Make It 'Til You Fake It*

Raphael Boleslavsky
University of Miami
r.boleslavsky@miami.edu

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

March 7, 2022

Abstract

We study a dynamic principal agent model of fraud and trust. The principal has limited power of commitment and wishes to accept a real project and reject a fake. The agent is either an ethical type that produces only a real project, or a strategic type that also has the ability to produce a fake. Producing a real project takes a positive and uncertain amount of time, while a fake project can be created instantaneously at some cost. We characterize the equilibrium, and explore two institutional remedies that improve the principal's welfare: opaque standards, and impediments in the approval process.

JEL Classifications: C73, D21, D82, L15, M42.

Keywords: deception, fraud, innovation, opaque standards, time pressure, trust.

*We thank Dan Bernhardt, Willie Fuchs, Alex Horenstein, Navin Kartik, Ayça Kaya, Kyungmin Kim, Nicolas Klein, George Mailath, Ricardo Serrano-Padial, Juuso Toikka and seminar audiences at Diego Portales University, Duke University, Indiana University, Penntheon (Penn State), Texas A&M, University of Chicago, University of Illinois, University of Miami, University of Pennsylvania, and University of Texas Austin (Economics and Finance).

1 Introduction

In 2003, Elizabeth Holmes dropped out of Stanford to found life-sciences company Theranos. For the next decade, the company worked to develop a technology that could run clinical tests using only tiny drops of blood, without drawing much attention or publicizing its findings. In 2013, Theranos—and Holmes herself—began to promote its technology heavily, claiming that it could conduct a wide variety of clinical tests more quickly, accurately, and cheaply than conventional methods.

In 2018 federal agencies charged Holmes and company President Ramesh Balwani of "raising more than $700 million from investors through an elaborate, years-long fraud in which they exaggerated or made false statements about the company's technology."[1] The DOJ claimed that the deception went beyond exaggerated statements, accusing Holmes and Balwani of conducting "misleading technology demonstrations... intended to cause potential investors to believe that blood tests were being conducted on Theranos proprietary analyzer; when[they] knew that [the] analyzer was running a 'null protocol'... to make [it] appear to be operating, but it was not."[2] Because of the complexity of the technology, Holmes and Balwani were able to produce a fake prototype that was sufficiently convincing to raise breathtaking amounts of capital from unwitting investors.

Though it is possible that Holmes believed that her company would eventually be able to deliver on her exaggerated claims, longtime observers have suggested that the fraud was likely motivated by short term gain. John Carreyrou, a reporter for the *Wall Street Journal* who was one of the first to raise questions about the company, believes that Holmes had an "all-consuming quest to be the second coming of Steve Jobs," which motivated her to "cut corners" when progress toward a real innovation stalled; "If there was collateral damage on her way to riches and fame, so be it" (Carreyrou, 2018).

While the Theranos fraud is an especially well-known example, the desire for short term gain can lead to misconduct in a variety of settings where malfeasence is hard to detect. For instance, facing pressure to complete a lucrative deal, a procurer may be tempted to purchase a counterfeit good or deal with a banned supplier, as in the case of AEY Inc., which illegally purchased Chinese munitions in order to fulfill a $300 million order from the United States government (Lawson, 2011). In response to competition from foreign manufacturers selling small, fuel-efficient vehicles in the 1970s, Ford Motor Company rushed to produce a similar car. Rapid development of the Pinto was achieved by skimping on safety testing, with disastrous results (Dowie, 1977). A realtor with access to other business opportunities

[1] Securities and Exchange Commission, Press Release 18-41.

[2] United States of America v. Elizabeth A. Holmes and Ramesh "Sunny" Balwani, 12(C).

may feel pressure to finish a deal quickly. Indeed, Levitt and Syverson (2008) note that "Real estate agents have an incentive to convince clients to sell their houses too cheaply and too quickly." Facing the risk of a hiring freeze, an academic department or government agency may be tempted to hire an under-qualified job candidate, rather than lose the slot.

