliefs

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

for $0 \leq \alpha, \beta \leq 1$. 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 following conditional weights on the above posterior beliefs:

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

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 $\left\{\mu'_s, \left\{\beta\left(\alpha \mathbf{1}_t + (1-\alpha)\phi_s\right) + (1-\beta)(1-\alpha)\mu'_s\right\}_{t \in \Theta}\right\}_{s \in S}$ 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 Blackwell-informative than (S, Π). 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 $\underline{\alpha} \leq \alpha \leq 1$, $\underline{\beta} \leq \beta \leq 1$.

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 $(S \cup \Theta, \Pi'')$. Then, $(S \cup \Theta, \Pi'')$ is strictly less informative than $(S \cup \Theta, \Pi')$. Appealing to the proof of Proposition 6, it is therefore also less informative than the equilibrium with information structure (S, Π).