Hauser, Daniel

Censorship and Reputation

Published in:
American Economic Journal: Microeconomics

DOI:
10.1257/mic.20200266

Published: 01/02/2023

Document Version
Publisher's PDF, also known as Version of record

Please cite the original version:
Hauser, D. (2023). Censorship and Reputation. *American Economic Journal: Microeconomics*, *15*(1), 497-528. https://doi.org/10.1257/mic.20200266

American Economic Journal: Microeconomics 2023, 15(1): 497–528
https://doi.org/10.1257/mic.20200266

Censorship and Reputation†

By Daniel N. Hauser*

I study how a firm manages its reputation by both investing in the quality of its product and censoring, hiding bad news from consumers. Without censorship, the threat of bad news provides strong incentives for investment. I highlight discontinuities in the firm's maximum equilibrium payoff that censorship creates. When censorship is inexpensive, the firm never invests and a patient firm's payoffs approach the lowest possible. In contrast, when censorship is moderately expensive, there exist equilibria where product quality is persistently high and payoffs approach the first-best, which can exceed the maximum equilibrium payoff if it was unable to censor. (*JEL* D21, D82, D83, G31, G32, L15)

The threat of bad news is a powerful motivator for a firm. From restaurants to auto mechanics, hotels to fossil fuel producers, no one wants their reputation tarnished by a negative news story or a bad review. A substantial literature in economics explores the incentives for unobserved investment in quality created by a firm's desire to maintain its reputation and avoid bad news. But what happens to this channel if the firm can censor bad news, preventing consumers from seeing negative information about its product? Can a firm build a reputation, even if consumers know that it can hide bad news from them? I study what happens to a firm's reputational incentives for investment when it can censor to manipulate how the market monitors it.

To study this problem, I incorporate censorship into the reputation model of Board and Meyer-ter-Vehn (2013). In this model a firm is selling a product with persistent, but stochastic, quality, which evolves according to the firm's level of investment. Consumers don't observe quality directly. Instead, they form beliefs based on news about current quality, which the firm can censor for a cost. I highlight two striking discontinuities in firm payoffs as the cost of censorship varies. Firstly, when censorship is cheap relative to the cost of investment, regardless of initial reputation, reputation effects collapse. Equilibrium quality and investment are low and, as the discount rate goes to zero, a firm's payoffs approach zero. Secondly, as soon as censorship becomes sufficiently expensive, payoffs jump to a positive level. In this case there exist equilibria where the firm's quality is always high in the long run and a

*Aalto University School of Business and Helsinki Graduate School of Economics (email: daniel.hauser@aalto.fi). Leslie Marx was coeditor for this article. Thanks to Aislinn Bohren, George Mailath, and Steven Matthews for their advice and support. I'm also grateful for suggestions from Simon Board, David Dillenberger, Jan Knoepfle, Pauli Murto, Mallesh Pai, Alessandro Pavan, Andrew Postlewaite, Julia Salmi, Philipp Strack, Juuso Välimäki, Rakesh Vohra, Yuichi Yamamoto, and seminar participants.

†Go to https://doi.org/10.1257/mic.20200266 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

patient firm's payoffs approach the first best. Finally, when the cost of censorship is infinite (i.e., censorship is impossible), a patient firm's payoff lies between the first best and zero and long-run quality can be both consistently high or consistently low with positive probability.

The substitutability of censorship and investment drives the dramatic collapse of incentives when censorship is inexpensive. If a firm couldn't censor, investing in quality is the only way to prevent bad news. Inexpensive censorship crowds out this investment. In the unique equilibrium the firm never has credible incentives to invest in quality. This makes it impossible for the firm to maintain a reputation. This sort of dynamic is prevalent across many industries. The Sugar Research Foundation paid three scientists $50,000 to downplay research that showed a connection between heart disease and sugar (O'Connor 2016). There is significant evidence that the tobacco industry knew as early as 1963 about the hazards of cigarettes but hid this research from the public (Hilts 1994b, a). Fossil fuel producers like Exxon knew about the harmful effects and actively hid this information from the public (Hall 2015), to this day resisting pressure to invest in alternative energy sources (Tabuchi 2020). In these situations suppressing information was much less expensive than developing a product without these defects, leading to low investment in improving quality.

In contrast, as soon as the cost of censorship exceeds the cost of investment, payoffs and reputation dynamics shift dramatically. Censorship gives a firm selling a low-quality product a way to avoid the large jumps in reputation—and the resulting decrease in incentives for future investment—caused by consumers adjusting their beliefs in response to bad news. While inexpensive censorship destroys the firm's investment incentives, sufficiently expensive censorship does not. A firm is motivated to produce a high-quality product to avoid paying the cost of censorship in the future. This introduces equilibria where, with probability one, in the long run, the firm consistently produces a high-quality product. For instance, a restaurant may benefit from some limited ability to hide bad reviews so that it isn't driven out of business by a few off days early on in its life cycle. But in the long run, it's more cost-effective to provide consumers with a good experience than to constantly suppress bad reviews. The threat of having to censor incentivizes investment. This leads to high quality in the long run and payoffs that approach the first best for a patient firm.

Both these cases contrast with the setting without censorship from Board and Meyer-ter-Vehn (2013). In that setting long-run quality is path dependent; with positive probability, the firm eventually establishes a persistent, high reputation. But, to generate credible incentives for investment, the firm must sometimes fall into a reputation trap and lose its reputation forever. These dynamics change if the firm can censor. On the one hand, censorship gives the firm an immediate way to prevent large reputational losses, which would otherwise push the firm's reputation down to a level where reputation incentives for future investment are much lower. By avoiding these reputation traps, the firm can potentially preserve credible investment incentives and build and maintain a high reputation. But if censorship is too cheap, the ability to prevent these drops in reputation destroys the firm's incentives to invest in the first place.

In the patient limit—the limit as the firm's discount rate goes to zero—the maximum average discounted payoff goes to (i) zero if censorship is inexpensive, (ii) the first best if censorship is moderately expensive, and (iii) lies between the two when censorship is very expensive. Long-run quality is low in the first case, high in the second, and both high and low with positive probability in the third. This result suggests that regulators, consumer protection agencies, and platform designers may benefit in the long run from making censorship more costly for firms, even if, at first glance, it seems that their intervention hasn't stopped firms from hiding bad news. But making censorship so onerous that it is effectively impossible may have unanticipated negative effects, especially on newly established firms. Early bad news can force the firm into a reputation trap and severely blunt any future investment in quality. There's evidence that seller reviews cluster almost entirely at the highest rating due to suppression of negative reviews (Tadelis 2016), but many online platforms seem to function reasonably well. An online platform like Yelp, Airbnb, or eBay should be very worried if it's incredibly easy to suppress negative reviews. But as long as this sort of suppression is sufficiently costly, sellers on the platform have incentives to invest in providing a high-quality service, even if the realized rating does not seem particularly informative.

Literature Review.—In this paper I extend the reputation for quality model developed in Board and Meyer-ter-Vehn (2013) by giving the firm the ability to hide bad news from the market. In the reputation for quality model, a firm is selling a product with persistent quality that changes at Poisson times based on the firm's level of investment. Board and Meyer-ter-Vehn (2013) show when consumers learn from bad news, the firm has strong incentives for investment at high reputations. This leads to path-dependent learning; with positive ex ante probability, a firm's reputation eventually converges either to the highest or lowest level and remains there permanently. The introduction of censorship changes these dynamics, either by depressing incentives for investment so much that the firm never invests in equilibrium and its reputation vanishes in the long run almost surely or by allowing the firm to avoid reputation traps, causing a patient firm to achieve approximately the first-best payoff if censorship is sufficiently expensive.

This reputation for quality model has been explored in other contexts in Halac and Prat (2016); Marinovic, Skrzypacz, and Varas (2018); and Hauser (2021). Marinovic, Skrzypacz, and Varas (2018) and Hauser (2021) study a closely related problem where the firm can reveal good news about its product to consumers. Hauser (2021) studies the interaction of a firm's ability to promote positive information about its product and exogenous bad news, showing that—like expensive censorship here—promotion can be used to eliminate reputation traps, but if promotion is too effective, it can reduce incentives for investment at high reputations, leading to reputation cycles. This paper is discussed in more detail in Section IIID.

There is a large literature on reputation, beginning in Kreps and Wilson (1982) and Milgrom and Roberts (1982) and surveyed in Mailath and Samuelson (2015). For the most part, in these papers the monitoring structure and firm type are fixed, and the distribution of signals can only be influenced directly through the firm's unobserved investment choices. The firm is willing to distort these choices to manipulate beliefs about its type. Reputation effects can discipline equilibrium behavior and beliefs, allowing players to achieve payoffs close to the Stackelberg bound in any equilibrium (Fudenberg and Levine 1989, 1992). In contrast, in the model considered in this paper, whether payoffs close to the Stackelberg bound can be achieved depends crucially on the cost of censorship.

Other papers have studied agents' incentives to reveal bad news to the public. There is a literature that focuses on how the threat of litigation incentivizes the disclosure of bad news (Marinovic and Varas 2016; Dye 2013; Daughety and Reinganum 2008). These papers focus on the incentives for disclosure alone, abstracting away from investment. While directly hiding bad news is not costly (the firm can always choose not to disclose), it is indirectly costly because the firm potentially faces sanctions in the future when the failure to disclose is uncovered. Firms may benefit from higher costs of hiding bad news (Marinovic and Varas 2016) since it leads to more positive beliefs after when the market does not observe any disclosure of bad news from the firm. In contrast, here I focus on a firm that is actively hiding bad news for a cost, and any information the firm hides (or the fact that it hid information in the first place) is never directly observed by the market. The benefits from making hiding bad news more expensive here, in contrast to this literature on litigation risk, primarily come from the more favorable market beliefs it generates about investment in quality in the future, as opposed to the more favorable beliefs about quality today.

Costly censorship has also been studied in political contexts. This literature mostly focuses on studying one-shot censorship decisions about whether or not the media discloses information about a politician's exogenous type. Besley and Prat (2006) consider a problem where a politician bargains with media outlets to hide bad news and finds that insufficient competition leads to media capture. Shadmehr and Bernhardt (2015) study a game where a ruler can hide media reports about its exogenous type and show that the ruler may benefit from committing to less censorship. Sun (2018) studies a dynamic problem where a ruler hides bad news over time about their fixed type and similarly finds that the ruler may benefit from committing to reduced censorship.

I. Model

The Firm.—Time is continuous, $t \in [0,\infty)$. There is a single long-lived firm with stochastic quality $\theta_t \in \{L,H\}$, $\Pr(\theta_0 = H) = x_0 \in (0,1)$. The firm has discount rate r. At each instant of time, the firm chooses a level of investment $a_t \in [0,1]$, for flow cost ca_t, $c > 0$. In addition a firm with low quality today chooses a level of censorship $\pi_t \in [0,1]$ for flow cost $k\pi_t$, $k > 0$.

Whenever $\theta_t = L$, bad news arrives via a Poisson process with arrival rate μ. If bad news arrives, it is revealed to consumers with probability $1 - \pi_t$. Quality evolves via Poisson shocks, as in Board and Meyer-ter-Vehn (2013). Specifically, there is a Poisson process with intensity $\lambda > 0$. Whenever there is an arrival of this process, θ_t becomes H with probability a_t, L with probability $1 - a_t$, and is fixed between arrivals. The firm observes θ_t; consumers do not. A firm's strategy is a stochastic process $(a_t,\pi_t)_{t=0}^{\infty}$ that determines the investment choice and level of censorship. These strategies are predictable processes with respect to the σ-algebra generated by the quality and news processes.[1] Throughout this paper I maintain the assumption $c < \lambda/(r+\lambda)$; otherwise, it would be optimal for the firm to not invest in quality even if θ_t and a_t were observable.