In this paper we present and analyze a dynamic principal-agent model of fraud and trust in which the agent is motivated by short term gain and fraud is difficult to detect. The principal has limited commitment power and wishes to approve a *real* project and reject a *fake*. Real projects and fakes have different arrival processes: real project development takes a positive and uncertain amount of time, while a fake project can be manufactured instantaneously at some cost. The agent faces pressure to perform quickly: he is rewarded when his project is approved, and he is impatient. The agent is of two possible types. Both types have the same ability to develop a real project, but they vary in their willingness to commit fraud. An ethical type is unwilling to produce a fake, while a strategic type can generate a costly fake at will. Furthermore, real projects and fakes are indistinguishable for the principal at the moment they are submitted. She only observes the time at which a project is proffered when deciding whether to accept or reject it. Thus, the time of project arrival plays a critical role in the principal's approval decision, permitting her to learn both about the agent's type and the project's authenticity.[3]

In the unique equilibrium of the game, both the principal and strategic type of agent play mixed strategies on a finite time interval at the beginning of the interaction, which we refer to as the *phase of doubt*. Specifically, during this interval the strategic agent searches for a real discovery, but also randomly creates a fake. Thus, when the principal receives a submission, she is uncertain whether it constitutes a real breakthrough by an ethical agent, a real breakthrough by a strategic agent, or a fake. In equilibrium, the principal employs a random approval policy over the phase of doubt. To offset the strategic agent's desire to commit fraud as early as possible, the principal's equilibrium approval probability increases over time. Furthermore, as time passes without receiving a submission, the principal's confidence that the agent is ethical grows. Thus, the strategic type can fake more frequently without triggering rejection. If no project is submitted, the principal's trust grows further, incentivizing even more frequent cheating by the strategic type. Indeed, the strategic type submits a fake with probability 1 by some finite date, at which point the phase of doubt ends. After that, the principal is fully confident that the agent is ethical and she approves any submission with probability 1.

[3]In a previous version of the paper, we extend the baseline model, endowing the principal with a costly auditing technology which she can deploy before deciding whether to approve or reject the project. If the auditing technology is not too costly, then access to it increases the principal's welfare. Details available on request.

Because the principal is indifferent between approving and rejecting projects that are submitted during the phase of doubt, her expected payoff is equal to her payoff from rejecting. In other words, she benefits from the arrival of a project if and only if the arrival occurs after the phase of doubt is over, when her trust in the agent is fully established. Since a submission is made after the phase of doubt only if (i) the agent actually is ethical and (ii) he does not make a discovery during the phase of doubt, the principal's equilibrium payoff is relatively low in the baseline model. The rest of the paper, therefore, is dedicated to investigating two institutional remedies that the principal can use to improve her situation.

First, we consider opacity of the principal's standards. We consider a setting in which the agent interacts with one of two types of principal with different tolerances for accepting fakes, showing that both the high-standard and low-standard types of principal are better off when the agent is uncertain about which type of principal he faces. In particular, when the principal is sufficiently likely to have high standards for approval, in equilibrium the agent behaves as if he *only* faces this type. Thus, the high standards principal obtains the same equilibrium payoff under opacity as under transparency. However, we show that the low-standards principal is strictly better off. In the complementary case, when the probability of the high standards principal is low, the phase of doubt is divided into two stages: an initial stage with aggressive cheating, followed by a second stage with mild cheating. In this case, opacity strictly benefits both types of principal.

Second, we consider impediments in the approval process. In particular, we study a "logjam" which prevents the principal from approving a project until it clears via a Poisson arrival.[4] The logjam effectively imposes a time-varying upper bound on the ex ante probability that a submission arriving at a given time is approved, which becomes less restrictive over time. In this case, the equilibrium structure is characterized by *two* phases of credibility surrounding a single phase of doubt. As in the main model, the phase of doubt is an interval (with duration identical to that of the main model), but unlike the main model, the phase of doubt begins at a strictly positive time, rather than at the beginning of the game. Intuitively, the agent expects the principal to be jammed with high probability early on, and therefore has less incentive to pay the cost of faking in order to rush his submission out. Because an unjammed principal accepts early submissions, the first credibility phase generates an expected benefit relative to the main model, where early arrivals generate no expected surplus. However, the second phase of credibility imposes a relative cost, since the principal is constrained by the logjam to accept a real project with probability less than one, while in the main model she

[4]In order to make the underlying forces as clear as possible, we focus on the case where the logjam clears privately. However, we have also analyzed the case in which it clears publicly, and have verified that the principal also can benefit in this case. Details available on request.

accepts any submission with probability 1 once the phase of doubt is over. We show that when the probability of a strategic agent is sufficiently large (small), the logjam benefits (harms) the principal.

In the next section, we review related literature. We present the model in Section 3. We characterize the equilibrium of the main model in Section 4. Sections 5 and 6 address the remedies. Concluding remarks appear in Section 7. Proofs can be found in the Appendix.

