Make It 'Til You Fake It: Supplement

Raphael Boleslavsky
Indiana University
rabole@iu.edu

Curtis R. Taylor
Duke University
curtis.taylor@duke.edu

June 29, 2023

Abstract

This supplement covers two model variants not included in the main draft.

1 Opaque Standards

In this section we analyze an environment in which the principal can be one of two types. With probability ν she has a high standard, θ_H, and with probability $1-\nu$, a low one, θ_L where $\theta_H > \theta_L$. With "transparent standards" the principal's type is observed by the agent at the beginning of the interaction. For each realization of the principal's type, the equilibrium is identical to the main model. With "opaque standards" the agent cannot observe the principal's type. This potential remedy is particularly salient in settings where the decision to approve is made by a group, and the allocation of real authority is unknown to the agent. For example, if two homeowners with different preferences decide whether to accept or reject an offer on their house, the realtor may not be sure which homeowner has the final say.

The goal is to characterize the equilibrium with opaque standards, and analyze its normative properties. Denote the acceptance strategy of each principal $a_i(\cdot)$, where $i \in \{H, L\}$, and let $a_U(t) \equiv \nu a_H(t) + (1-\nu)a_L(t)$ denote the expected probability of acceptance at time t, accounting for the agent's uncertainty about the principal's type.[1] The agent's expected payoff of selecting cheating time t is identical to his payoff in the main model, substituting the *expected* acceptance probability $a_U(\cdot)$ for the acceptance probability $a(\cdot)$ of the main model. Similar arguments to those in Lemma 4.3 (in the main text) establish that the agent mixes continuously on an interval from time zero to some finite threshold $\bar{t}_U$, defining a finite "phase of doubt." Furthermore, over this interval, the *expected* acceptance probability inherits the features of the acceptance probability in the main model: $a_U(\cdot)$ is strictly greater than ϕ, increasing, continuous, differentiable, and approaches one at the end of the phase of doubt. After the phase of doubt, the acceptance probability is one, $a_U(t) = 1$ for $t > \bar{t}_U$.

With opaque standards, the agent's mixing distribution is the same, regardless of what standard he actually faces. In other words, both types of principal face the same mixed strategy. Because the principal's type orders her payoff according to single-crossing, it cannot be that both types of principal are simultaneously indifferent between accepting and rejecting. Consequently, if one type of principal mixes in equilibrium, then the other strictly prefers accepting or rejecting. Furthermore, the low standards principal has a stronger incentive to accept: thus, whenever the high type mixes, the low type accepts, and whenever the low type mixes, the high type rejects. As we show in the following lemma, this ordering of the principal's incentives implies that under opaque standards, the phase of doubt is divided into two sub-phases. In the first (possibly degenerate) sub-phase the low standards principal

[1] Because the principal makes a decision only *after* the agent makes a submission and her type concerns her preferences, the agent does not update beliefs about the principal over time.

mixes and the high standards principal rejects. In the second sub-phase, the low standards principal accepts, and the high standards principal mixes.

Lemma 1.1 (Opaque standards equilibrium structure.). *In equilibrium with uncertain standards, there exists $\bar{t}_U \in (0, \infty)$ and $\widetilde{t}_U \in [0, \bar{t}_U)$ such that*

(i) the agent's cheating time is drawn from a continuous mixed strategy with no mass points or gaps supported on an interval $[0, \bar{t}_U]$.

(ii) for $t \in [\bar{t}_U, \infty)$, both types of principal accept the project, $a_L(t) = a_H(t) = 1$.

(iii) for $t \in [\widetilde{t}_U, \bar{t}_U)$ the low standards principal always accepts, $a_L(t) = 1$, and the high standards principal's acceptance strategy is strictly increasing, continuous, and differentiable almost everywhere, with $\lim_{t \to \bar{t}_U} a_H(t) = 1$.

(iv) if $\widetilde{t}_U > 0$, then for $t \in [0, \widetilde{t}_U)$ the high standards principal always rejects $a_H(t) = 0$, and the low standards principal's acceptance strategy $a_L(\cdot)$ is strictly increasing, continuous, and differentiable, with $\lim_{t \to \widetilde{t}_U} a_L(t) = 1$. Furthermore, $\lim_{t \to \widetilde{t}_U} a_H(t) = 0$.

To complete the characterization, we separately consider equilibria with a "one stage" structure, corresponding to the case $\widetilde{t}_U = 0$, and a "two stage" structure, corresponding to $\widetilde{t}_U > 0$. First, we introduce some additional notation that simplifies the exposition. For $i \in \{H, L\}$, let

$$\mu_i \equiv \lambda \frac{1 - \theta_i}{\theta_i} \qquad \bar{t}_i \equiv -\frac{\ln(1 - \sigma)}{\mu_i} \qquad \delta_U \equiv \frac{-\ln(1 - \frac{\nu \widehat{\phi}}{\widehat{\phi} - \phi})}{\rho} \qquad \nu^* \equiv (1 - \frac{\phi}{\widehat{\phi}})(1 - \exp(-\rho \bar{t}_H))$$

Note that μ_i is the equilibrium cheating rate when the agent observes that he faces a type $i \in \{H, L\}$ principal. In other words, it is the equilibrium cheating rate under transparency for the type i principal. Similarly, $\bar{t}_i$ is the duration of the phase of doubt under transparency with a type i principal. By implication, $\mu_L > \mu_H$ and $\bar{t}_L < \bar{t}_H$. As we will see, δ_U is the duration of the second stage in the two-stage equilibrium. Note that δ_U is well-defined whenever $\nu < 1 - \phi/\widehat{\phi}$. Finally, we will also see that the relationship between ν and ν^* determines whether the equilibrium has one or two stages of faking.

In a one stage equilibrium, the low standards principal accepts all arrivals, while the high standards principal mixes for arrivals before $\bar{t}_U$ and accepts thereafter. Because only the high type principal mixes in the phase of doubt, in such equilibria, the agent behaves as if he faces *only* the high type principal. Thus, the agent mixes over the same phase of doubt as in the main model, $[0, \bar{t}_H]$. By implication, the low standards principal strictly prefers acceptance in both phases ($\theta_L < \theta_H < 1$). Furthermore, from the agent's perspective, the expected acceptance probability is the same as in the main model. However, because the low standards principal always accepts, the high standards principal's acceptance strategy must be adjusted to maintain the same expected acceptance probability as in the main model, $\nu a_H(t) + (1-\nu) = a(t)$. Therefore, a one stage equilibrium exists only if ν is relatively large: if ν is small, then $1-\nu > a(0)$, which would imply $a_H(0) < 0$. Intuitively, if the probability of a low standards principal is high, then the probability that an early arrival is accepted is also high. Consequently, the strategic agent will be tempted to cheat early, even if he is rejected by the high type. Given the low probability of a high-standards principal, no adjustment in the high type's acceptance probability can offset this.

Proposition 1.1 (Opaque standards, One Stage.). *With opaque standards, a one phase equilibrium exists if and only if $\nu > \nu^*$, and it is characterized below. Furthermore, with opaque standards, no other one phase equilibrium exists.*

Strategies. *The agent's cheating time is drawn from distribution function*

$$F(t) = \frac{1}{\sigma}(1 - \exp(-\mu_H t))$$

supported on interval $[0, \bar{t}_H]$. If $t \in [0, \bar{t}_H]$, then the high type principal accepts with probability

$$a_H(t) = \frac{1}{\nu}\Big(\frac{\phi}{\widehat{\phi}} + (1 - \frac{\phi}{\widehat{\phi}})\exp\{-\rho(\bar{t}_H - t)\} - (1-\nu)\Big),$$

and with probability 1 otherwise. The low type principal always accepts, $a_L(t) = 1$. The expected acceptance probability $a_U(\cdot)$ is identical to the acceptance probability in the main model, with principal's standard known to be θ_H.

Beliefs. *If $t \in (0, \bar{t}_H)$, then $g(t) = \theta_H$, and $g(t) = 1$ otherwise.*

Payoffs. *The strategic agent's equilibrium payoff is $a_U(0) - \phi$, identical to the main model with principal's standard known to be θ_H. The high standards principal's payoff is*

$$V_H = (1-\theta_H)(1-\sigma)\int_{\bar{t}_H}^{\infty} \lambda \exp\{-(\rho+\lambda)t\}dt,$$

and the low type principal's payoff is

$$V_L = \lambda(1 - \frac{\theta_L}{\theta_H})\int_0^{\bar{t}_H} \exp\{-(\rho + \frac{\lambda}{\theta_H})t\}dt + (1-\theta)(1-\sigma)\int_{\bar{t}_H}^{\infty} \lambda\exp\{-(\rho+\lambda)t\}dt.$$

Normative Ranking. *In the one phase equilibrium with opaque standards, (i) the high type principal's payoff is the same as in the unique equilibrium with transparent standards. (ii) The low type principal's payoff is strictly higher than in the unique equilibrium with transparent standards.*

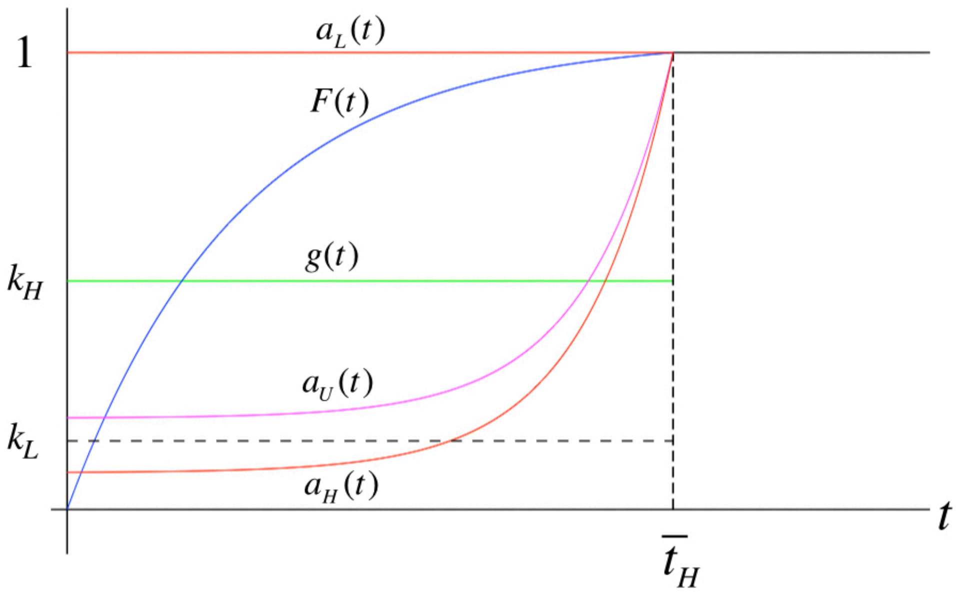

Figure 1: *Proposition 1.1*

When the principal is probable to have high standards, opacity (weakly) increases the payoffs of both types of principal. Because the agent is most likely to interact with the high standards principal, it is the high type's incentive to accept that restrains the strategic agent's incentive to cheat. In equilibrium, the strategic type effectively targets only the high type principal, completely ignoring the low type. In particular, the agent's strategy is identical to the baseline model, assuming that the principal's standard is known to be θ_H. Thus, the high type principal obtains the same equilibrium payoff with opacity as with transparency. In contrast, when the principal's realized standard is low, she accepts every submission, obtaining a positive payoff in both the doubt and credibility phases. However, the agent cheats more slowly with opacity than transparency if he faces the low type principal; thus, opacity also delays the onset of the credibility phase for the low type. The low type principal faces a tradeoff with opacity: a higher belief during the phase of doubt (and acceptance of all arrivals), but a longer phase of doubt. It turns out that the benefit generated by a higher belief during the phase of doubt outweighs the delay in restoring credibility; i.e., opacity strictly benefits the low type principal.