Consumers.—Consumers do not observe θ_t, a_t, or π_t directly. Consumers' only source of information is any uncensored bad news, which provides a signal about θ_t. Given the firm's strategy, the news process consumers observe effectively has intensity $\mu(1-\pi_t)1_{\theta_t=L}$. Let h_t denote the public history, which consists of the past public signal arrival times up to time t. Let $\mathbf{x}_t = \Pr(\theta_t = H|h_t)$, where the probability measure is the measure induced by the consumer beliefs about the firm's strategy.

Payoffs.—The firm receives a flow payoff of $\mathbf{x}_t$. The firm solves

$$\max_{\hat{a},\hat{\pi}} E_{\hat{a},\hat{\pi}}\left(\int_0^\infty e^{-rt}\left[\mathbf{x}_t - c\hat{a}_t - k\hat{\pi}_t 1_{\{\theta_t=L\}}\right]dt\right),$$

where the expectation is taken with respect to the actual probability measure induced by the firm's chosen investment and censorship levels, while $\mathbf{x}_t$ is determined by what consumers believe about the firm's investment and censorship choices.

Solution Concept.—I characterize Markov Perfect Equilibrium. These are Perfect Bayesian Equilibria where equilibrium strategies only depend on the market beliefs.

DEFINITION 1: *A Markov Perfect Equilibrium (MPE) consists of a pair* $a:[0,1] \to [0,1]$, $\pi:[0,1] \to [0,1]$, *which map market beliefs to the level of investment and censorship and believed strategies* $\tilde{a}:[0,1] \to [0,1]$ *and* $\tilde{\pi}:[0,1] \to [0,1]$*:*

(1) (a,π) *are sequentially rational for all* x_0.

(2) *Market beliefs*, $(\mathbf{x}_t)_{t\geq 0}$, *are consistent with Bayes' rule given strategies* $(\tilde{a},\tilde{\pi})$.

(3) *Beliefs are correct*, $a = \tilde{a}$, $\pi = \tilde{\pi}$

(4) *Given any history, if news arrives at time t, then* $\mathbf{x}_t = 0$ *and, for any* $s > t$, $\mathbf{x}_s$ *is consistent with Bayes' rule given believed strategies* $(\tilde{a},\tilde{\pi})$ *and prior* $\mathbf{x}_t = 0$.

The only off-path events in this model are arrivals of bad news when the market believes the firm is censoring at the maximum level. The fourth condition ensures

[1] As in Board and Meyer-ter-Vehn (2013), formally, there is a probability space (Ω, F, P), (i) a random variable θ_0 that determines initial quality, (ii) a sequence of independent uniform $[0,1]$ random variables that determine quality changes, (iii) a sequence of independent uniform random variables that determine whether news is censored, and (iv) the quality and news Poisson processes. To avoid some minor technical complications, it is convenient to allow both the low- and high-quality firms to choose π_t for flow cost to $k\pi_t 1_{\theta_t=L}$, so long as $\theta_t = H\pi_t$ effectively does nothing.

that if that occurs, beliefs update to zero and then continue to adjust using Bayes' rule with respect to the Markov beliefs. This restriction is natural, as news can only arrive if the firm is selling a low-quality product. In addition it ensures that the equilibrium respects the spirit of the Markovian restriction by restricting beliefs to only depend on the payoff-relevant state, both on and off path. Analogous restrictions have been made in similar models, for instance, in Marinovic, Skrzypacz, and Varas (2018).

From here on, I use MPE and equilibrium interchangeably. Fixing Markov beliefs about strategies $(\tilde{a}, \tilde{\pi})$, let $V(x, \theta)$ denote the firms' value function, that is

$$V(x,\theta) = \sup_{a,\pi} E_{(a,\pi)}\left(\int_0^\infty e^{-rt}\left[\mathbf{x}_t - ca_t - k\pi_t 1_{\{\theta_t = L\}}\right]dt \,\middle|\, \mathbf{x}_0 = x,\ \theta_0 = \theta\right),$$

where $\mathbf{x}_t$ is formed according to the believed strategies $(\tilde{a}, \tilde{\pi})$.[2]

II. Analysis

A. *Equilibrium Analysis*

Consumers learn about the firm's quality through arrivals of bad news. Given Markov beliefs $(\tilde{a}, \tilde{\pi})$, between arrivals of bad news beliefs are a deterministic process x_t that follows law of motion

$$\dot{x}_t = \underbrace{\lambda\left(\tilde{a}(x_t) - x_t\right)}_{\text{Quality breakthroughs}} + \underbrace{\mu\left(1 - \tilde{\pi}(x_t)\right)x_t(1 - x_t)}_{\text{Absence of news}}.$$

This expression accounts both for unobserved changes in quality and the inference consumers draw from no news. When news arrives at time t, consumers learn that $\theta_t = L$ and adjust their beliefs to zero.

Incentives to invest and censor are driven by the threat of bad news. The firm's incentives are determined by the value of high quality relative to low quality, $D(x) := V(x,H) - V(x,L)$, and the marginal cost of bad news, $\Delta(x) := V(x,L) - V(0,L)$.

LEMMA 1 (Sequential Rationality): *Fix Markov beliefs* $(\tilde{a}, \tilde{\pi})$. *Strategies* $(a_t, \pi_t)_{t=0}^{\infty}$ *are sequentially rational if and only if after any history, at almost all times,* a_t *solves*

$$\max_{a \in [0,1]} \lambda D(\mathbf{x}_t)a - ca$$

and π_t *solves*

$$\max_{\pi \in [0,1]} \mu\Delta(\mathbf{x}_t)\pi - k\pi.$$

[2] The law of motion for beliefs may be discontinuous. Therefore, there are some Markov strategies where the corresponding beliefs are not well defined (see Klein and Rady 2011). To resolve this, throughout this paper I restrict attention to believed strategies $(\tilde{a}, \tilde{\pi})$ and corresponding belief process $\mathbf{x}_t$ that satisfy the admissibility restrictions from Board and Meyer-ter-Vehn (2013) and are consistent with the discrete time approximation. See Appendix A1.

The firm invests in quality if the marginal value of quality exceeds the cost of investment and wants to censor news if the marginal loss from bad news exceeds the cost of censorship.

The firm has the strongest incentives to either invest or censor when beliefs are high because then it stands to lose the most from bad news. There is a tight relationship between $D(x)$ and $\Delta(x)$.

LEMMA 2 (Value of Quality): *Fix Markov beliefs* $(\tilde{a}, \tilde{\pi})$. *The value of quality* $D(x)$ *is*

$$D(x) = \int_0^{\infty} e^{-(r+\lambda)t}\big[\mu\big(1 - \pi(x_t)\big)\Delta(x_t) + k\pi(x_t)\big]dt.$$

Moreover, payoffs satisfy the following properties for any beliefs:

(1) $V(\cdot, H)$ *and* $V(\cdot, L)$ *are strictly increasing.*

(2) $V(x,H) \geq V(x,L)$ *for all x.*

(3) $\Delta(x)$ *is strictly increasing.*

(4) $D(x)$ *is weakly increasing and is strictly increasing in a neighborhood of any x where* $D(x) \neq k/(r+\lambda)$.

The loss from bad news $\Delta(x)$ and the value of quality $D(x)$ are both increasing in the firm's reputation. This is a consequence of the bad news monitoring structure. At higher reputations, bad news is more costly for the firm. Thus, the firm's incentives to censor are stronger, as are the firm's incentives for investment.

The ability to censor limits the difference between the value of high and low quality, $D(x)$. Since the firm can always censor bad news, the integrand, $\mu\big(1 - \pi(x_t)\big)\Delta(x_t) + k\pi(x_t)$, is at most k. The firm wants to have a high-quality product because then bad news no longer arrives, but censorship mitigates the threat of bad news.

Lemma 2 implies that the equilibrium can be expressed in cutoff strategies, simplifying the analysis.

PROPOSITION 1: *An MPE exists. If* $(1 + r/\lambda)c \neq k$, *then in any MPE there exist cutoffs* $x_a \in [0,1]$ *and* $x_\pi \in (0,1]$ *such that equilibrium strategies take the form*

$$a(x) = \begin{cases} 1 & \text{if } x > x_a \\ 0 & \text{if } x < x_a \end{cases},$$

$$\pi(x) = \begin{cases} 1 & \text{if } x > x_\pi \\ 0 & \text{if } x < x_\pi \end{cases}.$$

I call an equilibrium *full-shirk* if the firm does not invest with positive probability from any interior initial condition, i.e., $a(x) = 0$, and *full-work* if the firm invests at every reputation, i.e., $a(0) > 0$. In equilibrium the firm invests in quality and hides bad news to keep consumers from learning it is selling a bad product. Since the firm's payoffs are impacted the most by bad news when its reputation is high, it will censor and/or invest when it has established a high reputation.

B. *Patience and the Cost of Censorship*

The next two sections characterize the equilibrium set and establish the following result. Let

$$\mathcal{M}_r := \{x_0 V(x_0, H) + (1 - x_0) V(x_0, L) : \text{V is an MPE value function}\}$$

be the set of ex ante MPE payoffs for discount rate r, and let

$$V_r^* = \sup M_r$$

be the highest equilibrium payoff. To highlight the stark discontinuity in the value of V_r^* as k increases, I present the main result in terms of the limit of the highest average discounted payoff, rV_r^*, as $r \to 0$.[3]

THEOREM 1: *For any $x_0 \in (0,1)$, the highest average discounted payoff converges to*

(*i*) *the lowest possible payoff*, $\lim_{r \to 0} r V_r^* = 0$, *if* $c \geq k$,

(*ii*) *the first best*, $\lim_{r \to 0} r V_r^* = (1 - c)$, *if* $k > c$.

In the game where the firm cannot censor, the highest average discounted equilibrium payoff lies below $1 - c$ in the limit as $r \to 0$; payoffs lie strictly below $1 - c$ for a nonempty open set of parameters.

This result, and a similar result outside of the patient limit, will be established over the next three sections. The following definitions will be useful for establishing it.

DEFINITION 2: *An equilibrium* has a reputation trap *if with positive probability ex ante the firm's long-run reputation reaches and remains at zero*, *i.e.*, $\Pr(\underline{T} < \infty) > 0$, *where* $\underline{T} = \inf\{t : x_s = 0 \ \forall s > t\}$. *A firm is* in a reputation trap *if* $\Pr(\underline{T} < \infty) = 1$.

DEFINITION 3: *An equilibrium has* persistent perfect reputation *if with positive probability ex ante the firm's reputation goes to one in the long run*, *i.e.*, $\Pr(\lim_{t \to \infty} x_t = 1) > 0$. *This occurs almost surely if* $\Pr(\lim_{t \to \infty} x_t = 1) = 1$.

[3] The equilibrium characterization in the remainder of this paper establishes a similar, albeit harder to exposit, version of this same discontinuity outside of the limit.