2 Literature

At a broad level, our paper is related to a recent stream of work concerned with cheating, gaming, and subterfuge in principal-agent relationships. Barron, Georgiadis, and Swinkels (2019) consider the design of compensation contracts for agents who can "game the system" by gambling with intermediate output, thereby adding mean-preserving noise. In such an environment, the agent's wage must be a concave function of his output, necessitating linear ironing on intervals where the standard contract is convex. A different perspective on gaming is presented in Frankel and Kartik (2019), who study a signaling model in which agents differ both in their "natural actions" and in their "gaming ability." The authors show that actions convey muddled information about both dimensions and derive conditions under which an increase in the stakes tilts information provision toward gaming ability. Glazer, Herrera, and Perry (2019) study the informativeness of a product review when the evaluator may be a dishonest type, who can submit a fake review in order to make the product appear good. In equilibrium, the informativeness of reviews is compressed: past a cutoff, all positive reviews have the same effect on beliefs.[5]

The evolution of the principal's belief about the agent's integrity plays a key role in our analysis. In this sense, our work is connected to the literature on reputation in long term relationships. To our knowledge, ours is the first paper in this area that explores the link between the maturation of projects and the growth of reputational capital. Sobel (1985) considers a repeated cheap talk game in which the agent may be either a "friend" of the principal, with aligned preferences, or an "enemy," with opposing preferences. The enemy cultivates his reputation by sometimes issuing honest advice in periods with moderate stakes. When the stakes become sufficiently high, the enemy exploits his reputation by issuing a self-serving recommendation, thereby revealing his type. Bar-Isaac (2003) studies how reputation affects a monopolist's decision to abandon a market. In equilibrium, the good type of seller signals that his product is likely to be of high quality by staying in the market, despite an

[5]See also Perez-Richet and Skreta (2018) who consider the design of an optimal test when the agent has the ability to manipulate the process by which the test determines his type

unlucky run in which realized product quality is low. Ely and Välimäki (2003) study a model of advice in which a long-lived expert advises a sequence of short-lived principals, who observe past recommendations, but not past states. The authors highlight a perverse incentive, whereby the "good" advisor is disinclined to make recommendations that might make him appear to be the "bad" type, even if such recommendations are actually warranted. Deb, Mitchell, and Pai (2019) also explore a dynamic model of expertise. In each period, the agent privately observes the arrival of information before choosing whether to act on it. Only a good agent can acquire information, which can be either high or low quality. To maintain his reputation, a good agent is sometimes tempted to act on low quality information. Kolb and Madsen (2020) develop a dynamic principal agent model in which a principal runs a project, which may be implemented by a disloyal agent. The principal controls the evolution of the project stakes, which increase both the principal's flow benefit from honest performance and a disloyal agent's flow benefit from undermining. The principal detects undermining stochastically, and thus the evolution of stakes affects the principal's flow payoff and her ability to root out disloyalty.

While our paper focuses on an agent's ability to generate an artificial arrival, another strand of literature focuses on an agent's ability to suppress or delay an arrival, particularly in the context of information or news. Gratton, Holden, and Kolotilin (2018) study a dynamic persuasion model in which a stochastic arrival privately informs the sender of his type. Once the sender discloses that he has learned his type (without disclosing what it is), the receiver begins to draw informative signals about it. Early disclosure provides the receiver with more opportunities to learn about the sender and therefore signals good news. Shadmehr and Bernhardt (2015) analyze a ruler's incentive to suppress media reports, showing that the ruler can benefit from a commitment to censor less than he does in equilibrium. Sun (2018) considers a dynamic model of censorship, demonstrating that when the arrival of bad news is inconclusive, it is censored aggressively by the good type of ruler, which can improve information quality and lead to a Pareto improvement. In a different vein, Li, Matouschek, and Powell (2017) study power dynamics in a relational contract. In each period, the principal approves or vetoes an agent's recommended project, without observing whether her own preferred project is available. Thus, the agent can suppress the arrival of the principal's preferred project, hoping to implement his own.

Our analysis does not allow for transfers and limits the principal's commitment power, but it is nevertheless related to the literature on dynamic moral hazard contracts in which the agent's effort accelerates a project's arrival (Bergemann and Hege, 1998, 2005; Mason and Välimäki, 2015; Sun and Tian, 2018). In these papers, the agent's effort is costly but increases the arrival rate of a success. In contrast, in our analysis, cheating *increases* the arrival rate of a "success" while *decreasing* its quality to the point that the principal would prefer to reject. We are aware of only two dynamic contracting papers — Klein (2016) and Varas (2018) — that allow the agent to act in a similar manner.