We turn next to the two stage equilibrium, in which $\widetilde{t}_U > 0$. The second stage resembles the one stage equilibrium—the low type principal always accepts and high type mixes. In contrast to the one stage equilibrium, however, the high type's acceptance probability begins at zero, $a(\widetilde{t}_U) = 0$, finishing at one, $a(\overline{t}_U) = 1$. Furthermore, in the first phase, the high type principal rejects, while the low type principal mixes. The low type's acceptance probability is positive at time zero, and increases during the first stage hitting one at the transition time $\widetilde{t}_U$. The agent's mixed strategy, while continuous, also takes a different form in the two sub-phases. In the first sub-phase, the agent cheats at a faster rate, inducing mixing by the low type principal; in the second, the agent cheats more slowly, inducing mixing by the high type principal. The agent and principal indifference conditions over the two stages, combined with the appropriate continuity boundary conditions define a system of differential equations that characterize the equilibrium.

Proposition 1.2 (Opaque standards, Two Stages.). *With opaque standards, a two stage equilibrium exists if and only if $\nu < \nu^*$, and it is characterized below. Furthermore, with opaque standards, no other two stage equilibrium exists.*

Stage Transitions. *The transition times $\widetilde{t}_U, \bar{t}_U$ in the two stage equilibrium are*

$$\widetilde{t}_U = \bar{t}_L - \frac{\mu_H}{\mu_L}\delta_U \qquad \bar{t}_U = \widetilde{t}_L + \delta_U.$$

Furthermore, $0 < \widetilde{t}_U < \bar{t}_U < \bar{t}_H$.

Strategies. *The agent's cheating time is drawn from continuous distribution function*

$$F(t) = \begin{cases} \frac{1}{\sigma}(1 - \exp(-\mu_L t)) & \text{for } t \in [0, \widetilde{t}_U) \\ \frac{1}{\sigma}(1 - \exp(-\mu_H t - (\mu_L - \mu_H)\widetilde{t}_U)) & \text{for } t \in [\widetilde{t}_U, \bar{t}_U] \end{cases}$$

supported on $[0, \bar{t}_U]$. If $t \in [0, \widetilde{t}_U]$, then the high type principal always rejects, $a_H(t) = 0$, and the low type principal accepts with probability

$$a_L(t) = (\frac{1}{1-\nu})(\frac{\phi}{\widehat{\phi}} + (1 - \frac{\phi}{\widehat{\phi}} - \nu)\exp\{-\rho(\widetilde{t}_U - t)\}).$$

If $t \in [\widetilde{t}_U, \bar{t}_U]$, then the high type principal accepts with probability

$$a_H(t) = \frac{1}{\nu}\Big(\frac{\phi}{\widehat{\phi}} + (1 - \frac{\phi}{\widehat{\phi}})\exp\{-\rho(\bar{t}_U - t)\} - (1-\nu)\Big),$$

and the low type principal always accepts, $a_L(t) = 1$. If $t \geq \bar{t}_U$, then both types of principal always accept, $a_L(t) = a_H(t) = 1$.

Beliefs. *If $t \in (0, \widetilde{t}_U)$, then $g(t) = \theta_L$. If $t \in (\widetilde{t}_U, \bar{t}_U)$, then $g(t) = \theta_H$. Otherwise $g(t) = 1$.*

Payoffs. *The agent's equilibrium payoff is $a_U(0) - \phi$. The high standards principal's payoff is*

$$V_H = (1-\theta_H)(1-\sigma)\int_{\bar{t}_U}^{\infty} \lambda \exp\{-(\rho+\lambda)s\}ds.$$

The low type principal's payoff is

$$V_L = \exp(-(\mu_L - \mu_H)\widetilde{t}_U)(1 - \frac{\theta_L}{\theta_H})\int_{\widetilde{t}_U}^{\bar{t}_U} \lambda \exp(-(\rho + \frac{\lambda}{\theta_H})t)dt + (1-\theta_L)(1-\sigma)\int_{\bar{t}_U}^{\infty} \lambda \exp(-(\rho+\lambda)t)dt.$$

Normative Ranking. *In the two stage equilibrium with opaque standards, (i) the high type principal's payoff is strictly higher than in the unique equilibrium with transparent standards. (ii) The low type principal's payoff is strictly higher than in the unique equilibrium with transparent standards.*

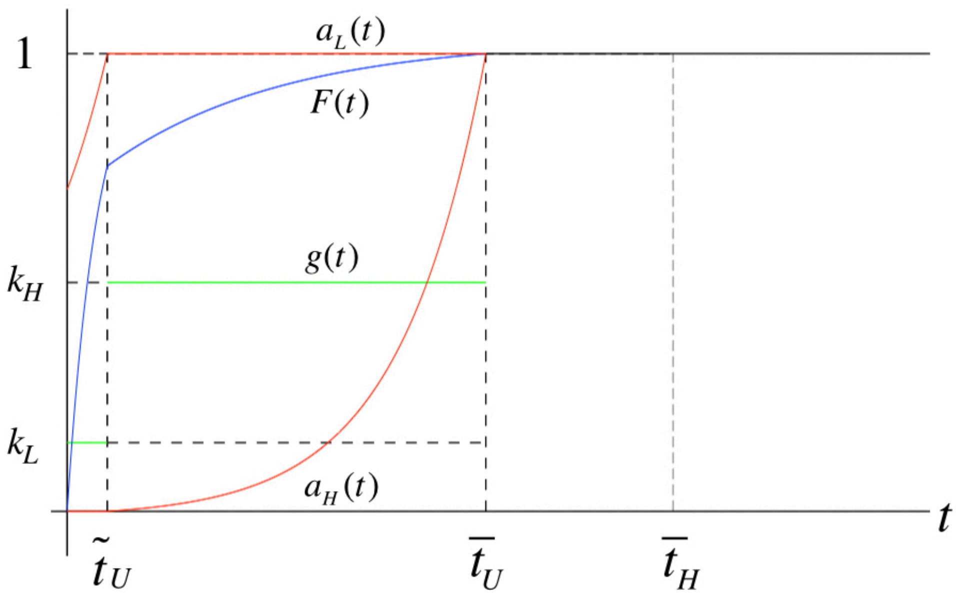

Figure 2: *Proposition 1.2. Note that $g(\cdot)$ jumps from θ_L to θ_H at $\tilde{t}_U$.*

When the probability of the low type principal is sufficiently high, then the agent does not ignore her—as he did in the one phase equilibrium. Instead, in the initial stage, $t \in [0, \tilde{t}_U)$, the agent cheats aggressively, gambling that the evaluator is the low type, who accepts even early arrivals with positive probability. As a consequence of this increase in the cheating rate, the high type principal is strictly better off than under transparency. Recall that under transparency, the agent cheats with rate μ_H throughout the phase of doubt. However, in the two stage equilibrium with opacity, the agent starts off cheating at rate $\mu_L > \mu_H$, and switches to rate μ_H at some interior time. Thus, the agent's overall credibility is restored more quickly, at time $\bar{t}_U < \bar{t}_H$. Because the high type principal's payoff is determined exclusively by the duration of the phase of doubt, she strictly benefits from opacity. The low type faces an initial phase with a high cheating rate and no surplus, a second phase with a lower cheating rate and positive surplus, and finally the restoration of full credibility. Though it takes longer for the agent's credibility to be restored fully, the low type expects positive surplus in the second phase. On net, she also benefits from opacity.

It is worth pointing out that endowing the principal with a disclosure technology by which she can verify her type to the agent at the outset does not undermine this equilibrium. That is, even if the principal has such a technology, an equilibrium exists in which neither type of principal uses it.[2]

2 External Impediments

In this section, we consider external impediments that prevent the principal from approving the project immediately. In particular, we suppose that the principal initially faces a "logjam." If a project arrives during the logjam, the principal observes the arrival time, but she must delay her approval decision until the logjam clears, an event which occurs at Poisson rate γ and is observed privately by the principal.[3] This situation might arise, for example, if the principal is occupied with other projects, or if the project must clear additional bureaucratic hurdles within the organization before it can be approved or implemented. As in the main model, payoffs are realized when the principal makes her decision. In this setting, it is convenient to interpret ρ as the rate at which the game ends rather than the rate of time preference. Then, by delaying the principal's decision, the logjam admits the possibility that the game ends before a submission can be approved.

There are multiple equilibria in this version of the game, but they are all expected payoff equivalent. Multiplicity arises because both types of the principal (unjammed and jammed) have the same beliefs about the project, and thus, the same approval incentives. In other words, given an arrival at time t, both types either strictly prefer to accept, strictly prefer to reject, or are indifferent. When both types are indifferent they may mix with different acceptance strategies in equilibrium—nevertheless the *expected probability of approval* is pinned down uniquely by the agent's indifference condition, which ensures that he is willing to mix.

With these qualifications in mind, we economize on notation and solve for a pooling equilibrium in which the unjammed and jammed types of principal use the same acceptance strategy, $a_J(t)$.[4] In other words, the unjammed type approves immediately with probabil-

[2]Because both types of principal prefer opacity to transparency (at least weakly), if the type i principal is expected not to disclose her type, then it is a best response for the type j principal not to disclose. In some cases, equilibria with partial disclosure also exist.

[3]If the agent can observe whether the principal is jammed, the principal may also benefit from the logjam in this case. We also have analyzed a case in which the jam can only clear after the agent submits a project. Thus, the jam is essentially a delay in the evaluation process, which can also benefit the principal. Details for both extensions available upon request.

[4]This also corresponds to an equilibrium of the model in which the principal does not observe whether she is jammed; e.g., the logjam comes from "red tape" in some other part of the organization.

ity $a_J(t)$ and the jammed type approves with this same probability as soon as she is able, provided the game does not end first.[5]

From the agent's perspective, the possibility of a logjam imposes an upper bound on the probability that a project is approved. In particular, if the project is submitted at time t and the principal would like to approve it, the probability that she is eventually able to do so is

$$\begin{aligned}\bar{a}(t) &\equiv \underbrace{1-\exp(-\gamma t)}_{\text{unjammed}} + \underbrace{\frac{\gamma}{\gamma+\rho}\exp(-\gamma t)}_{\text{jammed}} \\ &= 1 - \frac{\rho}{\rho+\gamma}\exp(-\gamma t).\end{aligned}$$

The first term is the probability that the logjam has cleared before t, in which case the principal can approve the project at the time it is submitted; the second term is the probability that the logjam is still present at time t, multiplied by the probability that the logjam clears before the game terminates.

To highlight our main case of interest, we focus on a logjam that clears slowly.

Assumption 1. *The rate at which the logjam clears is sufficiently low that $\bar{a}(t)$ is less than the equilibrium approval of the main model at time zero for all $\sigma \in (0,1)$,*

$$\frac{\gamma}{\gamma+\rho} < \frac{\phi}{\widehat{\phi}}.$$

To develop intuition for the effect of the logjam, consider a possible equilibrium in which the agent never fakes. In such a putative equilibrium, the principal always approves and hence the agent's best response is determined by the condition

$$\begin{aligned}u'(t) &= \exp(-(\lambda+\rho)t)\{\bar{a}'(t) - \rho\bar{a}(t) + \phi(\rho+\lambda)\} \\ &= \exp(-(\lambda+\rho)t)\rho\{\exp(-\gamma t) - (1-\frac{\phi}{\widehat{\phi}})\}.\end{aligned}$$

It follows that the agent would submit a fake at time $t^* \in (0,\infty)$, where

$$\exp(-\gamma t^*) = 1 - \frac{\phi}{\widehat{\phi}}.$$

Of course, if the agent did so, the principal would infer that a submission at this time is fake, which implies that such an equilibrium does not exist.[6] Unlike the main model, where

[5] The passage of time while waiting for a jam to clear does not change the principal's belief about whether a project that was submitted at t is real or fake.

[6] The corresponding calculation implies that the ethical type's payoff of submitting a project at time t is strictly decreasing when the acceptance probability is $\bar{a}(\cdot)$. Thus, the ethical type would like to submit a real project as soon as it arrives.

the agent would like to deviate by faking *immediately*, in the logjam model the agent would like to delay faking until $t^* > 0$. Intuitively, the agent expects that the principal is likely to be jammed initially, so there is less incentive for the agent to pay the cost of faking in order to rush his project out early.