In the long run, because of the bad news incentive structure, two possible outcomes can arise. A firm may acquire a reputation above x_a, produce a high-quality product, and achieve a persistent perfect reputation. In this case along almost all trajectories where $x_t \rightarrow 1$, in finite time product quality becomes high and stays high from then on, and the firm eventually starts, and never stops, investing in quality. Alternatively, if the firm's reputation reaches a low enough level, it has weak incentives to invest and falls into a reputation trap. Given the cutoff structure described in Lemma 2, if at time t a firm is in a reputation trap, it must be that $a(\mathbf{x}_s) = 0$ at all future times, and beliefs are monotonically decreasing along all trajectories, so quality eventually becomes low and the firm's reputation decays until reaching zero.

If the firm was unable to censor, as in the model of Board and Meyer-ter-Vehn (2013), the threat of reputation traps creates incentives. Firms invest at high reputations, both to avoid the immediate losses in reputation from bad news and to avoid falling into a reputation trap and losing incentives for future investment. In equilibrium the firm can achieve a persistent perfect reputation but also risks falling into a reputation trap. But with the introduction of censorship, the firm has another mechanism to avoid these losses. The firm could invest in producing a high-quality product to avoid bad news, but it could also hide that news directly. When censorship is sufficiently inexpensive $((1 + r/\lambda)c > k)$, it completely crowds out investment; the firm uses censorship to limit how much it could lose from having a low-quality product, but then the threat of bad news is never enough to support investment. In contrast, expensive censorship gives the firm a way to avoid reputation traps without destroying the incentives for investment they provide. The firm now wants to invest to avoid having to censor reputation-damaging bad news.

In the following three sections, I explore these dynamics by characterizing the set of equilibria. I then use that characterization to establish the three parts of Theorem 1. First, if censorship is inexpensive, the unique equilibrium is the full-shirk. Then, if the cost of censorship is high, a patient firm's payoffs approach the first best. Finally, in the game without censorship, payoffs may be bounded away from the first best.

III. Equilibrium Characterization

A. *Cheap Censorship*

In many settings censorship is relatively inexpensive. It may be inexpensive to incentivize consumers not to post bad reviews; it may be cheaper to cheat on emissions tests or bury research than to develop a cleaner engine. In this situation the ability to censor crowds out the firm's reputational incentives. When censorship is sufficiently inexpensive, there is a unique equilibrium, and in this equilibrium the firm never invests in quality.

PROPOSITION 2: *When censorship is sufficiently cheap*, $(1 + r/\lambda)c > k$, *there is a unique equilibrium. That equilibrium is full-shirk*, *i.e.*, *an equilibrium where the firm never invests in quality.*

The firm's ability to censor limits how much it can lose from low quality relative to high quality. Since the firm can always censor bad news, it never loses more than the cost of censorship, k, at any instant from low quality. Therefore, it would never invest when k is low enough relative to the cost of investment. If censoring bad news is inexpensive, then the firm substitutes censorship for investment.

This contrasts with both the game without censorship and the case of expensive censorship. In both these games the firm has strong incentives to invest at high reputations to avoid bad news. But inexpensive censorship destroys any incentives for investment, leading to full-shirk being the unique equilibrium.

This equilibrium is reminiscent of oil companies suppressing research about climate change or cigarette companies suppressing research about the health impacts of smoking. Instead of investing in safer and more effective products, these companies spent significant amounts of money to hide bad news. This was both bad for consumers and significantly damaged the firms' reputations in the long run.

Figure 1 illustrates the dynamics of this equilibrium. This full-shirk equilibrium has many undesirable properties from the firm's perspective. In the unique equilibrium reputation vanishes in finite time, along with the firm's profits and product quality.

LEMMA 3: *If* $c \geq k$, $\lim r V_r^* \to 0$.

Since the firm's reputation almost surely goes to zero in finite time in the unique equilibrium, the firm's average discounted payoffs must also go to zero as the firm becomes patient. In contrast, if the firm was unable to censor, this becomes the setting studied in Board and Meyer-ter-Vehn (2013). In that game there exist equilibria with persistent reputation building. This comparison is discussed in more detail in Section IIIC.

B. *Expensive Censorship*

Once censorship is sufficiently expensive, it becomes a powerful tool for the firm. Without censorship, if initial quality is low, $\theta_0 = L$, a firm is vulnerable to reputation traps. With positive probability ex ante, regardless of the firm's initial reputation, bad news arrives and pushes the firm's reputation to zero, where investment incentives are lowest. Investment and censorship work in tandem. Investment gives the firm a way to prevent bad news in the long run, while censorship lets the firm stop bad news in the short run, thereby avoiding the risk of ending up in a reputation trap.

When $k > (1 + r/\lambda)c$, there exist equilibria where the firm invests in quality.[4] As shown in the Proof of Proposition 1, there always exist equilibria of the form illustrated in Figure 2, where the firm stops investing at a lower reputation than it stops censoring.

[4] In the knife-edge case where $k = (1 + r/\lambda)c$, the equilibria may not be in cutoff strategies, as $\lambda D(x)$ can be kept constant and equal to c over an open interval of reputations. This nongeneric case unnecessarily complicates the analysis, so I omit it.

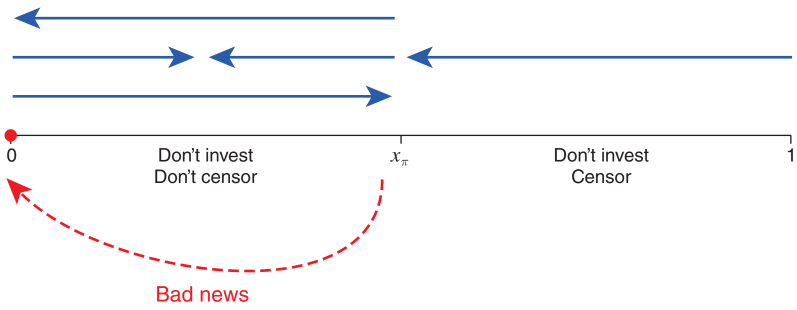

FIGURE 1. THE STRUCTURE OF A FULL-SHIRK EQUILIBRIUM

Notes: The solid arrows describe the drift of beliefs, while the dotted arrows illustrate how beliefs move after news; 0 is an absorbing state. In the leftmost region the direction of the drift is ambiguous and behaves according to one of the three arrows, depending on the relative size of λ and μ.

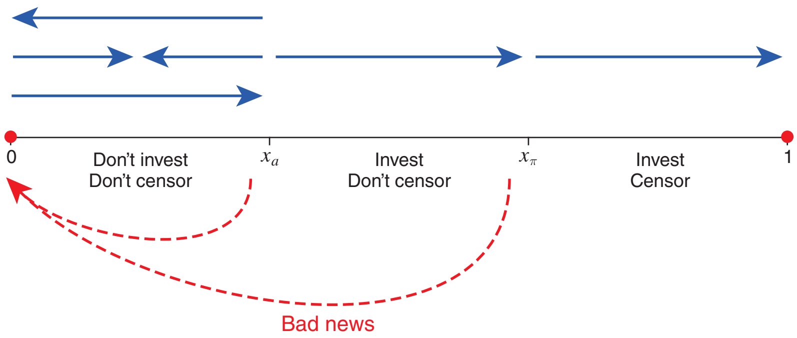

FIGURE 2. THE STRUCTURE OF AN EQUILIBRIUM WHERE THE FIRM STOPS INVESTING AT A LOWER REPUTATION THAN IT STOPS CENSORING

In this equilibrium the firm invests and censors when it has a high reputation, does not censor at intermediate reputations, and does not invest or censor at low reputations. Equilibria of this form highlight the three distinct long-run dynamics that may arise in equilibrium. When x_0 is below x_a, the firm is in a reputation trap. The firm's reputation decays over time, drifting toward zero until bad news arrives, which immediately lowers its reputation to zero. Between x_a and x_π, the firm's reputation is path dependent; there are reputation traps but also the possibility that the firm can establish a persistent perfect reputation; the firm's reputation converges to one with positive probability. Finally, when $x_0 > x_\pi$, the firm has a persistent perfect reputation almost surely.

A second type of equilibria, where $x_a \geq x_\pi$, may also exist. These are illustrated in Figure 3. In equilibria of this form, the long-run behavior of reputation is entirely determined by the initial reputation. If $x_0 < x_a$, then the firm is in a reputation trap—with probability one, its reputation will fall to zero and never recover. In contrast, if $x_0 > x_a$, then the firm has a persistent perfect reputation a.s.—with probability one, its reputation converges to one. In both of these types of equilibrium, firms with

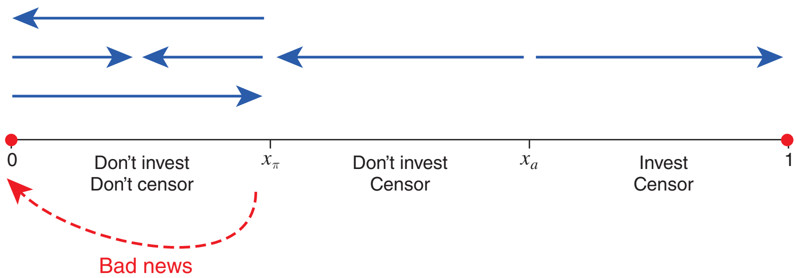

FIGURE 3. THE STRUCTURE OF AN EQUILIBRIUM WHERE THE FIRM STOPS INVESTING AT A HIGHER REPUTATION THAN IT STOPS CENSORING

sufficiently high initial reputations are able to secure the highest possible reputation in the long run with probability one.

Proposition 3 characterizes when equilibria where the firm a.s. attains a persistent perfect reputation exist.

PROPOSITION 3: *Suppose $k > (1 + r/\lambda)c$. There exists an equilibrium that is not full-work where the firm has persistent perfect reputation a.s. if and only if*

$$k \leq \frac{(\lambda + rx_0 - c(r+\lambda))\mu}{r(r+\lambda+\mu)}.$$

Moreover, if this condition holds, all equilibria either have persistent perfect reputation a.s. or at x_0 the firm is in a reputation trap, i.e., $\Pr(\underline{\tau} < \infty) = 1$, *where* $\underline{\tau} = inf\{t : x_s = 0 \ \forall s > t\}$.

This proposition establishes the existence of an equilibrium with persistent perfect reputation a.s.[5] For intermediate costs of censorship, the firm either has a persistent reputation or is in a reputation trap it cannot escape from. Consumers don't believe the firm is investing in quality, so the firm doesn't invest in quality, and its reputation vanishes. The equilibria with persistent reputation arise from the interaction between censorship and investment. The firm has the incentive to invest because, if product quality becomes low, it has to censor news to maintain its high reputation. Since censorship is expensive, the firm wants to avoid this, so it invests. This effect is strongest at high initial reputations, as there the firm stands to lose the most from losing its reputation, which in turn incentivizes censorship and investment. In equilibrium sufficiently expensive censorship provides the firm the credibility it needs to build up a persistent reputation. Moreover, to maintain this high reputation, the firm has incentives to invest in quality. In any equilibrium with persistent perfect

[5] The full-work equilibrium always has this property, but a full-work equilibrium only exists for very low costs of investment. I establish sufficient conditions for the nonexistence of such an equilibrium in Lemma A.5. This condition implies that full-work equilibria do not exist for any discount rate if $c > \mu/(2(\lambda+\mu))$.

reputation, $\Pr\left(a\left(\mathbf{x}_t\right) = 1\right) = 1$ for all t. Not only does the firm make high profits in this case; it also produces a high-quality product.[6]

These equilibria with persistent perfect reputation give payoffs that approach the first-best payoff. This result is what underlies the stark discontinuity in the maximum payoff. Even outside of the patient limit, a small increase in k so that $k > \left(1 + r/\lambda\right)c$ potentially leads to a large jump in the highest possible payoff the firm could achieve in equilibrium. Above the cutoff, in any equilibrium it is either the case that x_0 lies below x_π and above x_a, in which case the reputation dynamics are path dependent, or x_0 lies below both x_a and x_π, in which case the firm's reputation a.s. goes to zero in the long run.