In Klein (2016), the principal hires an agent to experiment by generating public information in the form of a state-contingent Poisson process. In addition to the experimentation technology, the agent has access to a specious technology which produces Poisson successes (that appear identical to the ones generated by the experimentation technology) at a rate that is independent of the state. Thus, a specious success is uninformative and worthless to the principal. The author shows that the optimal compensation contract backloads payments. By contrast, the agent in our model possesses a technology for generating a single fake project rather than a stream of false data. In this context, we find that an early arrival has no value to the principal, while a late (enough) arrival must be authentic.

In the contracting environment investigated by Varas (2018), the agent chooses in each instant whether to *work, shirk*, or *gamble*. Working generates high quality output after an uncertain amount of time and effort, while gambling generates an output of random quality that is difficult for the principal to verify. The optimal contract derived by Varas (2018) exhibits two phases: an initial phase of diminishing payments followed by a stationary phase in which the agent is not punished for production delays. The principal in Varas (2018) learns about project quality post–submission, while the principal in our setting learns about the integrity of the agent pre–submission. More generally, Varas (2018) underscores the limits of high–powered incentive contracts, whereas our findings point to the crucial role played by reputation and trust in a setting marked by limited commitment.

Finally, the institutional remedies we consider are related in some degree to a body of prior work. The first remedy that we study allows the principal to maintain private information about her preferences. Ederer, Holden, and Meyer (2018) study a related form of opacity in a multi-task moral hazard model, where the principal deters gaming by committing to randomize between different performance measures. The final remedy we consider forces the principal to postpone her decision until she has a private Poisson arrival. In the language of Aghion and Tirole (1997), the principal holds the formal authority, but she must wait for a Poisson arrival to acquire real authority. Limiting the principal's approval power incentivizes the strategic agent to delay faking and improves the principal's welfare. This result is reminiscent of Fuchs and Skrzypacz (2015, 2019), who show that shutting down a market at certain times changes the dynamic incentives to trade and enhances efficiency.

3 The Model

A principal (she) interacts with an agent (he) over an indefinite horizon. Time is continuous and both parties discount future payoffs at rate $\rho > 0$.[6] The agent develops a project over time that he submits to the principal for approval. The project can be developed using a technology that is either *authentic* or *fraudulent*. If the agent uses the authentic technology at time t, then a *real* project arrives at Poisson rate λ. The fraudulent technology allows the agent to instantly develop a *fake* project, at fixed cost $\phi \in [0, 1)$. Thus, the authentic technology is free but slow, while the fraudulent technology is costly (if $\phi > 0$) but fast.[7]

The agent is one of two types: *strategic* with probability $\sigma \in (0, 1)$ or *ethical* with probability $1 - \sigma$. The strategic agent can use either technology while the ethical agent can only use the authentic one.[8] The agent's type is private information, but the value σ is common knowledge. Below whenever we write of the agent choosing whether to generate a fake project, it should be understood that we are referring to the strategic type.

Once a project has been developed, it is instantly submitted to the principal for approval.[9] The state of the project — real or fake — is not directly observable upon submission. The principal only observes the time at which the project was submitted when deciding whether to approve or reject it.

The principal would like to approve real projects and reject fakes. Her payoffs are normalized so that approving a real project yields $1 - \theta > 0$ and approving a fake yields $-\theta < 0$. If she rejects a submission then she receives 0 regardless of its state. Preference parameter

[6]For most of our results, whether ρ is the rate of time preference or the rate at which the game ends does not matter. However, when we investigate the possibility of organizational logjams in Section 6, the latter interpretation is most natural.

[7]If the authentic technology imposed a positive flow cost on the agent, then he could engage in fraud in order to shirk, thereby avoiding the cost of authentic project development. Our goal is to study the dynamic incentive to commit fraud, which has received relatively little formal analysis, as opposed to the incentive to shirk, which has been researched extensively. Furthermore, incorporating a small positive flow cost for the authentic technology would have no impact on our results. Normalizing the expected flow cost under the real technology back to zero would manifest technically as a reduction in ϕ (calculations in Appendix).