This observation has implications for the equilibrium structure. When selecting his optimal cheating time, the agent can deviate from t^* both to later times ($t > t^*$) and to earlier times ($t < t^*$), which suggests that the support of the agent's equilibrium mixed strategy is an interval around t^*. Furthermore, if the logjam clears slowly, then t^* is large, which suggests that the bottom of the support is strictly positive.

Lemma 2.1. *(Logjam Equilibrium Structure). Let* $a(t) \equiv a_J(t)\bar{a}(t)$ *be the expected probability of approval. In an equilibrium with a logjam, there exist* $\overline{t}_J \in (t^*, \infty)$ *and* $\widetilde{t}_J \in (0, t^*)$ *such that*

(i) the agent's mixed strategy is drawn from a continuous mixed strategy with no mass points or gaps supported on interval $[\widetilde{t}_J, \overline{t}_J]$.

(ii) the principal's expected acceptance strategy $a(t)$ *is continuous and increasing for all* $t \geq 0$.

(iii) for $t \in (\widetilde{t}_J, \overline{t}_J)$, $a_J(t) < 1$ *so that* $a(t) < \overline{a}(t)$.

(iv) for $t \notin (\widetilde{t}_J, \overline{t}_J)$ $a_J(t) = 1$ *so that* $a(t) = \overline{a}(t)$.

This lemma highlights two main differences between the logjam equilibrium and the main model. First, with a logjam, there is an early phase ($t < \widetilde{t}_J$), in which the agent does not fake and the principal accepts with the maximum expected probability $\overline{a}(\cdot)$—there is no such phase in the main model. Because the principal is likely to be jammed at early times, there is less incentive for the agent to accelerate the arrival of his project by faking. Second, once the agent's credibility is fully restored ($t > \overline{t}_J$) the principal is constrained to accept with expected probability $\overline{a}(\cdot) < 1$. From an ex ante perspective, the first effect is beneficial to the principal, because she is able to accept early real arrivals some of the time, while the second effect is harmful, because it prevents her from accepting late real arrivals as often as she would like.

We complete the characterization with the following proposition.

Proposition 2.1. *(Logjam.) With a logjam, the pooling equilibrium of the game is characterized as follows.*

Strategies. *The agent's cheating time is drawn from continuous distribution function*

$$F(t) = \frac{1}{\sigma}(1 - \exp(-\mu(t - \widetilde{t}_J))$$

supported on interval $[\widetilde{t}_J, \overline{t}_J]$, *where* $\widetilde{t}_J$ *is such that*

$$\exp(-\gamma \widetilde{t}_J) = (1 - \frac{\phi}{\widehat{\phi}})(1 + \frac{\gamma}{\rho})\frac{1 - \exp(-\rho \overline{t})}{1 - \exp(-(\rho + \gamma)\overline{t})}$$

and $\overline{t}_J = \widetilde{t}_J + \overline{t}$. *If* $t \in (\widetilde{t}_J, \overline{t}_J)$, *then*

$$a(t) = \frac{\phi}{\widehat{\phi}} + (\overline{a}(\overline{t}_J) - \frac{\phi}{\widehat{\phi}}) \exp(-\rho(\overline{t}_J - t))$$

and $a(t) = \overline{a}(t)$ *otherwise.*

Beliefs. *If* $t \in (\widetilde{t}_J, \overline{t}_J)$, *then* $g(t) = \theta$, *and* $g(t) = 1$ *otherwise.*

Payoffs. *The strategic agent's equilibrium payoff is*

$$U_J = \int_0^{\widetilde{t}_J} \lambda \exp(-(\rho + \lambda)t)\overline{a}(t)dt + \exp(-(\rho + \lambda)\widetilde{t}_J)(\overline{a}(\widetilde{t}_J) - \phi).$$

The principal's ex ante equilibrium payoff is

$$V_J = (1 - \theta)(\int_0^{\widetilde{t}_J} \lambda \exp(-(\rho + \lambda)t)\overline{a}(t)dt + (1 - \sigma) \int_{\overline{t}_J}^{\infty} \lambda \exp(-(\rho + \lambda)t)\overline{a}(t)dt).$$

Normative Ranking. *The principal is strictly better off in the logjam equilibrium than in the equilibrium of the main model if* σ *is sufficiently large, and she is strictly worse off if* σ *is sufficiently small.*

With a logjam, the equilibrium has *two* phases of credibility, an early one $(0, \widetilde{t}_J)$, and a late one $(\overline{t}_J, \infty)$, with a single phase of doubt $[\widetilde{t}_J, \overline{t}_J]$, sandwiched between them. During the initial phase of credibility, the principal accepts all arrivals with maximum probability, but because the strategic type does not fake, the principal's belief about the agent's ethics does not evolve. When the phase of doubt is reached, the agent begins to fake at rate μ, as in the main model, which leaves the principal indifferent between accepting and rejecting. Because the principal's belief about the agent at time $\widetilde{t}_J$ is the same as at the beginning of the game and the agent fakes at the same rate during the phase of doubt, it takes the same amount of time for his credibility to be fully restored. That is, the duration of the phase of doubt is the same as in the main model, $\overline{t}_J - \widetilde{t}_J = \overline{t}$. Finally, we reach the second phase of credibility, in which the principal is confident that the agent is ethical and accepts all arrivals with maximum probability.

The initial phase of credibility generates a normative gain for the principal: an arrival in this phase is real and is accepted with expected probability $\bar{a}(\cdot)$. At the same time, the second phase of credibility generates a normative loss for the principal. Although the principal is

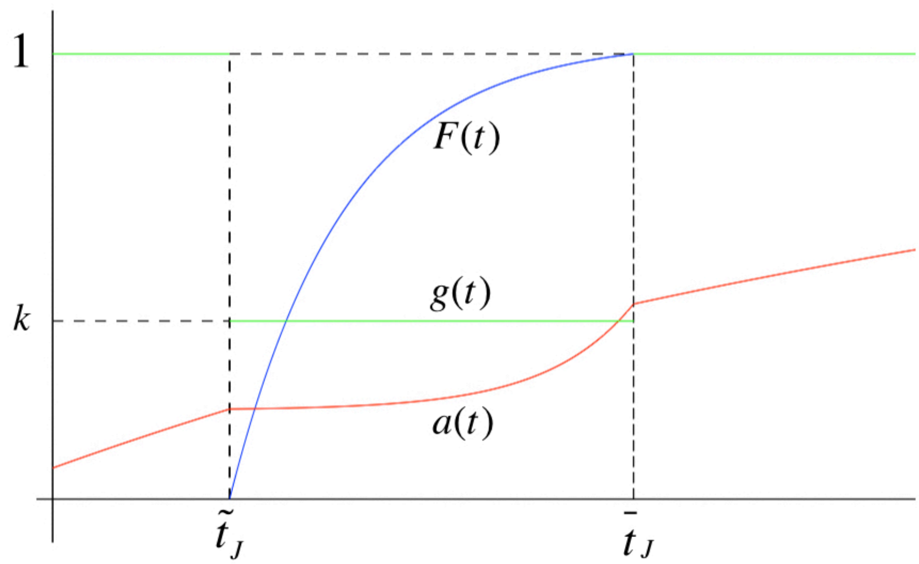

Figure 3: *Proposition 2.1. Note that* $g(\cdot)$ *jumps from* 1 *to* θ *at* $\widetilde{t}_J$ *and back up to 1 at* $\bar{t}_J$*.*

confident that such an arrival is real, the possibility of being jammed prevents her from accepting it with probability 1. In addition, the second phase of credibility is reached later than in the main model, at time $\bar{t}_J = \widetilde{t}_J + \bar{t}$, rather than at time $\bar{t}$. Thus, the positive surplus generated by such an arrival is discounted more heavily. The normative impact of the logjam depends on which of these effects dominates. When σ is large, the phase of doubt is very long in the main model and the principal's surplus approaches zero. We show that with the logjam, the duration of the *initial* phase of credibility approaches a strictly positive limit as σ gets large. Thus, the principal's expected surplus with the logjam is bounded away from zero. By implication, when she faces an agent who is likely to be strategic, the logjam benefits the principal. When σ becomes small, the duration of the phase of doubt collapses to zero. In the main model, the principal accepts all arrivals, which she is sure are real, and her payoff converges to the first best. With the logjam, she accepts all projects with maximum probability, and she is sure that they are real. However, her payoff is strictly smaller because she cannot approve all projects with probability 1. Consequently, the logjam is harmful if the agent is likely to be ethical.

A Proofs

A.1 Proofs for Opaque Standards

Proof of Lemma 1.1. Proof of (i). This point follows exactly from the arguments in the proof of Lemma 4.3 in the main text, replacing $a(\cdot)$ with $a_U(\cdot)$.

Step 0. We claim that for $t \in [0, \bar{t}_U)$, the expected acceptance probability $a_U(\cdot)$ is strictly greater than ϕ, strictly increasing, continuous, and differentiable almost everywhere, with $\lim_{t \to \bar{t}_U} a_U(t) = 1$. Furthermore, for $t \in [\bar{t}_U, \infty)$, the expected acceptance probability is 1, $a_U(t) = 1$. These points follow exactly from the arguments in the proof of Lemma 4.3 in the main text, replacing $a(\cdot)$ with $a_U(\cdot)$.

By analogy with the proof of Lemma 4.3 in the main text, let $\bar{t}_U \equiv \inf\{t : a_U(t) = 1\}$. Existence of $\bar{t}_U \in (0, \infty)$ is established by analogy with Lemma 4.3 in the main text.

Proof of (ii). From Step 0, if $t \geq \bar{t}_U$ then $a_U(t) = 1$, and hence, $\nu a_H(t) + (1 - \nu) a_L(t) = 1$. Because $a_L(t) \leq 1$ and $a_H(t) \leq 1$, it follows that $a_H(t) = a_L(t) = 1$.

Step 1. We show that for any $t \geq 0$, (a) if $a_H(t) > 0$, then $a_L(t) = 1$ and (b) if $a_L(t) < 1$, then $a_H(t) = 0$. Both (a) and (b) follow immediately from each type of principal's sequentially rational acceptance strategy, coupled with $\theta_H > \theta_L$.

Step 2. We show that for any $t \in [0, \bar{t}_U)$, exactly one of the following three conditions (A,B,C) must hold: (A) $a_H(t) \in (0, 1)$ and $a_L(t) = 1$, (B) $a_L(t) \in (0, 1)$ and $a_H(t) = 0$, (C) $a_L(t) = 1$ and $a_H(t) = 0$. From the definition of $\bar{t}_U$, we know that $a_U(t) < 1$ for $t < \bar{t}_U$. Hence, for such t, at least one of $a_i(t) < 1$ for $i \in \{H, L\}$.

If $a_L(t) < 1$, then $a_H(t) = 0$ from Step 1 (b). Furthermore, from Step 0, we know that $a_U(t) > \phi$. Coupled with $a_H(t) = 0$, this implies $a_L(t) > 0$. Hence, we have (B).

If $a_H(t) < 1$, then there are two possibilities. If $a_H(t) > 0$, then from Step 1 (a), we have $a_L(t) = 1$, case (A). If $a_H(t) = 0$, then we must have $a_L(t) > 0$ (otherwise $a_U(t) = 0$, contradicting Step 0). If $a_L(t) < 1$, then case (B). If $a_L(t) = 1$ then case (C).

Step 3. Consider $0 \leq t < t' < \bar{t}_U$. We show that (a) If $a_H(t) \in (0, 1)$ then $a_H(t) < a_H(t') < 1$ and $a_L(t') = 1$, (b) If $a_L(t') \in (0, 1)$ then $a_L(t) < a_L(t') < 1$ and $a_H(t) = 0$. From Step 0, we have $a_U(t') > a_U(t)$. Thus,

$$\nu a_H(t') + (1 - \nu) a_L(t') > \nu a_H(t) + (1 - \nu) a_L(t). \tag{1}$$

To prove claim (a), suppose that $a_H(t) \in (0, 1)$. By Step 1 (a), we have $a_L(t) = 1$. Hence,

(1) implies

$$\nu a_H(t') + (1-\nu)a_L(t') > \nu a_H(t) + (1-\nu)$$
$$\nu(a_H(t') - a_H(t)) > (1-\nu)(1-a_L(t')) \geq 0,$$

and hence, $a_H(t') > a_H(t) > 0$. From Step 2, we have $a_H(t') < 1$ and $a_L(t') = 0$ (Case A).