Corollary 1 concludes the characterization of the set of equilibria when censorship is expensive.

COROLLARY 1: *Suppose* $k > \left(1 + r/\lambda\right)c$.

- *If* $k > \left(\lambda + rx_0 - c\left(r + \lambda\right)\right)\mu/\left(r\left(r + \lambda + \mu\right)\right)$, *then every equilibrium is either full-work* $\left(a\left(0\right) > 0\right)$ *or has a reputation trap.*
- *If* $k \geq \left(1 - c\right)\left(r + \lambda\right)\mu/\left(r\left(r + \lambda + \mu\right)\right)$, *then the firm never censors in equilibrium.*

Finally, there exists an equilibrium where $x_a < 1$ *if and only if* $c \leq \mu\lambda/\left(\left(r+\mu\right)\left(r+\lambda\right)\right)$.[7]

Figure 4 illustrates the equilibrium characterization. Unsurprisingly, if censorship is very expensive, then it becomes hard to incentivize censorship in equilibria. Then reputation dynamics begin to resemble those in the setting where the firm cannot censor. In equilibrium either $x_\pi > x_0 > x_a$, in which case the firm's reputation converges to both zero and one with positive probability, or both $x_a > x_0$ and $x_\pi > x_0$, in which case the firm's reputation decays to zero in finite time a.s. But, the more patient the firm is, the more valuable preventing bad news is for the firm and the more willing the firm is to pay to censor bad news. So at any x_0, for a small enough r, there exists an equilibrium with persistent perfect reputation a.s.

COROLLARY 2: *The maximum average discounted equilibrium payoff* rV_r^* *converges to the first-best payoff,* $\left(1 - c\right)$, *as* $r \to 0$.

A sufficiently patient firm is always able to secure approximately the first-best payoff in some equilibrium. The more patient the firm is, the more it benefits from a persistent perfect reputation. This both makes investment credible and gives the firm incentives to censor early bad news until that investment is successful. So the firm avoids reputation traps, initially through censoring bad news, then through successfully maintaining high quality.

[6]Formally, $\Pr\left(\tau_\theta < \infty\right) = 1$, where $\tau_\theta = inf\left\{t : \theta_s = H \ \forall s > t\right\}$.

[7]The upper bound on c, $\mu\lambda/\left(\left(r+\mu\right)\left(r+\lambda\right)\right)$ is an upper bound on $\lambda D\left(x\right)$ for any possible beliefs, so as long as it is possible to incentivize investment, it can be incentivized if censorship is sufficiently expensive.

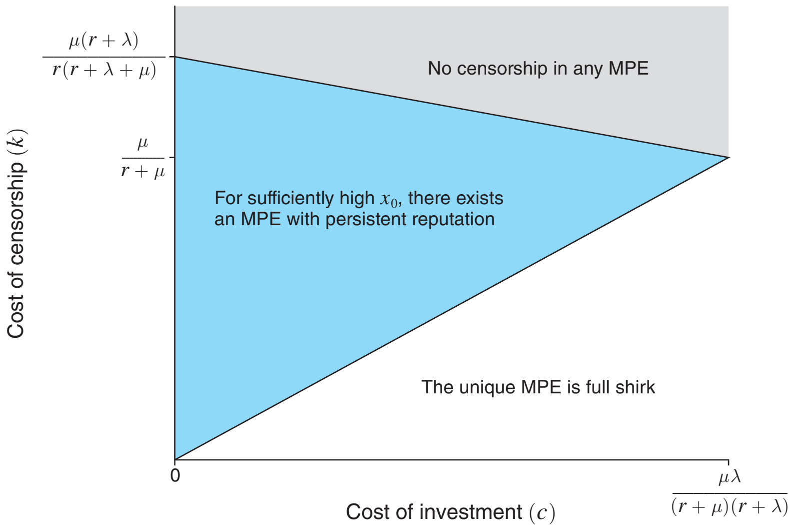

Figure 4. Markov Perfect Equilibria

Note: There do not exist equilibria with investment for $c > \mu\lambda/\big((r+\mu)(r+\lambda)\big)$ for any k.

C. *No Censorship*

To close the analysis, I show that without censorship, payoffs are bounded away from the first-best for an open interval of parameters.[8] In any equilibrium in this game, the firm still follows a cutoff strategy, i.e., there exists an x_a such that

$$a(x) = \begin{cases} 1 & \text{if } x > x_a \\ 0 & \text{if } x < x_a \end{cases}.$$

This leads to three qualitatively different types of equilibria: (i) shirk-work equilibria where $x_a \in [0,1)$ and $a(0) = 0$, (ii) full-shirk equilibria where $x_a = 1$, and (iii) full-work equilibria where $x_a = 0$ and $a(0) > 0$.

Of these three equilibria, only the full-work equilibria delivers persistent perfect reputation a.s. In both the full-shirk and shirk-work equilibria, the lowest reputation, $x = 0$, is an absorbing state. Since the firm cannot censor bad news, bad news arrives with positive probability ex ante in any of the equilibria. Thus, with positive probability, the firm falls into a reputation trap and never recovers. This possibility bounds equilibrium payoffs away from the first best in any equilibrium with a reputation trap. Proposition 4 characterizes one such bound and provides sufficient conditions to rule out the full-work equilibrium.

[8] While this bound on payoffs is new, it builds on the more detailed analysis of this model without censorship in Board and Meyer-ter-Vehn (2013).

PROPOSITION 4: *Suppose the firm cannot censor.*

(i) *In any shirk-work equilibrium, the average discounted payoff,* $r\big(x_0 V(x_0,H) + (1-x_0)V(x_0,L)\big)$, *is bounded above by* $(1-c)\big(x_0 + (1-x_0)(r+\lambda)/(r+\lambda+\mu)\big)$.

(ii) *In any full-shirk equilibrium, the average discounted payoff is bounded above by* $r(1/\lambda + 1/\mu)$.

(iii) *A full-work equilibrium doesn't exist for any* $r > 0$ *if* $c > \mu/\big(2(\lambda+\mu)\big)$.

This proposition establishes that, as $r \to 0$, payoffs in any equilibrium other than the full-work equilibrium converge to zero. Even in a shirk-work equilibrium, where persistent reputations are possible, the firm cannot achieve the first best, as bad news arrives with positive probability ex ante.[9] So, as long as a full-work equilibrium doesn't exist (e.g., when $c > \mu/\big(2(\lambda+\mu)\big)$) even as $r \to 0$, payoffs are bounded away from $1-c$. In that case

$$\limsup rV_r^* \leq (1-c)\left(x_0 + (1-x_0)\frac{\lambda}{\lambda+\mu}\right) < 1-c.$$

The term $(1-x_0)(\lambda)/(\lambda+\mu)$ captures the loss due to initially starting with a low reputation and falling into a reputation trap. This dynamic is important for incentivizing investment but also prevents the firm from achieving the first-best payoff.

Finally, combining these results gives Theorem 1, which I restate here.

THEOREM 1: *For any* $x_0 \in (0,1)$, *the highest average discounted payoff converges to*

(i) *the lowest possible payoff,* $\lim_{r\to 0} rV_r^* = 0$, *if* $c \geq k$,

(ii) *the first best,* $\lim_{r\to 0} rV_r^* = (1-c)$, *if* $k > c$.

In the game where the firm cannot censor, the highest average discounted equilibrium payoff lies below $1-c$ *in the limit as* $r \to 0$*; strictly below for a nonempty open set of parameters.*

In fact, while these differences are starkest in the patient limit, Propositions 3 and 2 imply a discontinuous change in the highest equilibrium payoff between the cheap and expensive censorship cases even outside of the patient limit. Figure 5 illustrates this discontinuity.

Outside of the limit, the comparison between payoffs in this environment and the environment without censorship (or where censorship is so prohibitively expensive that it's never used) is ambiguous. An impatient firm that faces a high cost

[9] Unsurprisingly, as in Lemma 3, full-shirk payoffs go to zero as $r \to 0$. A tighter bound than the bound described here is constructed in Lemma A.7.

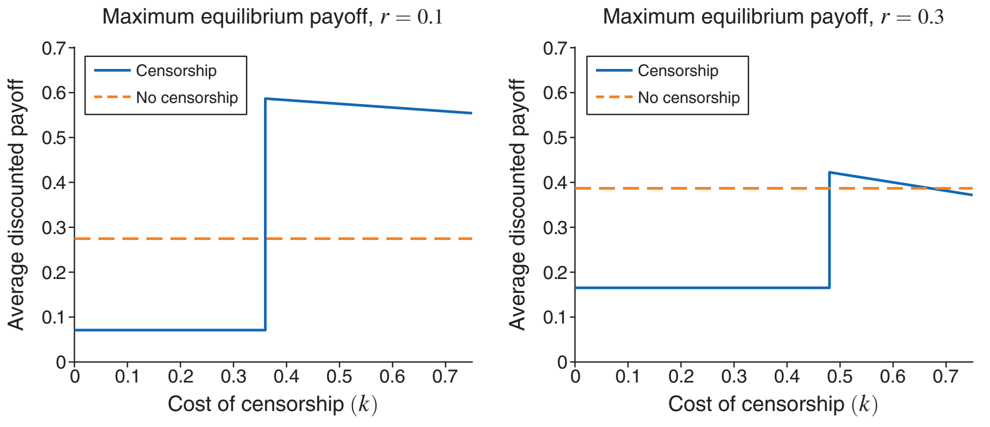

FIGURE 5. HIGHEST EQUILIBRIUM PAYOFF

Notes: $\mu = \lambda = 1$, $c = 0.3$, $x_0 = 0.2$, $r = 0.1$ in left figure, $r = 0.3$ in right figure. The dotted line is the highest payoff in the game from Board and Meyer-ter-Vehn (2013).

of censorship may prefer to commit to never censoring and benefit from the more favorable drift of beliefs.

D. *Bad News and Promotion*

In a related work Hauser (2021) studies how a firm's reputational concerns interact with its ability to disclose information about its product. Quality behaves as in the model studied here, and at random intervals a promotional opportunity arrives, when the firm can pay cost k and disclose its quality. Among the settings they study is an environment where consumers learn both through the firm's promotion and bad news the firm cannot influence.

While their setting has some similarities to the setting studied here, the results are technically and conceptually distinct. In the environment with censorship, incentives for investment are generated by the desire to avoid bad news and reputation traps, either directly or indirectly through the threat of having to censor any news that arrives, in some sense replacing the original incentives delivered by bad news. Promotion, in contrast, can't stop bad news from arriving. Instead, inexpensive promotion creates incentives for investment at low reputations that bad news could not deliver, as the firm can promote after successful investment. This eliminates reputation traps but potentially reduces incentives for investment at high reputations by reducing the threat of bad news, meaning quality and reputation cycle in the long run.

From a technical perspective, this paper and Hauser (2021) use different techniques to analyze the respective models. Allowing the firm to control the monitoring problem changes the firm's problem in two ways relative to Board and Meyer-ter-Vehn (2013). First, the news process is now endogenously determined by the firm, and second, the value functions vary discontinuously as equilibrium beliefs change; a small change in the relative positions of the believed cutoffs can lead to a large change in reputation dynamics.