To prove claim (b), suppose that $a_L(t') \in (0,1)$. By Step 1 (b) we have $a_H(t') = 0$. Hence, (1) implies

$$(1-\nu)a_L(t') > \nu a_H(t) + (1-\nu)a_L(t)$$
$$(1-\nu)(a_L(t') - a_L(t)) > \nu a_H(t) \geq 0$$

and hence, $a_L(t) < a_L(t') < 1$. From Step 2, we must have $a_L(t) > 0$ and $a_H(t) = 0$.

Step 4. We show that there exists some $t < \bar{t}_U$ such that $a_H(t) \in (0,1)$. Suppose not. From Step 2, we have that $a_H(t)$ for all $t < \bar{t}_U$. Thus, for all such t, we have $a_U(t) = (1-\nu)a_L(t) \leq (1-\nu) < 1$. By implication $\lim_{t\to\bar{t}_U} a_U(t) \leq 1-\nu < 1$, contradicting Part (ii).

Let $\widetilde{t}_U \equiv \inf\{t : a_H(t) > 0\}$.

Proof of (iii). From Step 4, we know that $\widetilde{t}_U < \bar{t}_U$. Thus, for any $t \in (\widetilde{t}_U, \bar{t}_U)$, there exists $t' = t - \epsilon$ such that $a_H(t') > 0$. Applying Step 3, we know that for any $t \in (\widetilde{t}_U, \bar{t}_U)$ we have $a_H(t) \in (0,1)$ and $a_L(t) = 1$. Thus, for such t, we have $a_U(t) = \nu a_H(t) + (1-\nu)$. Because $a_U(\cdot)$ is continuous, increasing, and differentiable on $[0, \bar{t}_U)$ and $\widetilde{t}_U < \bar{t}_U$, we have that $a_H(\cdot)$ is continuous, increasing, and differentiable for such t. Finally, from $\lim_{t\to\bar{t}_U} a_U(t) = 1$, we have $\lim_{t\to\bar{t}_U} a_H(t) = 1$.

Step 5. Suppose $\widetilde{t}_U > 0$. We show that $a_L(t) \in (0,1)$ and $a_H(t) = 0$ for $t < \widetilde{t}_U$. From the definition of $\widetilde{t}_U$, we have $a_H(t) = 0$ for $t < \widetilde{t}_U$. From Step 2, for all such t, we have either $a_L(t) \in (0,1)$ for $a_L(t) = 1$. Consider $t, t' < \widetilde{t}_U$ with $t' > t$. From Step 0, we have $a_U(t') > a(t)$, and hence, $a_L(t') > a_L(t)$. By implication $a_L(t) < 1$. Hence, $a_L(t) \in (0,1)$.

Step 6. Suppose $\widetilde{t}_U > 0$. We show that $a_L(t)$ is continuous, increasing, differentiable for $t < \widetilde{t}_U$. From Step 5, we have $a_H(t) = 0$ for $t < \widetilde{t}_U$. Therefore $a_U(t) = (1-\nu)a_L(t)$. Because $a_U(\cdot)$ is continuous, increasing, and differentiable, the conclusion follows.

Step 7. Suppose $\widetilde{t}_U > 0$. We show that $\lim_{t\to\widetilde{t}_U} a_L(t) = 1$ and $\lim_{t\to\widetilde{t}_U} a_H(t) = 0$. Let $t^- \equiv \widetilde{t}_U - \epsilon$ and $t + - \equiv \widetilde{t}_U + \epsilon$ for $\epsilon > 0$. Note that for $t > \widetilde{t}_U$ we have $a_L(t) = 1$. It is therefore obvious that $\lim_{\epsilon\to 0} a_L(t^+) = 1$. Similarly, for $t < \widetilde{t}_U$ we have $a_H(t) = 0$. Therefore it is obvious that $\lim_{\epsilon\to 0} a_H(t^-) = 0$. What remains to establish is

$$\lim_{\epsilon\to 0} a_L(t^-) = 1 \qquad \lim_{\epsilon\to 0} a_H(t^+) = 0.$$

Note that continuity of $a_U(\cdot)$ at $\widetilde{t}_U$ implies

$$\lim_{\epsilon\to 0}[a_U(t^-) - a_U(t^+)] = 0.$$

Substituting $a_L(t^+) = 1$ and $a_H(t^-) = 0$ we have,

$$\lim_{\epsilon\to 0}[(1-\nu)a_L(t^-) - \nu a_H(t^+) - (1-\nu)] = 0 \Rightarrow \lim_{\epsilon\to 0}[(1-\nu)(a_L(t^-) - 1) - \nu a_H(t^+)] = 0.$$

Because $a_L(\cdot) \le 1$ and $a_H(\cdot) \ge 0$ the result follows.

Proof of (iv). Follows from Steps 5-7. □

Proof of Proposition 1.1. We construct a one phase equilibrium, consistent with Lemma 1.1, showing that such an equilibrium exists if and only if $\nu > (1 - \frac{\phi}{\widehat{\phi}})(1 - \exp(-\rho \bar{t}_H))$, and that the only such equilibrium is the one characterized in the statement of the proposition. To this end, consider the one phase structure, characterized by Lemma 1.1 with $\widetilde{t}_U = 0$.

Strategies. If an equilibrium with the one phase structure exists, then for all $t \in [0, \bar{t}_U)$ we have $a_H(t) \in (0,1)$. By implication,

$$g(t) = \theta_H \Rightarrow \mu(t) = \lambda \frac{1-\theta_H}{\theta_H} \Rightarrow F(t) = \frac{1}{\sigma}(1 - \exp(-\lambda \frac{1-\theta_H}{\theta_H} t)),$$

where the last step uses point (i) of Lemma 1.1 to rule out a mass point in the agent's mixed strategy, thereby identifying an integration constant. Using point (i) of Lemma 1.1, we have $F(\bar{t}_U) = 1$, implying $\bar{t}_U = \bar{t}_H$ as stated in the proposition.

For the acceptance probability, note that the agent's expected payoff of waiting to cheat until time t is

$$u(t) = \int_0^t \exp(-(\lambda+\rho)s)(\nu a_H(s) + (1-\nu))ds + \exp(-(\lambda+\rho)t)(\nu a_H(t) + (1-\nu) - \phi),$$

where we have used Lemma 1.1 (iii) to establish $a_H(t) \in (0,1)$ and $a_L(t) = 1$. Furthermore, since $a_H(\cdot)$ is differentiable for $t < \bar{t}_U$. It follows that

$$u'(t) = 0 \iff \nu a_H'(t) - \rho(\nu a_H(t) + (1-\nu)) + \phi(\rho + \lambda) = 0.$$

Solving, we have

$$a_H(t) = \frac{1}{\nu}\Big(\frac{\phi}{\widehat{\phi}} - (1-\nu)\Big) + \kappa \exp\{\rho t\},$$

where κ is an integration constant. Using the boundary condition $a_H(\bar{t}_H) = 1$ we find

$$\begin{aligned} a_H(t) &= \frac{1}{\nu}(\frac{\phi}{\widehat{\phi}} - (1-\nu)) + \Big(1 - \frac{1}{\nu}(\frac{\phi}{\widehat{\phi}} - (1-\nu))\Big) \exp\{-\rho(\bar{t}_H - t)\} \\ &= \frac{1}{\nu}(\frac{\phi}{\widehat{\phi}} + \Big(1 - \frac{\phi}{\widehat{\phi}}\Big) \exp\{-\rho(\bar{t}_H - t)\} - (1-\nu)). \end{aligned}$$

Therefore, such an equilibrium exists provided two additional conditions,

$$a'_H(t) > 0 \iff \frac{1}{\nu}(\frac{\phi}{\widehat{\phi}} - (1-\nu)) < 1 \iff \phi < \widehat{\phi}$$

and

$$a_H(0) \geq 0 \iff \nu \geq (1 - \frac{\phi}{\widehat{\phi}})(1 - \exp\{-\rho \bar{t}_H\}).$$

The first of these is Assumption 1, the second is stated in the proposition.

Payoffs. The strategic agent's payoff is equal to his payoff of submitting at time 0, which is

$$a_U(0) - \phi = \nu a_H(t) + (1-\nu) - \phi = \frac{\phi}{\widehat{\phi}} + (1 - \frac{\phi}{\widehat{\phi}})\exp(-\rho \bar{t}_H) - \phi,$$

which is identical to the strategic agent's payoff in the main model, when facing a principal with known standard θ_H. The high type principal gets payoff zero from any arrival inside the phase of doubt, and a payoff of one if the phase of credibility is reached. Thus, the high type principal's payoff is

$$(1-\theta_H)(1-\sigma)\int_{\bar{t}_H}^{\infty} \lambda \exp(-(\lambda+\rho)s)ds,$$

exactly as in the main model where the principal's payoff is known to be θ_H. The low type principal's payoff, in this equilibrium is

$$\int_0^{\bar{t}_H} \exp(-(\rho+\lambda)t)(\lambda(1-\sigma F(t))(1-\theta_L) - \sigma f(t)\theta_L)dt + (1-\theta_L)(1-\sigma)\int_{\bar{t}_H}^{\infty} \lambda \exp(-(\rho+\lambda)t)dt.$$

In this equilibrium, $1 - \sigma F(t) = \exp(-r_H t)$ and $\sigma f(t) = r_H \exp(-r_H t)$, where $r_H \equiv \lambda \frac{1-\theta_H}{\theta_H}$. Therefore, the low type principal's payoff is

$$V_L = \int_0^{\bar{t}_H} \exp(-(\rho+\lambda+r_H)t)(\lambda(1-\theta_L) - r_H\theta)dt + (1-\theta_L)(1-\sigma)\int_{\bar{t}_H}^{\infty} \lambda \exp(-(\rho+\lambda)t)dt.$$

Substituting for r_H, we have

$$\begin{aligned} V_L &= \int_0^{\bar{t}_H} \exp(-(\rho + \frac{\lambda}{\theta_H})t)(\lambda(1-\theta_L) - \theta_L \lambda \frac{1-\theta_H}{\theta_H})\,\mathrm{d}t + (1-\sigma)(1-\theta_L)\int_{\bar{t}_H}^{\infty} \lambda \exp(-(\rho+\lambda)t)\,\mathrm{d}t \\ &= \lambda(1 - \frac{\theta_L}{\theta_H})\int_0^{\bar{t}_H} \exp(-(\rho + \frac{\lambda}{\theta_H})t)\,\mathrm{d}t + (1-\sigma)(1-\theta_L)\int_{\bar{t}_H}^{\infty} \lambda \exp(-(\rho+\lambda)t)\,\mathrm{d}t. \end{aligned}$$

Normative Analysis. Point (i) is obvious, since the high type principal's payoff is the same as under transparency, where she is known to be the high type.