In this paper I take advantage of the structure of the belief process to construct an equilibrium and show that other equilibria are similar above the investment cutoff x_a. To do this, I solve a pair of auxiliary problems where the value functions are well behaved and whose solutions together pin down the promotion and investment cutoffs consistent with the MPE where $x_\pi \geq x_a$.

In contrast, in the game with promotion studied in Hauser (2021), the belief dynamics are less well behaved. There are many different possible belief dynamics that may arise in equilibrium, including not only reputation traps and persistent reputation but also the possibility of long-run reputation cycles. This makes more explicit equilibrium characterizations difficult, as these different belief dynamics have very different implications for payoffs. While an explicit construction is difficult, promotion provides a powerful potential deviation; the firm can always invest in quality and then promote when quality is high to restore a high reputation. This observation allows for the construction of bounds on payoffs that hold across all MPEs, which they use to rule out both reputation traps and persistent perfect reputations. There isn't a similar deviation that I can exploit in the environment with censorship; the firm has no way to credibly signal to consumers that it has produced a high-quality product beyond the inference consumers draw from the absence of bad news. In particular, if the market believes that the firm isn't investing at zero, then there's no way the firm can meaningfully alter the belief dynamics through any deviation.

IV. Conclusion

Censorship dramatically changes reputation dynamics. It allows firms selling low-quality products to avoid bad news and preserve their reputations. Inexpensive censorship eliminates investment incentives and drives payoffs to zero. In contrast, expensive censorship allows the firm to build a persistent reputation, and a patient firm can achieve the first-best payoff. This manifests as a discontinuous change in the highest possible payoff. So even a small, unanticipated shock that makes censorship cheaper can have a large negative impact on quality.

In addition expensive censorship can lead to higher levels of investment than eliminating censorship altogether. These results suggest that policies that make censorship more difficult, even if they do not seem to discourage censorship much in the short run, can have large effects in the long run. A policymaker may want to make censorship more expensive but not so incredibly difficult that a firm cannot censor bad news at all, to encourage investment. That's not to say that censorship is always beneficial if it is sufficiently expensive. An impatient firm may benefit from not being able to censor, or a policymaker who is worried that the product's price does not internalize the negative costs of selling a low-quality product (for instance, due to negative externalities) may want to make censorship impossible to make it easier to detect low-quality products, even if this discourages future investment. That said, expensive censorship can be a powerful tool to drive investment in quality; it maintains the strong incentives for investment created by reputation traps but also provides the firm with a tool to avoid falling into a reputation trap.

Appendix

A1. *Admissibility*

As noted in Klein and Rady (2011) and Board and Meyer-ter-Vehn (2013), there are Markov strategies (a, π) for which the differential equation

$$\dot{x}_t = \lambda\big(a(x_t) - x_t\big) + \mu\big(1 - \pi(x_t)\big)x_t(1 - x_t)$$

has no solution or admits multiple solutions. In order to resolve this, I impose additional admissibility restrictions on believed strategies. These restrictions are identical to the restrictions placed in Board and Meyer-ter-Vehn (2013) and Hauser (2021). If beliefs follow the law of motion $\dot{x}_t = g(x_t)$, then the drift $g(x)$ must satisfy one of the following conditions at any point of discontinuity:

(1) $g(x) = 0$

(2) $g(x) > 0$ and $\tilde{\pi}(x), \tilde{a}(x)$ are right continuous at x,

(3) $g(x) < 0$ and $\tilde{\pi}(x), \tilde{a}(x)$ are left continuous at x,

and beliefs can be partitioned into a finite set of intervals such that both the believed investment and censorship choices are Lipschitz continuous on the interior of all these intervals and satisfy the above conditions at the boundaries. Under these conditions, the Picard Lindelof theorem implies that the belief process admits a solution (see Board and Meyer-ter-Vehn (2013, section A.1)), and whenever it admits multiple solutions, I select the one consistent with the discrete time approximation. Note that these restrictions are placed on believed strategies, not on the firm's actual strategy. No restrictions are placed on possible deviations the firm could make.

A2. *Preliminary Proofs*

LEMMA A.1: *Fix any admissible Markov beliefs $(\tilde{a}, \tilde{\pi})$ and pair of initial conditions $x_0 > x_0'$ and let $(x_t)_{t=0}^{\infty}$ and $(x_t')_{t=0}^{\infty}$ be the corresponding stochastic processes, then $x_t \geq x_t'$ for all $t < \tau$, where τ is the first time news arrives.*

PROOF:

Consider the belief trajectories in the absence of news, x_t and x_t'. These are continuous in t. If $x_t = x_t'$ at any t, then for any $s > t$, x_s and x_s' both solve

$$\dot{x}_s = \lambda\big(\tilde{a}(x_s) - x_s\big) + \mu\big(1 - \tilde{\pi}(x_s)\big)x_s(1 - x_s)$$

with the same initial condition. Since the solution to this is the unique solution that is consistent with the discrete time approximation, x_s' and x_s must be equal. So if it is ever the case that $x_t = x_t'$, the two processes must be the same from then on, so $x_t \geq x_t'$. ∎

PROOF OF LEMMA 1:

Fixing admissible Markov beliefs $(\tilde{a}, \tilde{\pi})$, truncating at the first arrival on any process, the firm's payoff from this strategy can be rewritten as

$$V(x,\theta) = sup \int_0^\infty e^{-(r+\lambda+\mu 1_{\{\theta=L\}})t} \Big[x_t + 1_{\{\theta=L\}}\big(\mu\pi_t\Delta(x_t) - k\pi_t + \mu V(0,L)\big)$$
$$+ \lambda a_t D(x_t) + \lambda V(x_t, L) - c\,a_t\Big] ds,$$

where x_t solves $\dot{x}_t = \lambda\big(\tilde{a}(x_t) - x_t\big) + \mu\big(1 - \tilde{\pi}(x_t)\big)x_t(1-x_t)$ with $x_0 = x$.

This strategy is optimal if $a(x)$ solves

$$\max_{a\in[0,1]} \lambda D(x)a - ca$$

and the optimal $\pi(x)$ solves

$$\max_{\pi\in[0,1]} \mu\Delta(x)\pi - k\pi$$

since these maximize the integrand pointwise. Similarly, any strategy that does not satisfy that equation a.e. gives a value strictly lower than a strategy that does. ∎

PROOF OF LEMMA 2:

Given admissible Markov beliefs $(\tilde{a}, \tilde{\pi})$, fix a sequentially rational pair of strategies (a, π). Lemma 5 of Board and Meyer-ter-Vehn (2013) shows that

$$V(x_t,\theta) - V(x_{t'},\theta) = \int_t^{t'} \Big[x_s - 1_{\{\omega=L\}}\big(\mu\big(1-\pi(x_s)\big)\Delta(x_s) + k\pi(x_s)\big)$$
$$+ \lambda\big(a(x_s)D(x_s) + V(x_s,L) - V(x_s,\theta)\big) - ca(x_s)\Big] ds.$$

Subtracting the two value functions gives

$$D(x_t) - D(x_{t'}) = \int_t^{t'} \Big[\mu\big(1-\pi(x_s)\big)\Delta(x_s) + k\pi(x_s) - \lambda D(x_s)\Big] ds.$$

Applying lemma 5 of Board and Meyer-ter-Vehn (2013) again,

$$D(x_t) = \int_t^\infty e^{-(r+\lambda)(s-t)} \Big[\mu\big(1-\pi(x_s)\big)\Delta(x_s) + k\pi(x_s)\Big] ds.$$

Now we can show the monotonicity properties:

$V(x,\theta)$ *Strictly Increasing in* x.—Fixing θ_0, for any two initial conditions, $x > x'$, the firm facing consumers with prior $x_0 = x$ could instead follow the strategy it would have followed if it faced consumers with prior $x_0 = x'$. This would induce the same probability measure over signals and quality. Moreover, until bad news arrives, the firm that faced consumers with a higher prior has a higher reputation at any point in time by Lemma A.1, and has a strictly higher reptuation for an initial interval of time, and thus receives a higher flow payoff. Since the firm's payoffs must be greater than the payoff from mimicking the x'-strategy, $V(x,\theta)$ is increasing in x.

$V(x,\theta)$ Is Increasing in θ.—The high-quality firm gets a weakly greater payoff than the low-quality firm since it can generate the exact same distribution over future quality changes as a firm that starts with low quality. Moreover, since bad news never arrives when quality is high, the firm's flow payoffs must be higher at every time t when compared with a low-quality firm that started with the same initial reputation. So $V(x,H) \geq V(x,L)$.

$\Delta(x)$ Is Strictly Increasing.—The cost of news $\Delta(x) = V(x,L) - V(0,L)$ is strictly increasing since $V(x,L)$ is strictly increasing.

$D(x)$ Is Increasing.—Finally, since

$$D(x) = \int_0^{\infty} e^{-(r+\lambda)t}\left[\mu\left(1 - \pi(x_t)\right)\Delta(x_t) + k\pi(x_t)\right]dt,$$

where $\pi(x_t)$ maximizes

$$\pi(x_t)\left[\mu\Delta(x_t) - k\right].$$

Recall that $\mu\Delta(x_t) - k$ is strictly increasing. Thus, by Lemma 1, if beliefs start at a larger x_0, then the integrand is larger pointwise. So $D(x)$ is increasing, strictly so unless the $\pi_t = \mu$ at all future instances of time, in which case $D(x) = k/(r+\lambda)$.

Continuity of Value Functions.—As long as the belief differential equation has a solution, the value functions satisfy some standard continuity properties. Lemma A.2 establishes that value functions are only discontinuous at the points where the believed strategies are not only discontinuous but cause the drift to change signs.

LEMMA A.2: *Under any admissible Markov beliefs, functions $t \mapsto V(x_t,\theta)$ and $t \mapsto x_t$ are continuous. Moreover, $V(x,\theta)$ is continuous at any x where*

(*i*) *There exists an $\epsilon > 0$ and a $\delta > 0$ such that if $|x - x'| < \delta$, then the drift of beliefs at x' is greater than ϵ.*

(*ii*) *There exists an $\epsilon < 0$ and a $\delta > 0$ such that if $|x - x'| < \delta$, then the drift of beliefs at x' is less than ϵ.*

(*iii*) *There exists a $\delta > 0$ such that if $|x - x'| < \delta$, the drift of beliefs at x' is* 0.

PROOF:

Appendix A.1 of Board and Meyer-ter-Vehn (2013) establishes that given admissible beliefs, a solution exists to the belief ODE $\dot{x}_t = g(x)$ and $t \mapsto x_t$ is continuous.

Fix a pair of sequentially rational strategies (a,π). Truncating at the first arrival, for any times $t < t'$,

$$\begin{aligned}
&\left|V(x_t,\theta) - V(x_{t'},\theta)\right| \\
&\qquad = \left|\int_t^{t'} e^{-(r+\mu 1_{\{\theta=L\}}+\lambda)(s-t)}\Big[x_s + \lambda\big(a_s V(x_s,H) + (1-a_s)V(x_s,L)\big)\right. \\
&\qquad\qquad + 1_{\{\theta=L\}}\big(\pi_s\big(\mu V(x_s,L) - k\big) + (1-\pi_s)\mu V(0,L)\big) - c\,a_s\Big]ds \\
&\qquad\qquad \left. + \left(e^{-(r+\mu 1_{\{\theta=L\}}+\lambda)(t'-t)} - 1\right)V(x_{t'},\theta)\right| \\
&\qquad \le \left(1 + \frac{2\lambda + r + \mu}{r}\right)|t - t'|,
\end{aligned}$$

where the inequality follows from x_t being bounded above by 1, $V(x_t)$ being bounded above by $1/r$, and $1 - e^{-(r+\mu 1_{\{\theta=L\}}+\lambda)(t'-t)}$ being bounded below by $(r+\mu+\lambda)|t-t'|$.

Suppose there is some δ where $g(y) > \epsilon > 0$ in a neighborhood for all $y \in (x - \delta, x + \delta)$. Let $f(t)$ be function implicitly defined by $t \mapsto x_t$ with $x_0 = x - \delta/2$. This is a strictly increasing function when restricted to the interval defined by $\left[0, f^{-1}(x + \delta/2)\right]$. So it has a continuous inverse on this interval, and therefore $V(x,\theta)$ is continuous at x since it is the composition of a continuous function and $t \mapsto V(x_t,\theta)$, which is continuous.

If $g(\cdot)$ is constantly zero on some interval $(\underline{x},\bar{x})$ containing x, then the firm's problem in this interval simplifies to

$$V(y,\theta) = \max_{a\in[0,1],\pi\in[0,1]} E\left(\int_0^\infty e^{-rt}\left(y - k\pi 1_{\{\theta_t=L\}} - ca\right)\middle|\theta_0 = \theta\right),$$

for any $y \in (\underline{x},\bar{x})$. The objective function is continuous in y, so by the maximum theorem, this is continuous at x. ∎

A3. *Equilibrium Construction*

PROOF OF PROPOSITION 1:

The cutoff structure of equilibrium is a direct implication of Lemmas 1 and 2. $D(x)$ and $\Delta(x)$ are increasing. The optimal $\pi(x)$ and $a(x)$ solve

$$\max_{\pi\in[0,1]} \mu\Delta(x)\pi - k\pi$$

and

$$\max_{a\in[0,1]} \lambda D(x)a - ca,$$

respectively. Moreover, $\Delta(x)$ is strictly increasing, and at any x, $D(x)$ is either strictly increasing or

$$D(x) = \int_0^\infty e^{-(r+\lambda)} k dt = \frac{k}{r+\lambda} \neq c/\lambda,$$

so the solutions to these maximization problems define cutoff strategies. Suppose there was an equilibrium that wasn't in cutoff strategies, (a, π). Then at any x_0, this must agree with the solution to those maximization problems a.e. Let x_0 be a point where it doesn't. Let $g(x_t) = \dot{x}_t$. If $g(x_0) = 0$, then beliefs are constant so $\left(a(x_t), \pi(x_t)\right)$ doesn't agree with the maximization problem a.e. If $g(x_0) > (<)$ 0, then (a, π) are right (left) continuous by admissibility, so they can't solve the maximization problems a.e.

Finally, note that if $D(x) \leq k/(r+\lambda)$ so if $(1 + r/\lambda)c > k$, then investment is not sequentially rational.

It remains to establish existence. Depending on where the costs lie relative to each other, the equilibria may have very different structures. To establish existence, I break the problem up into these different cases and construct equilibria in each specific case.

LEMMA A.3: *If $(1 + r/\lambda)c > k$, there exists an MPE where the firm never invests.*

PROOF:

Here, I construct an equilibrium without investment and argue it must be unique. At the censorship cutoff, one of two things can happen; either (i) the drift of beliefs is negative at reputations above the censorship cutoff and positive at reputations below it, in which case the drift at the cutoff must be zero or (ii) the drift is negative on both sides of the cutoff. In either case if the initial belief is the censorship cutoffs, beliefs never reach a belief above the initial belief at any point in the future.

As noted above, for any beliefs,

$$D(x) = \int_0^\infty e^{-(r+\lambda)t} \left[\mu(1-\pi_t)\Delta(x_t)\right) + k\pi_t] dt \leq k/(r+\lambda) < c/\lambda$$

since sequential rationality immediately bounds the integrand above by k, so $a(x) = 0$ is optimal for any beliefs. This immediately implies that $V(0, L) = 0$ in any equilibrium. It remains to find a x^*_π such that under the beliefs consistent with that cutoff, $\mu V(x^*_\pi, L) = k$ (or $x^*_\pi = 1$, and $\mu V(x, L) < k$ for all L)

To avoid the potential discontinuity issues, I characterize the optimal strategy under a set of auxiliary beliefs and show that it also describes the equilibrium strategy. Let beliefs be

$$\tilde{a}(x) = 0$$
$$\tilde{\pi}(x) = \max\left\{0, 1 - \frac{\lambda}{\mu(1-x)}\right\}.$$

Under these beliefs, the drift is continuous and always weakly negative. Starting at any x_0, these beliefs induce the same future belief trajectories as the beliefs consistent with the cutoffs $x_a = 1, x_\pi = x_0$ since under both sets of beliefs at x_π either (i) the drift is zero and/or (ii) beliefs are drifting down at x_π and the firm is believed to be playing $\pi = 0$ at every belief that is reached in the absence of news from x_0 (and after bad news beliefs are constantly zero).[10] Let $V_{nc}(x, \theta)$ be the value function that corresponds to these beliefs. This is strictly increasing and zero at zero. Let

$$\begin{aligned} x_\pi^* &= \sup\left\{x : \mu V_{nc}(x, L) \le k\right\} \\ \pi(x_\pi^*) &= \max\left\{0, 1 - \frac{\lambda}{\mu(1 - x_\pi^*)}\right\} \\ a(x) &= 0. \end{aligned}$$

This, combined with the corresponding beliefs, characterizes an equilibrium. By construction, under beliefs consistent with some cutoff x_π, $V(x_\pi, L) = V_{nc}(x_\pi, L)$. Moreover, by Lemma A.2, V_{nc} is continuous everywhere, except possibly at $\hat{x} := 1 - \lambda/\mu$, where the sign of the drift is negative at larger x's and zero at smaller x's.

In fact, it is continuous at this point as well. Fix any $\epsilon > 0$ and note that for any two initial conditions x_0, x_0' and corresponding optimal strategies π and π' and beliefs trajectories x_t, x_t', it follows from the optimality of strategies that

$$E_\pi \int_0^\infty e^{-rt}\left[x_t - x_t'\right]dt \ \ge\ V_{nc}(x_0, L) - V_{nc}(x_0', L) \ \ge\ E_{\pi'} \int_0^\infty e^{-rt}\left[x_t - x_t'\right]dt,$$

where the first inequality comes from the player with prior x_0' deviating to play the same strategy as the x_0 player after every history, and the second comes from the x_0 player making a similar deviation. The previous inequality implies that for any x' such that $|x' - \hat{x}| < r\epsilon$, then, since beliefs that start at x' are either drifting toward $\hat{x}$ for all t or are constant,

$$\left|V_{nc}(x', L) - V_{nc}(\hat{x}, L)\right| \le \epsilon,$$

so it is continuous everywhere. So either $\mu V_{nc}(x_\pi^*, L) = k$ or $x_\pi^* = 1$ and $\mu V_{nc}(x, L) < k$ for all x.

Now I consider the case where $(1 + r/\lambda)c \le k$.

Here, equilibria may exist with both investment and censorship. To show existence, I construct an equilibrium where $x_\pi^* \ge x_a^*$, whenever the firm is censoring, it is also investing. The main difficulty here is not only are value functions discontinuous but that as the believed cutoffs vary, the firm's payoffs can change discontinuously, making it difficult to apply standard fixed point arguments. To construct an equilibrium, I construct two "auxiliary games" where either the censorship or investment strategies are fixed and then establish a linkage between the optimal strategies under those beliefs and an MPE. The value functions from these games will be used extensively in the analysis of this case.

[10] When $1 - \lambda/(\mu(1 - x_0)) < 0$, beliefs drift down until $x = 1 - \lambda/\mu$ and then stop drifting under both belief specifications.

First, consider the game where the firm is free to choose any censorship strategy but investment is restricted so that $a(x) = \tilde{a}(x) = 1_{\{x>0\}}$. Let V_{fw} be the value function when beliefs are consistent with

$$\tilde{a}(x) = 1_{\{x>0\}}$$

$$\tilde{\pi}(x) = 1.$$

These are the beliefs a firm would face if they were believed to be censoring everywhere and investing everywhere except zero.

Define D_{fw} and Δ_{fw} analogously. Let

$$x_\pi^* = \max\{x \in [0,1] : \mu V_{fw}(x,L) \leq k\}.$$

I'll show that this cutoff describes an equilibrium of this auxiliary game.

Second, we consider the game where the firm is restricted to follow the equilibrium censorship strategy described by cutoff x_π^* in the first game. Let $V_y(x)$ correspond to the value function generated by beliefs consistent with cutoffs x_π^* and $x_a = y$ (with $\tilde{a}(0) = 0$, $\tilde{a}(y) = 1$ if $y > 0$). Define D_y and Δ_y analogously. Let

$$x_a^* = \max\{x \in [0, x_{\pi^*}] : \lambda D_0(x) \leq c\};$$

this turns out to be an equilibrium of the game with the restricted censorship strategy.

LEMMA A.4: *If $(1 + r/\lambda)c \leq k$, there exists an equilibrium where $x_\pi = x_\pi^*$ and $x_a = x_a^*$ and $a(0) = 0$.*

PROOF:

First consider the belief dynamics that generate V_{fw}. The drift of beliefs is positive everywhere except zero. Moreover, on $(0,1]$ the value function V_{fw} is continuous and strictly increasing and at $0 V_{fw}(0,\theta) = 0$. So x_π^* is well defined and either $\mu V_{fw}(x_\pi^*, L) = k$ or $x_\pi^* = 1$ and $V_{fw}(1,L) < k$. Moreover, $V_{fw}(x_\pi^*, L)$ is equal to the value in auxiliary game where the firm is believed to play cutoff x_π^*, as the believed strategies only differ on $(0, x_\pi^*)$, which can't be reached under any strategy profile since beliefs drift up at x_π^* and zero is an absorbing state.

Now consider the other game. Given effort cutoff y, beliefs that start at or above y never reach a nonzero belief below y, so under any strategy, the firm faces exactly the same continuation beliefs as they would if they were working everywhere except zero. Moreover, since $D_y(x)$ is continuous above y, on the interior

$$D_0(y) = \lim_{x \to y+} D_y(x) = D_y(y).$$

So $x \mapsto D_x(x)$ is strictly increasing and continuous on $(0,1)$, $D_0(0) = 0$ and for $x \geq x_\pi^*$, payoffs agree with the *fw* payoffs and

$$D_x(x) = \int_0^\infty e^{-(r+\lambda)t} k dt \geq c/\lambda.$$

So x_a^* exists and is by definition less than x_π^*. By construction, if $x_\pi^* \neq 1$,

$$a(x) = 1_{\{x \geq x_a^*\}}$$

$$\pi(x) = 1_{\{x \geq x_\pi^*\}}$$

is an equilibrium since $V(x_\pi^*, L) = V_{fw}(x_\pi^*, L)$ and $D(x_a^*) = D_{x_a^*}(x_a^*)$. Similarly, if $x_\pi^* = 1$ and $x_a^* < 1$, then $a(x)$ defined as above and $\pi(x) = 0$ is an equilibrium by the same logic (since 1 is absorbing under both $\tilde{\pi}(1) = 1$ and $\tilde{\pi}(1) = 0$ if $\tilde{a}(1) = 1$).