(ii) We seek to show that this is larger than the payoff in the main model, V, given in Proposition ??. Note first that for $\theta_H = \theta_L$, the two expressions are equal, i.e. $V_L = V$. Differentiating with respect to θ_H, we have

$$\frac{dV_L}{d\theta_H} = \frac{\lambda\theta_L}{\theta_H^2}\int_0^{\bar{t}_H} \exp(-(\rho+\frac{\lambda}{\theta_H})t)\,\mathrm{d}t$$
$$+\lambda(1-\frac{\theta_L}{\theta_H})[\exp(-(\rho+\frac{\lambda}{\theta_H})\bar{t}_H)\frac{d\bar{t}_H}{d\theta_H}+\int_0^{\bar{t}_H}\exp(-(\rho+\frac{\lambda}{\theta_H})t)\frac{\lambda}{\theta_H^2}tdt]-(1-\sigma)(1-\theta_L)\lambda\exp(-(\rho+\lambda)\bar{t}_H)\frac{d\bar{t}_H}{d\theta_H}.$$

Note that

$$(1-\sigma)\exp(-(\rho+\lambda)\bar{t}_H) = \exp(\{1+(\rho+\lambda)\frac{\theta_H}{\lambda(1-\theta_H)}\}\ln(1-\sigma)) = \exp(-(\rho+\frac{\lambda}{\theta_H})\bar{t}_H),$$

and

$$\frac{d\bar{t}_H}{d\theta_H} = -\frac{\lambda(1-\theta_H)+\lambda\theta_H}{\lambda^2(1-\theta_H)^2}\ln(1-\sigma) = -\frac{1}{\lambda(1-\theta_H)^2}\ln(1-\sigma) = \frac{\bar{t}_H}{\theta_H(1-\theta_H)}.$$

Substituting and simplifying, we have

$$\frac{dV_L}{d\theta_H} = \frac{\lambda\theta_L}{\theta_H^2}\int_0^{\bar{t}_H}\exp(-(\rho+\frac{\lambda}{\theta_H})t)\,\mathrm{d}t - \frac{\lambda\theta_L}{\theta_H^2}\exp(-(\rho+\frac{\lambda}{\theta_H})\bar{t}_H)\bar{t}_H + \lambda(\frac{\theta_H-\theta_L}{\theta_H})\int_0^{\bar{t}_H}\exp(-(\rho+\frac{\lambda}{\theta_H})t)\frac{\lambda}{\theta_H^2}tdt.$$

Noting that the last integral is strictly positive, we have

$$\frac{dV_L}{d\theta_H} > \frac{\lambda\theta}{\theta_H^2}[\int_0^{\bar{t}_H}\exp(-(\rho+\frac{\lambda}{\theta_H})t)\,\mathrm{d}t - \exp(-(\rho+\frac{\lambda}{\theta_H})\bar{t}_H)\bar{t}_H] > 0,$$

where the last inequality follows because $\exp(-(\rho+\lambda/\theta_H)t)$ is a decreasing function, and thus, its average value on interval $[0,\bar{t}_H]$ is larger than its value at the right endpoint. Because (i) V_L is increasing in θ_H, (ii) $\theta_H = \theta \Rightarrow V_L = V$, and (iii) $\theta_H > \theta_L$, we have that in the one stage auditing equilibrium $V_L > V$. □

Proof of Proposition 1.2. We construct a two stage equilibrium, consistent with Lemma 1.1, showing that such an equilibrium exists if and only if $\nu < (1-\frac{\underline{\phi}}{\bar{\phi}})(1-\exp(-\rho\bar{t}_H))$, and that the only such equilibrium is the one characterized in the statement of the proposition. To this end, consider the two phase structure, characterized by Lemma 1.1 with $\widetilde{t}_U > 0$.

Strategies and Phase Transitions. From Lemma 1.1, in phase 1, we have $a_L(t) \in (0,1)$, and hence

$$\mu(t) = \mu_L \Rightarrow F(t) = \frac{1}{\sigma}(1-\exp\{-\mu_L t\}),$$

where use has been made of the fact that $F(0) = 0$ (i.e., $F(\cdot)$ has no mass point) which allows us to determine that the integration constant in the solution is 1.

For the acceptance strategy in phase 1, we substitute $a_U(t) = (1-\nu)a_L(t)$ into the agent's indifference condition to obtain

$$(1-\nu)a_L'(t) - \rho(1-\nu)a_L(t) + \phi(\rho+\lambda) = 0.$$

Solving, with boundary condition $a_L(\widetilde{t}_U) = 1$, we have

$$\begin{aligned} a_L(t) &= (\frac{1}{1-\nu})\frac{\phi}{\widehat{\phi}} + (1 - (\frac{1}{1-\nu})\frac{\phi}{\widehat{\phi}})\exp\{-\rho(\widetilde{t}_U - t)\} \\ &= (\frac{1}{1-\nu})[\frac{\phi}{\widehat{\phi}} + (1 - \frac{\phi}{\widehat{\phi}} - \nu)\exp\{-\rho(\widetilde{t}_U - t)\}]. \end{aligned}$$

From Lemma 1.1, in phase 2, we have $a_H(t) \in (0,1)$, and hence,

$$\mu(t) = \mu_H \Rightarrow F(t) = \frac{1}{\sigma}(1 - \kappa_2 \exp\{-\mu_H t\}),$$

where κ_2 is an integration constant. From the boundary condition $F(\bar{t}_U) = 1$ we have

$$\frac{1}{\sigma}(1 - \kappa_2\exp\{-\mu_H \bar{t}_U\}) = 1 \Rightarrow \kappa_2 = (1-\sigma)\exp\{\mu_H \bar{t}_U\}.$$

Note that $\kappa_2 > 0$. Thus, $F(\cdot)$ is increasing in the second phase.

The differential equation for the agent's indifference condition in phase 2 is identical to the differential equation for the indifference condition in the one phase equilibrium, characterized in Proposition 1.1. Following the same argument with boundary condition $a_H(\bar{t}_U) = 1$, we have

$$a_H(t) = \frac{1}{\nu}\Big(\frac{\phi}{\widehat{\phi}} + (1 - \frac{\phi}{\widehat{\phi}})\exp\{-\rho(\bar{t}_U - t)\} - (1-\nu)\Big).$$

To determine the phase transitions, $\widetilde{t}_U, \bar{t}_U$, we use continuity of $F(\cdot)$ at $\widetilde{t}_U$, and the boundary condition $a_H(\widetilde{t}_U) = 0$, both of which come from Lemma 1.1.

$$\begin{aligned} \frac{1}{\sigma}(1 - \exp\{-\mu_L \widetilde{t}_U\}) &= \frac{1}{\sigma}(1 - (1-\sigma)\exp\{\mu_H(\bar{t}_U - \widetilde{t}_U)\}) \\ \frac{1}{\nu}\Big(\frac{\phi}{\widehat{\phi}} + (1 - \frac{\phi}{\widehat{\phi}})\exp\{-\rho(\bar{t}_U - \widetilde{t}_U)\} - (1-\nu)\Big) &= 0 \end{aligned}$$

Solving, we have

$$\widetilde{t}_U = \bar{t}_L - \frac{\mu_H}{\mu_L}\delta_U \qquad \bar{t}_U = \bar{t}_L + (1 - \frac{\mu_H}{\mu_L})\delta_U,$$

where $\delta_U = -\ln(1 - \frac{\nu\widehat{\phi}}{\widehat{\phi}-\phi})/\rho$. Note that for this system to have any solution, we must have $\nu < 1 - \phi/\widehat{\phi}$, so that δ_U is well-defined. Additional details are available in the Online Supplement.

Thus, we have a unique candidate for the two stage equilibrium. This candidate is indeed an equilibrium if and only if $\widetilde{t}_U > 0$.

Claim 1. We show that $\widetilde{t}_U > 0 \iff \nu < (1 - \frac{\phi}{\widehat{\phi}})(1 - \exp\{-\rho \bar{t}_H\})$. Note that

$$\widetilde{t}_U > 0 \iff \bar{t}_L > \frac{\mu_H}{\mu_L}\delta_U \iff \bar{t}_H > \delta_U \iff -\rho\bar{t}_H < \ln(1 - \frac{\nu\widehat{\phi}}{\widehat{\phi} - \phi}) \iff$$
$$\exp\{-\rho\bar{t}_H\} < 1 - \frac{\widehat{\phi}\nu}{\widehat{\phi} - \phi} \iff \nu < (1 - \frac{\phi}{\widehat{\phi}})(1 - \exp\{-\rho\bar{t}_H\}).$$

It follows that the two stage equilibrium exists when $\nu < (1-\frac{\phi}{\widehat{\phi}})(1-\exp\{-\rho\bar{t}_H\})$, and in this equilibrium the strategies and phase transitions are the ones given in the proposition. One additional claim is made in the phase transitions part of the proposition, which we verify.

Claim 2. We show that if $\nu < (1 - \frac{\phi}{\widehat{\phi}})(1 - \exp\{-\rho\bar{t}_H\})$, then $\bar{t}_U < \bar{t}_H$. From claim 1, for such ν we have

$$\widetilde{t}_U > 0 \Rightarrow \bar{t}_L > \frac{\mu_H}{\mu_L}\delta_U \Rightarrow \bar{t}_H > \delta_U \Rightarrow \bar{t}_H(1 - \frac{\mu_H}{\mu_L}) > (1 - \frac{\mu_H}{\mu_L})\delta_U \Rightarrow$$
$$-\ln(1-\sigma)\frac{\mu_L - \mu_H}{\mu_H\mu_L} > (1 - \frac{\mu_H}{\mu_L})\delta_U \Rightarrow -\ln(1-\sigma)(\frac{1}{\mu_H} - \frac{1}{\mu_L}) > (1 - \frac{\mu_H}{\mu_L})\delta_U \Rightarrow$$
$$\bar{t}_H - \bar{t}_L > (1 - \frac{\mu_H}{\mu_L})\delta_U \Rightarrow \bar{t}_H > \bar{t}_L + (1 - \frac{\mu_H}{\mu_L})\delta_U = \bar{t}_U$$

Claim 3. We show that $\kappa_2 = (1-\sigma)\exp\{\mu_H\bar{t}_U\} = \exp\{-(\mu_L - \mu_H)\widetilde{t}_U\}$. Consequently, the agent's strategy can be written as stated in the proposition. Note that

$$\mu_H\bar{t}_U + \ln(1-\sigma) = \mu_H(-\frac{\ln(1-\sigma)}{\mu_L} + (1 - \frac{\mu_H}{\mu_L})\delta_U) + \ln(1-\sigma) =$$
$$(1 - \frac{\mu_H}{\mu_L})\ln(1-\sigma) + \mu_H(1 - \frac{\mu_H}{\mu_L})\delta_U =$$
$$(\mu_L - \mu_H)(\frac{\ln(1-\sigma)}{\mu_L} + \frac{\mu_H}{\mu_L}\delta_U) = -(\mu_L - \mu_H)\widetilde{t}_U.$$

The claim follows by applying $\exp(\cdot)$ to both sides of the equation.

Beliefs. Follows immediately from Lemma 1.1 and the principal's sequentially rational acceptance decision.

Payoffs. The strategic agent is indifferent about submitting a fake at all times inside the phase of doubt, and thus, the agent's payoff is the payoff of submitting a fake at time zero. Thus, the strategic agent's payoff is

$$a_U(0) - \phi = (1-\nu)a_L(0) - \phi = \frac{\phi}{\widehat{\phi}} + (1 - \frac{\phi}{\widehat{\phi}} - \nu)\exp(-\rho\widetilde{t}_U) - \phi.$$

The high type principal either rejects or mixes at all times $t \in [0, \bar{t}_U)$. Thus, the high type principal's payoff is

$$V_H = (1-\theta_H)(1-\sigma)\int_{\bar{t}_U}^{\infty} \lambda \exp\{-(\rho+\lambda)t\}dt.$$

To calculate the low type principal's payoff, note that the low type mixes in phase 1, and accepts in phase 2. Thus, the low type's payoff is

$$V_L = \int_{\widetilde{t}_U}^{\bar{t}_U} \exp(-(\rho+\lambda)t)(\lambda(1-\sigma F(t))(1-\theta_L)-\sigma f(t)\theta_L)dt+(1-\theta_L)(1-\sigma)\int_{\bar{t}_U}^{\infty} \lambda \exp(-(\rho+\lambda)t)dt$$

where $F(\cdot)$ and $f(\cdot)$ are the agent's CDF and PDF of the agent's mixed strategy. From the equilibrium characterization, we have $(1-\sigma F(t)) = (1-\sigma)\exp(\mu_H(\bar{t}_U - t))$ and $\sigma f(t) = (1-\sigma)\mu_H \exp(\mu_H(\bar{t}_U - t))$, which implies that the low type principal's payoff is

$$(1-\sigma)\exp(\mu_H \bar{t}_U)\int_{\widetilde{t}_U}^{\bar{t}_U} \exp(-(\rho+\frac{\lambda}{\theta_H})t)(\lambda(1-\theta_L)-\mu_H\theta_L)dt + (1-\theta_L)(1-\sigma)\int_{\bar{t}_U}^{\infty} \lambda \exp(-(\rho+\lambda)t)dt =$$

$$(1-\sigma)\exp(\mu_H \bar{t}_U)(1-\frac{\theta_L}{\theta_H})\int_{\widetilde{t}_U}^{\bar{t}_U} \lambda \exp(-(\rho+\frac{\lambda}{\theta_H})t)dt + (1-\theta_L)(1-\sigma)\int_{\bar{t}_U}^{\infty} \lambda \exp(-(\rho+\lambda)t)dt$$

Using Claim 3 to simplify the leading term, we have

$$\exp(-(\mu_L-\mu_H)\widetilde{t}_U)(1-\frac{\theta_L}{\theta_H})\int_{\widetilde{t}_U}^{\bar{t}_U} \exp(-(\rho+\frac{\lambda}{\theta_H})t)dt + (1-\theta_L)(1-\sigma)\int_{\bar{t}_U}^{\infty} \lambda \exp(-(\rho+\lambda)t)dt.$$

Normative Comparison. That the high type principal strictly benefits from opaque standards in this equilibrium follows immediately from $\bar{t}_U < \bar{t}_H$, proved in Claim 2.