Finally, if both cutoffs are equal to one, then $a(x) = \pi(x) = 0$ is an equilibrium since under these beliefs, x_t is drifting down at any x_0, as opposed to being constant and equal to one if $x_0 = 1$ (as it is in the problems that determine $V_{fw}(1,\theta)$ and $D_0(1)$). Given beliefs where consistent with no censorship or investment, $V(1,L) < V_{fw}(1,L) \leq k/\mu$. Similarly, $D(1) \leq \int_0^\infty e^{-(r+\lambda)t} \mu V_{fw}(1,L) dt = D_0(1) \leq c/\lambda$, so $a(1) = 0$ is optimal. ∎

A4. *Equilibrium Characterization and Proof of Theorem 1*

PROOF OF PROPOSITION 2:

This follows immediately from Lemma A.3. By the sequential rationality condition, $D(x)$ is bounded above by

$$D(x) \leq \frac{k}{r+\lambda},$$

so if $c > k\lambda/(r+\lambda)$, the firm never invests.

Recall that in any equilibrium $V_{nc}(x_\pi, L) = V(x_\pi, L)$ and $V_{nc}(x_\pi, L)$ is strictly increasing. So the equilibrium cutoff is unique. ∎

PROOF OF LEMMA 3:

Equilibrium payoffs in any full-shirk equilibrium are bounded below by zero. It remains to show that they are bounded above by something that goes to zero as $r \to 0$. Intuitively, bad news arrives in finite time almost surely, so most of the time, the firm is receiving a payoff of zero.

In equilibrium at x_0 either (i) beliefs are drifting down or (ii) beliefs are drifting up until they reach the point where either the firm wants to start censoring or $0 = x_t(\mu(1 - x_t) - \lambda)$. In the second case $\mu V(x_0, L) \leq k$, and as $D(x) \leq k/(r+\lambda)$, $V(x_0, H) \leq k/\mu + k/(r+\lambda)$.

Suppose beliefs are drifting down at x_0. Let $\tau = \inf\{t : x_t \leq x_\pi\}$. Above the censorship cutoff $\dot{x}_t = -\lambda x_t$, so $x_t = x_0 e^{-\lambda t}$. Therefore, payoffs satisfy

$$\begin{aligned} V(x_0, L) &\leq \int_0^\tau x_0 e^{-rt} e^{-\lambda t} dt + e^{-r\tau}\frac{k}{\mu} = \frac{x_0}{r+\lambda}\left(1 - e^{-(r+\lambda)\tau}\right) + e^{-r\tau}\frac{k}{\mu} \\ &\leq \frac{x_0}{r+\lambda} + \frac{k}{\mu}. \end{aligned}$$

Therefore, $V(x_0, L) \leq x_0/(r+\lambda) + k/\mu$ and $V(x_0, H) \leq (x_0+k)/(r+\lambda) + k/\mu$, so $\lim_{r\to 0} rV_r^* = 0$. ∎

PROOF OF PROPOSITION 3:

If the initial reputation, x_0, is above $\max(x_a, x_\pi)$, then $\dot{x}_t = \lambda(1 - x_t)$. This differential equation with initial condition x_0 is solved by $x_t = 1 - (1 - x_0)\, e^{-\lambda t}$. We can thus simplify V_{fw} from Proposition 1 by noting that, except at zero (where both are equal to zero),

$$\begin{aligned} V_{fw}(x, H) &= \int_0^\infty e^{-rt}\big[1 - (1-x)\, e^{-\lambda t} - c\big]dt = \frac{1-c}{r} - \frac{1-x}{r+\lambda}; \\ V_{fw}(x, L) &= \int_0^\infty e^{-(r+\lambda)t}\big[1 - (1-x)\, e^{-\lambda t} - c - k + \lambda V_{fw}(x_t, H)\big]dt \\ &= \frac{1-c}{r} - \frac{k}{r+\lambda} - \frac{1-x}{r+\lambda}. \end{aligned}$$

These are exactly the payoffs a firm would receive in any equilibrium starting at any $x_0 > \max\{x_a, x_\pi\}$.

Only if: For the firm to have persistent perfect reputation in an equilibrium where $a(0) = 0$, it must be that $x_0 \geq \max\{x_a, x_\pi\}$. Suppose such an equilibrium exists but $k > (\lambda + rx_0 - c(r+\lambda))\mu/(r(r+\lambda+\mu))$. Then $V(x_0, L) = V_{fw}(x_0, L)$ and $V(0, L) = 0$. But then it must be the case that

$$k \leq \mu\big(V(x_0, L) - V(0, L)\big)$$

$$k \leq \mu\left(\frac{1-c}{r} - \frac{k}{r+\lambda} - \frac{1-x_0}{r+\lambda}\right)$$

$$k \leq \frac{\big(\lambda + rx_0 - c(r+\lambda)\big)\mu}{r(r+\lambda+\mu)},$$

so this equilibrium cannot exist.

If: The equilibrium constructed in the Proof of Proposition 1 sets $x_\pi = \max\{x : \mu V_{fw}(x, L) \leq k\}$, so $x_0 \geq x_\pi \geq x_a$ in that equilibrium if $\mu V_{fw}(x_0, L) \geq k$, which can be rewritten as

$$k \leq \frac{\big(\lambda + rx_0 - c(r+\lambda)\big)\mu}{r(r+\lambda+\mu)}.$$

Finally, suppose there exists another equilibrium where $\Pr(\underline{\tau} < \infty) < 1$, where $\underline{\tau} = \inf\{t : x_s = 0 \ \forall s > t\}$ but the firm also does not have persistent perfect reputation a.s. For this to be possible, it must be that $x_\pi > x_0 \geq x_a > 0$. This means that $\mu V(x_0, L) \leq k$. But the firm can always deviate and censor all bad news and invest in quality, so

$$V(x_0, L) \geq \int_0^\infty e^{-(r+\lambda)t}\left[x_t - c - k + \lambda V(x_t, H)\right]dt > V_{fw}(x_0, L),$$

where the second inequality follows from the fact that $x_t > 1 - (1 - x_0)\,e^{-\lambda t}$ by Gronwall's inequality.[11] But then $\mu V(x_0, L) - V(0, L) > k$, so this cannot be an equilibrium. ∎

PROOF OF COROLLARY 1:

The first bullet is an immediate consequence of Proposition 3. The second bullet follows from the following observations. By the previous lemma, there is no equilibrium where the firm both invests and censors at *any* $x \in [0, 1)$ if $k \geq (1 - c)(r + \lambda)\mu/(r(r + \lambda + \mu))$. Suppose the firm censored but didn't invest at some x_0. Then censorship would still be optimal at x_0 if consumers believed the firm was investing and censoring at x_0, i.e., let $y_t = 1 - (1 - x_0)\,e^{-\lambda t}$, $y_t > x_t$ for all $t > 0$ (since x_t is drifting down whenever it's above x_π, while y_t is always drifting up), so payoffs would increase pointwise if the firm followed the same strategies but beliefs followed the trajectory described by y_t. Using the notation from Lemma A.4, these are the payoffs the firm would achieve at x_0 if they were believed investing and censoring everywhere except at zero, $V_{fw}(x_0, \theta)$. It must be that

$$k/\mu \leq V(x_0, L) < V_{fw}(x_0, L).$$

Therefore, censoring is also optimal at x_0 in the *fw* problem. But then

$$k \geq \frac{(\lambda + r x_0 - c(r + \lambda))\mu}{r(r + \lambda + \mu)} = V_{fw}(x_0, L)\mu > k,$$

which is a contradiction.

To see the final statement, there are two cases to consider, either (i) $k < (1 - c)(r + \lambda)\mu/(r(r + \lambda + \mu))$ and Proposition 3 implies this immediately or (ii) $k \geq (1 - c)(r + \lambda)\mu/(r(r + \lambda + \mu))$. Note that for $c > \mu\lambda/((r + \mu)(r + \lambda))$, $(1 + r/\lambda) > (1 - c)(r + \lambda)\mu/(r(r + \lambda + \mu))$, so the region where the firm can both invest and censor in equilibrium is empty.

[11] Let $y_t = 1 - x_t$. $\dot{y}_t \leq -y_t\lambda$ (strictly so between x_a and x_π), so $1 - x_t \leq (1 - x_0)\,e^{-\lambda t}$ (strict when $x_t \in (x_a, x_\pi)$).

In this second case if $x_a < 1$, then the firm has incentive to invest at reputation 1, so its payoffs in equilibrium are bounded by

$$V(1,L) \geq \int_0^\infty e^{-(r+\lambda+\mu)t}\left[1 - c + \lambda V(1,H)\right]dt = \frac{(r+\lambda)(1-c)}{r(r+\lambda+\mu)},$$

where the inequality follows from the firm playing never censoring and $V(0,L) \geq 0$. Therefore,

$$D(1) \leq \frac{1-c}{r} - \frac{(r+\lambda)(1-c)}{r(r+\lambda+\mu)} = \frac{\mu(1-c)}{r(r+\lambda+\mu)}.$$

So it is impossible for the firm to invest at one and thus invest in any equilibria if

$$\frac{\mu\lambda(1-c)}{r(r+\lambda+\mu)} < c$$

$$\frac{\mu\lambda}{(r+\lambda)(r+\mu)} < c.$$

Finally, to show that there's always an equilibrium with investment for high enough x_0, recall that there always exists an equilibrium where $a(0) = 0$ by the construction in Lemma A.4. Moreover, by the equilibrium construction in Lemma A.4, the firm invests at any point where $\lambda D_0(x) > c$, where $D_0(x)$ is the difference between the $\theta = H$ and $\theta = L$ value at x when the firm is believed to be investing everywhere except zero. In such an equilibrium

$$D_0(1) = D(1) = \frac{1-c}{r} - \frac{(r+\lambda)(1-c)}{r(r+\lambda+\mu)} = \frac{\mu(1-c)}{r(r+\lambda+\mu)},$$

so the firm invests at 1 if $\mu\lambda/\left((r+\lambda)(r+\mu)\right) > c$, and since $D_0(x)$ is continuous except at zero, $x_a < 1$. ∎

PROOF OF COROLLARY 2:

By Proposition 3, there exists an equilibrium where the firm has persistent perfect reputation if $k < \left(\lambda + rx_0 - c(r+\lambda)\right)\mu/\left(r(r+\lambda+\mu)\right)$. The right-hand side of this expression goes to ∞ as $r \to 0$, so for any $x_0 \in (0,1)$, there exists an $\bar{r}$ such that if $r < \bar{r}$, then an equilibrium with persistent perfect reputation exists. In such an equilibrium

$$rV(x_0,L) = r\left(\frac{1-c}{r} - \frac{k}{r+\lambda} - \frac{1-x_0}{r+\lambda}\right) \to 1-c$$

and

$$rV(x_0,H) = r\left(\frac{1-c}{r} - \frac{1-x_0}{r+\lambda}\right) \to 1-c,$$

so $\lim_{r\to 0} rV_r^* = 1-c$. ∎

LEMMA A.5: *Both the game with censorship and the game without censorship do not have any full-work equilibria if* $c > \lambda\mu(\mu+\lambda)/\big((r+\lambda)(r+\lambda+\mu) \times (r+2\lambda+2\mu)\big)$.

PROOF:

Suppose a full-work equilibrium exists. Then, in such an equilibrium, $\lambda D(0) \geq c$. In this equilibrium (with equality in the game without censorship)

$$V(0,L) \geq \int_0^\infty e^{-(r+\lambda+\mu)t}\big[x_t - c + \mu V(0,L) + \lambda V(x_t,H)\big]dt$$

So

$$\frac{r+\lambda}{r+\lambda+\mu}V(0,L) = \int_0^\infty (x_t - c)\left[\frac{\lambda}{\mu+\lambda}e^{-rt} + \frac{\mu}{\mu+\lambda}e^{-(r+\lambda+\mu)t}\right]dt.$$

$$\begin{aligned}
D(0) &= V(0,H) - V(0,L) \\
&= \int_0^\infty x_t\left(e^{-rt} - \frac{r+\lambda+\mu}{r+\lambda}\left[\frac{\lambda}{\mu+\lambda}e^{-rt} + \frac{\mu}{\mu+\lambda}e^{-(r+\lambda+\mu)t}\right]\right)dt \\
&\leq \int_0^\infty \big(1 - e^{-(\lambda+\mu)t}\big)\left(\frac{r\mu}{(r+\lambda)(\mu+\lambda)}e^{-rt} - \frac{r\mu}{(r+\lambda)(\mu+\lambda)}e^{-(r+\lambda+\mu)t}\right)dt \\
&= \frac{\mu(\mu+\lambda)}{(r+\lambda)(r+\lambda+\mu)(r+2\lambda+2\mu)},
\end{aligned}$$

where the second inequality follows from $\dot{x}_t \leq (1-x_t)(\lambda+\mu)$ and Gronwall's inequality.[12] So full-work is not an equilibrium if $c > \lambda\mu(\mu+\lambda)/\big((r+\lambda) \times (r+\lambda+\mu)(r+2\lambda+2\mu)\big)$. Note, as we take $r \to 0$, if $c > \mu/\big(2(\lambda+\mu)\big)$, then full-work cannot be an equilibrium for small enough r. ∎

PROOF OF PROPOSITION 4:

LEMMA A.6: *In any shirk-work equilibrium*

$$r\big(x_0 V(\hat{x},H) + (1-x_0)V(\hat{x},L)\big) \leq (1-c)\left(x_0 + (1-x_0)\frac{r+\lambda}{r+\lambda+\mu}\right).$$

PROOF:

Fix any equilibrium with $x_a \in (0,1)$ and let $\hat{x} = \max(x_0,x_a)$. The value from low quality satisfies

$$\begin{aligned}
V(\hat{x},L) &= \int_0^\infty e^{-(r+\lambda+\mu)t}\left[x_t - c + \lambda\int_t^\infty e^{-r(s-t)}(x_s - c)ds\right]dt \\
&= \int_0^\infty (x_t - c)\left[\frac{\lambda}{\lambda+\mu}e^{-rt} + \frac{\mu}{\lambda+\mu}e^{-(r+\lambda+\mu)t}\right]dt.
\end{aligned}$$

[12] $y_t = x_t - 1$, $\dot{y}_t \leq -y_t(\lambda+\mu)$, so by Gronwall, $x_t - 1 \leq -e^{-(\lambda+\mu)t}$.

So ex ante payoffs are bounded above by

$$
\begin{aligned}
& r\big(x_0 V(\hat{x},H) + (1-x_0)V(\hat{x},L)\big) \\
& = r\int_0^\infty (x_t - c)\left[\left(x_0 + (1-x_0)\frac{\lambda}{\lambda+\mu}\right)e^{-rt} + (1-x_0)\frac{\mu}{\lambda+\mu}e^{-(r+\lambda+\mu)t}\right]dt \\
& \leq r\int_0^\infty \big(1-(1-\hat{x})\,e^{-(\mu+\lambda)t} - c\big) \\
& \qquad \times \left[\left(x_0 + (1-x_0)\frac{\lambda}{\lambda+\mu}\right)e^{-rt} + (1-x_0)\frac{\mu}{\lambda+\mu}e^{-(r+\lambda+\mu)t}\right]dt \\
& = (1-c)\left(x_0 + (1-x_0)\frac{r+\lambda}{r+\lambda+\mu}\right) - (1-\hat{x})\frac{r}{r+\lambda+\mu} \\
& \qquad \times \left(x_0 + (1-x_0)\frac{r+2\lambda+\mu}{r+2\lambda+2\mu}\right),
\end{aligned}
$$

which, by just dropping the negative term, implies the desired bound. It follows immediately from Lemma A.5 that for sufficiently small r if $c > \mu/2(\lambda+\mu)$, then full-work cannot be an equilibrium.

To complete the Proof of Proposition 4 and (and Theorem 1), all that remains is showing the payoffs in the full-shirk equilibrium converge to zero as $r \to 0$. This is an immediate consequence of Lemma 5 if any other equilibrium exists, as the payoffs in any full-shirk equilibrium must be lower than the payoffs in any equilibrium with investment (the firm in the other equilibrium could always deviate to never investing and would receive a higher payoff than in the full-shirk equilibrium due to more favorable market beliefs). In fact, in the limit as $r \to 0$, it's clear that payoffs in a full-shirk equilibrium go to zero. To complete the argument for the case where the full-shirk equilibrium is the unique equilibrium, I construct a very simple bound. ∎

LEMMA A.7: *In a full-shirk equilibrium*

$$
r\big(x_0 V(x_0,H) + (1-x_0)V(x_0,L)\big) \leq r\left(x_0\frac{(r+\mu+\lambda)\hat{x}}{(r+\lambda)(r+\mu)} + (1-x_0)\frac{\hat{x}}{r+\mu}\right),
$$

where $\hat{x} = \max\{x_0, 1-\lambda/\mu\}$.

PROOF:

Let $\hat{x} = \max\{x_0, 1-\lambda/\mu\}$. In any full-shirk equilibrium, at x_0, until bad news arrives, either beliefs are drifting down or beliefs are drifting up until they reach $\hat{x}$, where the drift of beliefs is zero. So for all t, $x_t \leq \hat{x}$. Therefore,

$$
V(x_0,L) \leq \int_0^\infty e^{-(r+\mu)t}\hat{x}dt = \frac{\hat{x}}{r+\mu},
$$

and

$$
V(x_0,H) \leq \int_0^\infty e^{-(r+\lambda)t}\left[\hat{x} + \lambda\frac{\hat{x}}{r+\mu}\right]dt = \frac{(r+\mu+\lambda)\hat{x}}{(r+\lambda)(r+\mu)}.
$$

So

$$x_0 V(x_0,H) + (1-x_0)V(x_0,L) \le x_0 \frac{(r+\mu+\lambda)\hat{x}}{(r+\lambda)(r+\mu)} + (1-x_0)\frac{\hat{x}}{r+\mu}.$$

Note that $r\left[x_0(r+\mu+\lambda)\hat{x}/\left((r+\lambda)(r+\mu)\right) + (1-x_0)\hat{x}/(r+\mu)\right]$ goes to zero as $r \to 0$.

This lemma immediately implies the bound in Proposition 4. The expression

$$\left(x_0 \frac{(r+\mu+\lambda)\hat{x}}{(r+\lambda)(r+\mu)} + (1-x_0)\frac{\hat{x}}{r+\mu}\right)$$

is decreasing in r and increasing in $\hat{x}$, so plugging in $r = 0, \hat{x} = 1$ leads to the bound from the proposition.

Therefore, in any equilibrium that is not full-work, as we take $r \to 0$, payoffs stay bounded away from the first best, which when combined with Lemma 3, Corollary 2, and Proposition 4, gives Theorem 1. ∎

REFERENCES

Besley, Timothy, and Andrea Prat. 2006. "Handcuffs for the Grabbing Hand? Media Capture and Government Accountability." *American Economic Review* 96 (3): 720–36.

Board, Simon, and Moritz Meyer-ter-Vehn. 2013. "Reputation for Quality." *Econometrica* 81 (6): 2381–2462.

Daughety, Andrew F., and Jennifer F. Reinganum. 2008. "Products Liability, Signaling, and Disclosure." *Journal of Institutional and Theoretical Economics JITE* 164 (1): 106–26.

Dye, Ronald. 2013. "Voluntary Disclosure and the Duty to Disclose." Working Paper.

Fudenberg, Drew, and David K. Levine. 1989. "Reputation and Equilibrium Selection in Games with a Patient Player." *Econometrica* 57 (4): 759–78.

Fudenberg, Drew, and David K. Levine. 1992. "Maintaining a Reputation when Strategies are Imperfectly Observed." *Review of Economic Studies* 59 (3): 561–79.

Halac, Marina, and Andrea Prat. 2016. "Managerial Attention and Worker Performance." *American Economic Review* 106 (10): 3104–32.

Hall, Shannon. 2015. "Exxon Knew About Climate Change Almost 40 Years Ago." *Scientific American*, October 26, 2015. https://www.scientificamerican.com/article/exxon-knew-about-climate-change-almost-40-years-ago/.

Hauser, Daniel N. 2021. "Promoting a Reputation for Quality." Unpublished.

Hilts, Philip. 1994a. "Scientists Say Cigarette Company Suppressed Findings on Nicotine." *New York Times*, April 29, 1994. https://www.nytimes.com/1994/04/29/us/scientists-say-cigarette-company-suppressed-findings-on-nicotine.html.

Hilts, Philip. 1994b. "Tobacco Company was Silent on Hazards." *New York Times*, May 7, 1994. https://www.nytimes.com/1994/05/07/us/tobacco-company-was-silent-on-hazards.html.

Klein, Nicolas, and Sven Rady. 2011. "Negatively Correlated Bandits." *Review of Economic Studies* 78 (2): 693–732.

Kreps, David M., and Robert Wilson. 1982. "Reputation and Imperfect Information." *Journal of Economic Theory* 27 (2): 253–79.

Mailath, George J., and Larry Samuelson. 2015. "Reputations in Repeated Games." In *Handbook of Game Theory with Economic Applications*, Vol. 4, edited by H. Peyton Young and Shmuel Zamir, 165–238. Amsterdam: Elsevier.

Marinovic, Iván, and Felipe Varas. 2016. "No News is Good News: Voluntary Disclosure in the Face of Litigation." *RAND Journal of Economics* 47 (4): 822–56.

Marinovic, Iván, Andrzej Skrzypacz, and Felipe Varas. 2018. "Dynamic Certification and Reputation for Quality." *American Economic Journal: Microeconomics* 10 (2): 58–82.

Milgrom, Paul, and John Roberts. 1982. "Predation, Reputation, and Entry Deterrence." *Journal of Economic Theory* 27 (2): 280–312.

O'Connor, Anahad. 2016. "How the Sugar Industry Shifted Blame to Fat." *New York Times*, September 12, 2016. https://www.nytimes.com/2016/09/13/well/eat/how-the-sugar-industry-shifted-blame-to-fat.html.

Shadmehr, Mehdi, and Dan Bernhardt. 2015. "State Censorship." *American Economic Journal: Microeconomics* 7 (2): 280–307.

Sun, Yiman. 2018. "A Dynamic Model of Censorship." Unpublished.

Tabuchi, Hiroko. 2020. "A Secret Recording Reveals Oil Executives Private Views on Climate Change." *New York Times*, September 12, 2020. https://www.nytimes.com/2020/09/12/climate/methane-natural-gas-flaring.html.

Tadelis, Steven. 2016. "Reputation and Feedback Systems in Online Platform Markets." *Annual Review of Economics* 8: 321–40.