We prove that the low type principal's payoff is higher in the two stage equilibrium than in the baseline model in three steps.

In Step 1 we show that there exists $\widetilde{\sigma}$ such that the two stage equilibrium obtains iff $\sigma > \widetilde{\sigma}$ and that the low type principal's payoff in the two stage equilibrium approaches her payoff in the 1-phase equilibrium as $\sigma \downarrow \widetilde{\sigma}$. Because we showed above that the principal's payoff is strictly higher in the one stage opaque standards equilibrium than in the baseline model, we conclude that there exists $\epsilon > 0$ such that her payoff in the two stage auditing equilibrium is also higher than in the baseline model for all $\sigma \in (\widetilde{\sigma}, \widetilde{\sigma}+\epsilon)$.

In Step 2 we show that if the low type principal's payoff is higher in the 2-stage equilibrium than in the baseline model for any value of σ, then it is higher for all larger values as well.

In Step 3, we combine Steps 1 and 2 to show that the principal's payoff is higher in the two stage equilibrium than in the baseline model.

Step 1: We show that for any σ such that the two stage equilibrium exists, there exists

$\sigma' < \sigma$ such that (i) the two stage equilibrium exists at σ', and (ii) at σ' the principal's payoff in the two stage equilibrium is higher than her payoff in the baseline model.

Consider parameters at which the two stage equilibrium exists; by Proposition 1.2, we have $\nu < \nu^*$. Note that

$$\nu < \nu^* = (1 - \frac{\phi}{\widehat{\phi}})(1 - \exp(-\rho \bar{t}_H)) \iff \nu < 1 - \frac{\phi}{\widehat{\phi}} \quad \text{and} \quad \bar{t}_H \text{ is sufficiently large.}$$

Recalling that $\bar{t}_H$ is monotone increasing in σ, we have that the two stage equilibrium exists whenever $\nu < 1 - \frac{\phi}{\widehat{\phi}}$ and $\sigma > \widetilde{\sigma}$, for some $\widetilde{\sigma} \in (0, 1)$. By implication, if $\sigma > \widetilde{\sigma}$, then the two stage equilibrium exists for all $\sigma \in (\widetilde{\sigma}, \sigma)$.

Next, we argue that as $\sigma \downarrow \widetilde{\sigma}$, the low type principal's payoff in the two stage equilibrium approaches her payoff in the one stage equilibrium. Note that in the two stage equilibrium, the low type principal's payoff is

$$V_L = \exp(-(\mu_L - \mu_H)\widetilde{t}_U)(1 - \frac{\theta_L}{\theta_H}) \int_{\widetilde{t}_U}^{\bar{t}_U} \exp(-(\rho + \frac{\lambda}{\theta_H})t)dt + (1 - \theta_L)(1 - \sigma) \int_{\bar{t}_U}^{\infty} \lambda \exp(-(\rho + \lambda)t)dt.$$

A straightforward modification of Claim 1 shows that as $\sigma \downarrow \widetilde{\sigma}$, we have $\widetilde{t}_U \downarrow 0$. By implication, $\bar{t}_L \to \frac{\mu_H}{\mu_L}\delta_U$, and hence $\bar{t}_H \to \delta_U$. Furthermore $\bar{t}_U = \bar{t}_L + (1 - \frac{\mu_H}{\mu_L})\delta_U$. Substituting, we have $\bar{t}_U \to \delta_U$. Combining, we have $\bar{t}_U \to \bar{t}_H$. By routine simplification, as $\sigma \downarrow \widetilde{\sigma}$, the low type principal's payoff in the two stage equilibrium approaches

$$(1 - \frac{\theta_L}{\theta_H}) \int_0^{\bar{t}_H} \lambda \exp(-(\rho + \frac{\lambda}{\theta_H})t)dt + (1 - \theta_L)(1 - \widetilde{\sigma}) \int_{\bar{t}_H}^{\infty} \lambda \exp(-(\rho + \lambda)t)dt,$$

which is the low type principal's payoff in the one phase equilibrium that obtains at $\widetilde{\sigma}$. From Proposition 1.1, this payoff strictly exceeds the low type principal's payoff in the baseline model. By implication, there exists $\epsilon > 0$ such that for an $\sigma' \in (\widetilde{\sigma}, \widetilde{\sigma} + \epsilon)$, the low type principal's payoff in the two stage equilibrium at σ' strictly exceeds her payoff in the baseline model. Hence, for any $\sigma > \widetilde{\sigma}$, there exists $\sigma' \in (\widetilde{\sigma}, \sigma)$ at which the low type principal's payoff in the two stage equilibrium is higher than her payoff in the baseline model.

Step 2: Consider $\sigma > \widetilde{\sigma}$. We show that if the principal's payoff is higher in the two stage equilibrium with auditing than in the baseline model at σ, then the same is true for all $\sigma'' > \sigma$.

Consider the payoff difference between the two stage equilibrium and the baseline model,

$$\exp(-(\mu_L - \mu_H)\widetilde{t}_U)(1 - \frac{\theta_L}{\theta_H}) \int_{\widetilde{t}_U}^{\bar{t}_U} \exp(-(\rho + \frac{\lambda}{\theta_H})t)\,\mathrm{d}t$$
$$+ (1 - \theta_L)(1 - \sigma) \left[\int_{\bar{t}_U}^{\infty} \lambda \exp(-(\rho + \lambda)t)\,\mathrm{d}t - \int_{\bar{t}_L}^{\infty} \lambda \exp(-(\rho + \lambda)t)\,\mathrm{d}t \right].$$

We simplify the previous expression in order to isolate σ. To keep the exposition organized, we proceed line-by-line.

We simplify the first line.

$$\exp(-(\mu_L - \mu_H)\widetilde{t}_U)(1 - \frac{\theta_L}{\theta_H})[\exp(-(\rho + \frac{\lambda}{\theta_H})\widetilde{t}_U) - \exp(-(\rho + \frac{\lambda}{\theta_H})\bar{t}_U)].$$

Note that

$$-(\mu_L - \mu_H) = -\lambda\frac{\theta_H(1-\theta_L) - \theta_L(1-\theta_H)}{\theta_H\theta_L} = -\lambda\frac{\theta_H - \theta_L}{\theta_H\theta_L} = \frac{\lambda}{\theta_H} - \frac{\lambda}{\theta_L}.$$

Substituting, and using $\bar{t}_U = \widetilde{t}_U + \delta_U$, we have

$$\exp(-(\rho + \frac{\lambda}{\theta_L})\widetilde{t}_U)(1 - \frac{\theta_L}{\theta_H})[1 - \exp(-(\rho + \frac{\lambda}{\theta_H})\delta_U)].$$

Using $\widetilde{t}_U = \bar{t}_L - \frac{\mu_H}{\mu_L}\delta_U$, we have

$$\exp(-(\rho + \frac{\lambda}{\theta_L})\bar{t}_L)\exp((\rho + \frac{\lambda}{\theta_L})\frac{\mu_H}{\mu_L}\delta_U)(1 - \frac{\theta_L}{\theta_H})[1 - \exp(-(\rho + \frac{\lambda}{\theta_H})\delta_U)].$$

Substituting the definition of $\bar{t}_L$, the first line is

$$(1-\sigma)^{(\rho + \frac{\lambda}{\theta_L})\frac{\theta_L}{\lambda(1-\theta_L)}}\exp((\rho + \frac{\lambda}{\theta_L})\frac{\mu_H}{\mu_L}\delta_U)(1 - \frac{\theta_L}{\theta_H})[1 - \exp(-(\rho + \frac{\lambda}{\theta_H})\delta_U)] =$$
$$(1-\sigma)^{\frac{\rho\theta_L + \lambda}{\lambda(1-\theta_L)}}\exp((\rho + \frac{\lambda}{\theta_L})\frac{\mu_H}{\mu_L}\delta_U)(1 - \frac{\theta}{\theta_H})[1 - \exp(-(\rho + \frac{\lambda}{\theta_H})\delta_U)] =$$
$$\kappa_1(1-\sigma)^{\frac{\rho\theta_L + \lambda}{\lambda(1-\theta_L)}},$$

where $\kappa_1 \equiv \exp((\rho + \frac{\lambda}{\theta_L})\frac{\mu_H}{\mu_L}\delta_U)(1 - \frac{\theta_L}{\theta_H})[1 - \exp(-(\rho + \frac{\lambda}{\theta_H})\delta_U)]$ is independent of σ.

Next, we simplify the second line.

$$(1-\theta_L)(1-\sigma)[\int_{\bar{t}_U}^{\infty}\lambda\exp(-(\rho+\lambda)t)\,\mathrm{d}t - \int_{\bar{t}_L}^{\infty}\lambda\exp(-(\rho+\lambda)t)\,\mathrm{d}t] =$$
$$(1-\theta_L)(1-\sigma)\int_{\bar{t}_U}^{\bar{t}_L}\lambda\exp(-(\rho+\lambda)t)\,\mathrm{d}t = (1-\theta_L)(1-\sigma)\frac{\lambda}{\lambda+\rho}[\exp(-(\rho+\lambda)\bar{t}_U) - \exp(-(\rho+\lambda)\bar{t}_L)].$$

Substituting $\bar{t}_U = \bar{t}_L + (1 - \frac{\mu_H}{\mu_L})\delta_U$, we have

$$(1-\theta_L)(1-\sigma)\frac{\lambda}{\lambda+\rho}\exp(-(\rho+\lambda)\bar{t}_L)[\exp(-(\rho+\lambda)(1 - \frac{\mu_H}{\mu_L})\delta_U) - 1].$$

Note that

$$(1-\sigma)\exp(-(\rho+\lambda)\bar{t}_L) = \exp(1 - (\rho+\lambda)\frac{\theta_L}{\lambda(1-\theta_L)})\ln(1-\sigma)) =$$
$$\exp(-(\rho + \frac{\lambda}{\theta_L})(-\frac{\theta_L}{\lambda(1-\theta_L)})\ln(1-\sigma)) = \exp(-(\rho + \frac{\lambda}{\theta_L})\bar{t}_L).$$

Continuing the simplification,

$$
\begin{aligned}
&(1-\theta_L)\frac{\lambda}{\lambda+\rho}\exp(-(\rho+\frac{\lambda}{\theta_L})\bar{t}_L)[\exp(-(\rho+\lambda)(1-\frac{\mu_H}{\mu_L})\delta_U)-1] \\
&(1-\theta_L)\frac{\lambda}{\lambda+\rho}[\exp(-(\rho+\lambda)(1-\frac{\mu_H}{\mu_L})\delta)-1](1-\sigma)^{\frac{\rho\theta_L+\lambda}{\lambda(1-\theta_L)}} = \\
&\kappa_2(1-\sigma)^{\frac{\rho\theta_L+\lambda}{\lambda(1-\theta_L)}},
\end{aligned}
$$

where $\kappa_2 \equiv (1-\theta_L)\frac{\lambda}{\lambda+\rho}[\exp(-(\rho+\lambda)(1-\frac{\mu_H}{\mu_L})\delta_U)-1]$ is independent of σ. Combining terms, the payoff difference as a function of σ is simply

$$(\kappa_1+\kappa_2)(1-\sigma)^{\frac{\rho\theta_L+\lambda}{\lambda(1-\theta_L)}}.$$

Therefore, the payoff difference is positive if and only if $\kappa_1+\kappa_2>0$. Thus, if the payoff difference is positive for some value of $\sigma>\widetilde{\sigma}$, then it is also positive for $\sigma''\in(\sigma,1]$.

Step 3. We show that the low type principal's payoff is higher in the two stage equilibrium than in the baseline model. Consider $\sigma>\widetilde{\sigma}$. From Step 1, there exists $\sigma'\in(\widetilde{\sigma},\sigma)$ such that the principal's payoff in the two stage equilibrium at σ' is higher than in the baseline model. Applying Step 2, the low type principal's payoff in the two stage auditing equilibrium at $\sigma>\sigma'$ is also higher than in the baseline model. □

A.2 Proofs for Logjam

Proof of Lemma 2.1. Step 1. Let $S\equiv\{t: a(t)<\overline{a}(t)\}$. We show that S is an open set, i.e. S is a countable union of open intervals $S=\cup(t_k,t_{k+1})$.

Consider some t such that $a(t)<\overline{a}(t)$, i.e. the constraint is slack. By implication, we must have $g(t)=\theta$, and hence $f(t)>0$.

If there exists a small $\delta>0$ such that $a(t')<\overline{a}(t')$ for all $t'\in(t-\delta,t+\delta)$, then the set S is open, and the claim is established. To derive a contradiction, suppose that for all $\delta>0$, there exists some $t'\in[t-\delta,t+\delta]$ such that $a(t')=\overline{a}(t')$. Let $\overline{\epsilon}\equiv\exp(-(\rho+\lambda)t)(\overline{a}(t)-a(t))>0$ and let $\epsilon\equiv\overline{\epsilon}/4$. From $a(\cdot)\in[\phi,1)$ and continuity of exponential, there exists some $\delta'>0$ such that the following conditions hold for $t'\in(t-\delta',t+\delta')$,

$$
\begin{aligned}
-\epsilon\le\int_{t'}^{t}\lambda\exp(-(\rho+\lambda)s)a(s)ds\le\epsilon \\
-\epsilon\le(\exp(-(\rho+\lambda)t)-\exp(-(\rho+\lambda)t'))(a(t')-\phi)\le\epsilon \\
-\epsilon\le\exp(-(\rho+\lambda)t)(\overline{a}(t)-\overline{a}(t'))\le\epsilon.
\end{aligned}
$$

Furthermore, by assumption, there exists some $t' \in [t-\delta', t+\delta']$ such that $a(t') = \overline{a}(t')$.

$$u(t) - u(t') = \int_{t'}^{t} \lambda \exp(-(\rho+\lambda)s)a(s)ds + (\exp(-(\rho+\lambda)t) - \exp(-(\rho+\lambda)t'))(a(t') - \phi) + \\ \exp(-(\rho+\lambda)t)[\overline{a}(t) - \overline{a}(t') - (\overline{a}(t) - a(t)].$$

It follows that $u(t) - u(t') \leq 3\epsilon - \bar{\epsilon} = -\frac{\bar{\epsilon}}{4}$, which contradicts $f(t) > 0$.

Step 2. We show that $a(\cdot)$ is continuous at all $t \geq 0$.

First (1), consider $t \in S$, i.e. $a(t) < \overline{a}(t)$. From Step A we have that $(t-\delta, t+\delta) \in S$ for some $\delta > 0$, and hence, $f(t') > 0$ for all $t' \in (t-\delta, t+\delta)$. By implication, $u(t) - u(t') = 0$ for all $t' \in (t-\delta, t+\delta)$. It follows that for all such t',

$$\int_{t'}^{t} \lambda \exp(-(\rho+\lambda)s)a(s)ds + (\exp(-(\rho+\lambda)t) - \exp(-(\rho+\lambda)t')(a(t) - \phi) + \\ \exp(-(\rho+\lambda)t')(a(t) - a(t')) = 0.$$

Taking the limit as $t' \to t$, we find $a(t) - \lim_{t' \to t} a(t') = 0$, establishing continuity at t.

Second (2), consider t such that $a(t) = \overline{a}(t)$, and assume that t is in the interior of S^C, i.e. there exists a small δ such that for all $t' \in (t-\delta, t+\delta)$, we have $a(t') = \overline{a}(t')$. In this case continuity of $a(\cdot)$ at t follows immediately from continuity of $\overline{a}(\cdot)$ on interval $(t-\delta, t+\delta)$.

Third (3), consider t such that $a(t) = \overline{a}(t)$, and assume that t is on the boundary of S, so that in any interval $(t-\delta, t+\delta)$ there exists some t'_δ such that $a(t'_\delta) < \overline{a}(t'_\delta)$ and some t''_δ such that $a(t''_\delta) = \overline{a}(t''_\delta)$. We will show that for a sufficiently small δ, we have $|\overline{a}(t) - a(t')| < \epsilon$, whenever $t' \in (t-\delta, t+\delta)$, regardless of whether $t' \in S$ or $t' \in S^C$.

(i) Consider $t' \in S^C$. From continuity of $\overline{a}(\cdot)$ at t, we know that for any $\epsilon > 0$ there exists δ_C such that $|\overline{a}(t) - \overline{a}(t')| < \epsilon$ for any $t' \in (t-\delta_C, t+\delta_C)$.

(ii) Consider $t' \in S$, i.e. $a(t') < \overline{a}(t')$. We have

$$u(t') - u(t) = \int_{t}^{t'} \lambda \exp(-(\rho+\lambda s))a(s)ds + (\exp(-(\rho+\lambda)t') - \exp(-(\rho+\lambda)t))(a(t') - \phi) \\ + \exp(-(\rho+\lambda)t)(a(t') - \overline{a}(t)).$$

Because $a(\cdot) \in (\phi, 1)$ and the exponential is continuous and $\overline{a}(\cdot)$ is continuous, some $\delta_S > 0$ exists such that the following inequalities hold for all $t' \in (t-\delta_S, t+\delta_S)$:

$$\int_{t}^{t'} \lambda \exp(-(\rho+\lambda s))a(s)ds \leq \frac{\epsilon}{2}\exp(-(\rho+\lambda)t),$$
$$(\exp(-(\rho+\lambda)t') - \exp(-(\rho+\lambda)t))(a(t') - \phi) \leq \frac{\epsilon}{2}\exp(-(\rho+\lambda)t),$$

Recall that $t' \in S$, and hence $f(t') > 0$. For such t', we have

$$0 \leq u(t') - u(t) \leq \exp(-(\rho+\lambda)t)(\epsilon + a(t') - \overline{a}(t)),$$

and hence $\overline{a}(t) - a(t') \leq \epsilon$.

(iii) Next, note that $\overline{a}(t) - a(t') \geq \overline{a}(t) - \overline{a}(t') \geq -\epsilon$ for $t' \in (t - \delta_C, t + \delta_C)$.

Let $\delta^* = \min\{\delta_C, \delta_S\}$, and consider $t' \in (t - \delta^*, t + \delta^*)$. If $t' \in S^C$, then $|\overline{a}(t) - a(t')| = |\overline{a}(t) - \overline{a}(t')| < \epsilon$, from (i). If $t' \in S$, then (ii) implies that $\overline{a}(t) - a(t') \leq \epsilon$ and (iii) implies $\overline{a}(t) - a(t') \geq -\epsilon$, and hence, $|\overline{a}(t) - a(t')| < \epsilon$. It follows that $a(\cdot)$ is continuos at t.

Points (1)-(3) establish continuity of $a(\cdot)$ at all t in the interior of S, the interior of S^C, and the boundary of S, thereby proving the result.

Step 3. We show that $u(\cdot)$ is continuous. For any t, t' we have

$$u(t') - u(t) = \int_t^{t'} \lambda \exp(-(\rho + \lambda s))a(s)ds + (\exp(-(\rho+\lambda)t') - \exp(-(\rho+\lambda)t))(a(t') - \phi) \\ + \exp(-(\rho+\lambda)t)(a(t') - \overline{a}(t)).$$

Taking the limit as $t \to t'$ and using the continuity of $a(\cdot)$ yields the result.

Step 4. Suppose that for $t \in (t_L, t_H)$ we have $a(t) = \overline{a}(t)$, where $t_H < t^*$. We show that $u(\cdot)$ is strictly increasing on (t_L, t_H). After substituting $a(\cdot) = \overline{a}(\cdot)$ to calculate $u(\cdot)$, the result follows from straightforward differentiation.

Definition. Let $u^* \equiv \max_t u(t)$.

Step 5. We show that if $t' < t < t^*$ and $u(t) < u^*$, then $u(t') < u^*$. Consider $X = \{t'' \in [0, t'] : u(t'') = u^*\}$. If X is empty, then the result holds. By way of contradiction, suppose X is nonempty. Set X has an upper bound, and therefore it has a least upper bound, denoted T. Continuity of $u(\cdot)$ (Step 3) implies (1) $u(T) = u^*$, and (2), that $T < t$. It follows that for all $t' \in (T, t)$, we have $u(t') < u^*$ and thus, $f(t') = 0$. Therefore, for all such t', we have $a(t') = \overline{a}(t')$. Using Step 4, we have that $u(\cdot)$ is strictly increasing on (T, t), and hence $u(T) < u(t) < u^*$, resulting in a contradiction.

Step 6. Suppose that for $t \in (t_L, t_H)$ we have $a(t) = \overline{a}(t)$, where $t_L > t^*$. We show that $u(\cdot)$ is strictly decreasing on (t_L, t_H). Similar to Step 4.

Step 7. We show that if $t^* < t < t'$ and $u(t) < u^*$, then $u(t') < u^*$. Similar to Step 5.

Step 8. There exist $\widetilde{t}_J, \overline{t}_J$ with $0 \leq \widetilde{t}_J \leq t^* \leq \overline{t}_J \leq \infty$ such that (1) $u(t) < u^*$ for $t \in [0, \widetilde{t}_J)$, (2) $u(t) = u^*$ for $t \in [\widetilde{t}_J, \overline{t}_J]$, (3) $u(t) < u^*$ for $t \in (\overline{t}_J, \infty)$. Follows from Steps 5 and 7.

Step 9. We show that if $t \in (\widetilde{t}_J, \overline{t}_J)$ then $a(t) < \overline{a}(t)$ and $f(t) > 0$. From Step 8, we know that $u(t) = u^*$ for $t \in (\widetilde{t}_J, \overline{t}_J)$. Following a similar argument to the one in Lemma 4.1 it is possible to show that $a(\cdot)$ is differentiable on this interval, and that

$$a'(t) = \rho(a(t) - \frac{\phi}{\hat{\phi}}) \Rightarrow a(t) = \frac{\phi}{\hat{\phi}} + \kappa \exp(\rho t).$$

Suppose $t' \in (\widetilde{t}_J, \bar{t}_J)$ and $a(t') = \overline{a}(t')$. We consider two cases. (1) If $\kappa \leq 0$, then $a(t)$ is weakly decreasing. Thus, $a(t') = \overline{a}(t')$ implies $a(t'') \geq a(t')$ for $t'' \in (\widetilde{t}_J, t')$. By assumption $a(t') = \overline{a}(t') > \overline{a}(t'')$, where the last inequality follows because $\overline{a}(t)$ is a strictly increasing function. Thus, we have shown $a(t'') > \overline{a}(t'')$, violating the constraint. (2) If $\kappa > 0$, then $a(\cdot)$ is a strictly increasing convex function on $(\widetilde{t}_J, \bar{t}_J)$. Recall that $\overline{a}(\cdot)$ is a strictly increasing concave function. If $a'(t') < \overline{a}'(t')$, then the constraint $a(\cdot) \leq \overline{a}(\cdot)$ is violated in a small interval $(t' - \delta, t')$. Similarly if $a'(t') > \overline{a}'(t')$ then the constraint is violated in a small interval $(t', t' + \delta)$. Finally if $a'(t') = \overline{a}'(t')$, then $a(\cdot)$ and $\overline{a}(\cdot)$, can be separated by a tangent line at t', and hence $a(\cdot) > \overline{a}(\cdot)$ around t', violating the constraint. Thus, we have shown that $t' \in (\widetilde{t}_J, \bar{t}_J) \Rightarrow a(t') < \overline{a}(t')$. By implication $g(t') = \theta$, and hence $f(t') > 0$.

Step 10. We show (1) if $\widetilde{t}_J > 0$, then $a(t) = \overline{a}(t)$ for $t \in [0, \widetilde{t}_J]$, and (2) if $\bar{t}_J < \infty$ then $a(t) = \overline{a}(t)$ for $t \in [\bar{t}_J, \infty)$. From Step 8, we know that $u(t) < u^*$ on intervals $[0, \widetilde{t}_J)$ and $(\bar{t}_J, \infty)$. Thus, $f(t) = 0$ on such intervals, and consequently, $a(\cdot) = \overline{a}(\cdot)$. By continuity of $a(\cdot)$ at $\widetilde{t}_J$, $\bar{t}_J$, we have $a(\bar{t}_i) = \overline{a}(\bar{t}_i)$ for $i \in \{H, L\}$.

Step 11. We show that $\bar{t}_J < \infty$. Suppose that $\bar{t}_J = \infty$. From Step 9, we have $a(t) < \overline{a}(t)$ for $t \in (\widetilde{t}_J, \bar{t}_J)$. Thus, for $t \to \infty$ we have $g(t) = \theta$, and hence $f(t) = \mu(1 - \sigma F(t)) \geq \mu(1 - \sigma) > 0$. It follows that $F(t) \to \infty$ as $t \to \infty$.

Step 12. We show that $\widetilde{t}_J < \bar{t}_J$. Suppose $\widetilde{t}_J = \bar{t}_J$. Combined with the conditions $\widetilde{t}_J \leq t^* \leq \bar{t}_J$, we have $\widetilde{t}_J = t^* = \bar{t}_J$. Because $\phi/\hat{\phi} < 1$, we have $t^* > 0$. From Step 8, $u(t) < u^*$ for $t \neq t^*$. Thus, for $t \neq t^*$ we have $a(t) = \overline{a}(t)$, and by continuity of $a(\cdot)$, the same is true at t^*. A straightforward calculation reveals that the agent's optimal cheating time t^*, and thus $g(t^*) = 0$. By implication $a(t^*) = 0$, contradicting continuity of $a(\cdot)$.

Step 13. We show that $a(\cdot)$ is increasing. From Step 12, we know that $\widetilde{t}_J < \bar{t}_J$. From Step 8 we know that $u(t) = u^*$ on $[t_L, t_H]$. Using the agent's indifference condition, we have $a(t) = \phi/\hat{\phi} + \kappa \exp(\rho t)$. Thus, $a(\cdot)$ is either strictly increasing, strictly decreasing, or monotone on this interval. Next, note that whether $\bar{t}_J > 0$ or $\bar{t}_J = 0$, we have $a(\bar{t}_J) \leq \overline{a}(\bar{t}_J) < \overline{a}(\bar{t}_J)$, where the last inequality follows from Step 12 combined with the fact that $\overline{a}(\cdot)$ is strictly increasing. Furthermore, from Steps 10 and 11 we have $\overline{a}(\bar{t}_J) = a(\bar{t}_J)$ for some $\bar{t}_J \in (\widetilde{t}_J, \infty)$. Thus, we have shown that $a(\widetilde{t}_J) < a(\bar{t}_J)$. Since the monotonicity of $a(\cdot)$ does not change on $(\widetilde{t}_J, \bar{t}_J)$, it must be increasing on this interval. Applying Step 10, it follows that outside this interval $a(t) = \overline{a}(t)$, which is itself an increasing function.

Step 14. We show that $\widetilde{t}_J > 0$. Suppose $\widetilde{t}_J = 0$. Using Step 9, we have $a(0) < \overline{a}(0) =$

$\gamma/(\gamma+\rho) < \frac{\phi}{\hat{\phi}}$. The agent's indifference condition on $[\widetilde{t}_J, \overline{t}_J]$ implies $a(t) = \phi/\hat{\phi} + \kappa\exp(\rho t)$. Thus, $a(0) < \phi/\hat{\phi}$ implies $\kappa < 0$. By implication, $a(\cdot)$ is decreasing on $(\widetilde{t}_J, \overline{t}_J)$, contradicting Step 13.

Step 15. We show that $\widetilde{t}_J < t^*$; that $\overline{t}_J > t^*$ is proved in a similar way. Suppose that $\widetilde{t}_J = t^*$. We know that $\overline{t}_J > \widetilde{t}_J$. From the agent's indifference condition on $[\widetilde{t}_J, \overline{t}_J]$, we have $a(t) = \phi/\widehat{\phi} + \kappa\exp(\rho t)$. Combined with Step 10, we have

$$\frac{\phi}{\widehat{\phi}} + \kappa\exp(\rho t^*) = 1 - \frac{\rho}{\rho+\gamma}\exp(-\gamma t^*).$$

From the definition of t^*, we have

$$\frac{\phi}{\widehat{\phi}} + \kappa\exp(\rho t^*) = 1 - \frac{\rho}{\rho+\gamma}(1 - \frac{\phi}{\widehat{\phi}}) \Rightarrow \kappa\exp(\rho t^*) = \frac{\gamma}{\gamma+\rho}(1 - \frac{\phi}{\widehat{\phi}}).$$

Next, consider the derivative of $a(\cdot)$ and $\overline{a}(\cdot)$ at t^*. We have

$$\begin{aligned}
&a'(t^*) = \rho\kappa\exp(\rho t^*), \qquad \overline{a}(t^*) = \frac{\rho\gamma}{\rho+\gamma}\exp(-\gamma t^*) \Rightarrow \\
&a'(t^*) = \frac{\rho\gamma}{\gamma+\rho}(1 - \frac{\phi}{\widehat{\phi}}), \qquad \overline{a}(t^*) = \frac{\rho\gamma}{\rho+\gamma}(1 - \frac{\phi}{\widehat{\phi}}) \Rightarrow \\
&a'(t^*) = \overline{a}(t^*).
\end{aligned}$$

Because (1) $a(\cdot)$ is increasing and convex (2) $\overline{a}(\cdot)$ is increasing and concave, (3) $a(t^*) = \overline{a}(t^*)$ and (4) $a'(t^*) = \overline{a}'(t^*)$, we have $a(\cdot) > \overline{a}(\cdot)$ for $t \in (t^*, \overline{t}_J)$, violating the logjam constraint.

Proof of (i). That the agent's mixing distribution has no mass points follows from continuity of $a(\cdot)$ (Step 2). Combined, Steps 8, 11, 12, and 14 establish the rest of the claim.

Proof of (ii). Proved in Steps 2 and 13.

Proof of (iii). Follows from Steps 10-15.

Proof of (iv). Follows from Step 9 and 11-15. □

Proof of Proposition 2.1. Strategies. Based on Lemma 2.1, there exist $\{\widetilde{t}_J, \overline{t}_J\}$ with $0 < \widetilde{t}_J < t^* < \overline{t}_J < \infty$ such that for $t \in (\widetilde{t}_U, \overline{t}_U)$, equilibrium strategies are characterized by the differential equations $\mu(t) = \mu$ and $a'(t) - \rho a(t) + \phi(\rho+\lambda) = 0$ with boundary conditions, $F(\widetilde{t}_J) = 0$, $F(\overline{t}_J) = 1$, $a(\widetilde{t}_J) = \overline{a}(\widetilde{t}_J)$,$a(\overline{t}_J) = \overline{a}(\overline{t}_J)$.. Solving the first differential equation and boundary conditions, we have

$$F(t) = \frac{1}{\sigma}(1 - \exp(-\mu(t - \widetilde{t}_J)), \qquad \overline{t}_J = \widetilde{t}_J + \overline{t}.$$

Solving the second differential equation, we have

$$a(t) = \frac{\phi}{\hat{\phi}} + \kappa\exp(\rho t),$$

where κ is an integration constant. Thus, the remaining boundary conditions become

$$\frac{\phi}{\hat{\phi}} + \kappa \exp(\rho \widetilde{t}_J) = 1 - \frac{\rho}{\rho+\gamma} \exp(-\gamma \widetilde{t}_J)$$

$$\frac{\phi}{\hat{\phi}} + \kappa \exp(\rho(\widetilde{t}_J + \bar{t})) = 1 - \frac{\rho}{\rho+\gamma} \exp(-\gamma(\widetilde{t}_J + \bar{t})).$$

Solving, we have

$$\kappa = \exp(-\rho \widetilde{t}_J)(\overline{a}(\widetilde{t}_J) - \frac{\phi}{\hat{\phi}}),$$

$$\exp(-\gamma \widetilde{t}_J) = (1 - \frac{\phi}{\hat{\phi}})(1 + \frac{\gamma}{\rho}) \frac{\exp(\rho \bar{t}) - 1}{\exp(\rho \bar{t}) - \exp(-\gamma \bar{t})}.$$

Obviously, $\widetilde{t}_J > 0$ and $\bar{t}_J < \infty$. That $\widetilde{t}_J < t^* < \bar{t}_J$ follows from a straightforward application of L'Hopital's rule. Note that, if we solve the second equation for κ, we find $\kappa = \exp(-\rho \bar{t}_J)(\overline{a}(\bar{t}_J) - \frac{\phi}{\hat{\phi}})$, which corresponds to the expression of the acceptance strategy presented in the proposition.

Beliefs. Obvious.

Payoffs. The agent is indifferent between faking at all times inside $[\widetilde{t}_J, \bar{t}_J]$, and therefore his equilibrium payoff is $u(\widetilde{t}_J)$ as stated in the proposition. In the next part of the proof, we give a simpler expression for the agent's equilibrium payoff. For the principal's payoff, not that an arrival inside $(\widetilde{t}_J, \bar{t}_J)$ delivers no surplus, while an arrival outside of this time interval is known to be real, and is accepted with probability $\overline{a}(\cdot)$, delivering payoff $(1-\theta)$. Furthermore, an arrival comes at $t > \bar{t}_J$ only if the agent is ethical.

Normative Implications. Consider the principal's payoff. Note that as $\sigma \to 1$, we have $\bar{t} \to \infty$, and hence,

$$\widetilde{t}_J \to -\frac{1}{\gamma} \ln[(1 - \frac{\phi}{\hat{\phi}})(1 + \frac{\gamma}{\rho})] = -\frac{1}{\gamma} \ln(\frac{1 - \frac{\phi}{\hat{\phi}}}{1 - \frac{\gamma}{\gamma+\rho}}).$$

Under the maintained assumption that $\gamma/(\gamma+\rho) < \phi/\widehat{\phi}$, this limit is strictly positive. Therefore, with a logjam, the principal's payoff is bounded away from zero as $\sigma \to 1$, while it converges to zero in the main model. Now consider $\sigma \to 0$. We have $\bar{t} \to 0$. Using L'Hopital's rule,

$$\frac{\exp(\rho \bar{t}) - 1}{\exp(\rho \bar{t}) - \exp(-\gamma \bar{t})} \to \frac{\rho}{\rho+\gamma},$$

which in turn implies that $\widetilde{t}_J \to t^*$, $\bar{t}_J \to t^*$. It follows that the principal's payoff approaches

$$(1-\theta) \int_0^\infty \lambda \exp(-(\lambda+\rho)t) \overline{a}(t) dt < (1-\theta) \int_0^\infty \lambda \exp(-(\lambda+\rho)t) dt,$$

which is the limit as $\sigma \to 0$ in the main model. □