[8]The assumption that the ethical agent cannot use the fake technology is for expositional convenience. Suppose that we assume that the ethical type can access the fake technology, but that the ethical type is more patient, has a smaller payoff when a fake is approved, or has a larger cost of faking. If the differences across these dimensions are sufficient, then the ethical type would not fake in equilibrium.

[9]This assumption rules out equilibria in which both types of agent are compelled to delay by the principal's off–path belief that submissions arriving at certain times must be fake. A refinement in the spirit of D1 would eliminate such off–path beliefs, ruling out equilibria involving delays. In addition, in all model variants, the equilibrium that we derive is robust to allowing the agent to delay the submission of the project.

$\theta \in (0, 1)$ thus represents the principal's tradeoff between type I and type II errors. The strategic agent would like his project to be accepted regardless of its state, obtaining a gross benefit of 1 from acceptance, and 0 from rejection.[10]

4 Equilibrium Characterization

In this section, we characterize the weak Perfect Bayesian equilibrium (henceforth equilibrium) of the game and show that it is generically unique.[11] An equilibrium consists of strategies for the agent and the principal and a belief function for the principal regarding the state of a submitted project, such that (i) the agent's strategy is optimal given the principal's acceptance strategy, (ii) the principal's acceptance strategy is sequentially rational given her beliefs, (iii) the principal's belief about a submitted project is derived from Bayes' rule.[12]

Strategies. A pure strategy for the strategic agent is a choice of a "cheating time" $t \in \{\mathbb{R}_+ \cup \infty\}$ at which he will submit a fake project if a real one has not yet arrived. A mixed strategy for a strategic agent is a probability measure over finite cheating times represented by cumulative distribution function $F(\cdot)$.[13] A strategy for the principal is an acceptance function $a(\cdot)$ on the domain $\mathbb{R}_+$, which specifies the probability with which a submission at time t is approved.

Beliefs. If a project is submitted at time t, the principal's belief that it is real must be derived by Bayes' rule as the probability of a real arrival at t given an arrival at t.

Lemma 4.1 (Beliefs). *If the strategic agent submits a fake project according to the cumulative distribution function $F(\cdot)$ with density $f(\cdot)$, then, the probability that a submission at time t is real is*

$$g(t) = \frac{\lambda}{\lambda + \mu(t)}, \tag{1}$$

[10]The analysis and results are similar if the agent's payoff from having a fake project approved is smaller than his payoff from having a real project approved. Details available upon request.

[11]Multiple equilibria do exist iff $\phi = \frac{\rho}{\rho+\lambda}$. To streamline presentation, we abstract from this and other non-generic cases.

[12]Because a real project arrives at a positive rate and is assumed to be submitted immediately, all times $t \geq 0$ are on the equilibrium path. Thus, our characterization does not exploit the freedom to specify off-path beliefs granted by weak PBE.

[13]Strictly speaking, the strategic agent can choose never to cheat with positive probability, thereby allocating some probability mass to $t = \infty$. However, given Assumption 1 below, the strategic agent cheats with probability 1 in finite time in equilibrium.

where

$$\mu(t) \equiv \frac{\sigma f(t)}{1 - \sigma F(t)}.$$

The function $\mu(\cdot)$ is the hazard rate of a fake arrival: it is the likelihood that a fake arrival is generated at time t, given that one was not generated earlier. It is important to point out that $\mu(\cdot)$ is the hazard rate of a fake *from the principal's perspective*, because it accounts for her uncertainty about the agent's type, reflected in the parameter σ. Note that λ is the hazard rate of a real arrival, (which can be generated by either type of agent) and thus the principal's belief that an arrival at time t is real is simply the ratio of the hazard rate of a real arrival to the sum of the hazard rates of a real and fake arrival.

Principal's Decision. If the principal believes that a project that arrives at time t is real with probability $g(t)$, then her expected payoff from approving it is

$$g(t)(1-\theta) - (1-g(t))\theta = g(t) - \theta,$$

and therefore, the principal's sequentially rational acceptance strategy must satisfy

$$a(t) = \begin{cases} 1 & \text{if} \quad g(t) > \theta \\ [0,1] & \text{if} \quad g(t) = \theta \\ 0 & \text{if} \quad g(t) < \theta. \end{cases} \tag{2}$$

Agent's Decision. If the strategic agent adopts a pure strategy in which he will cheat at time t given that no real arrival has been generated by that point, then – holding fixed the principal's strategy – the agent's expected payoff is

$$u(t) = \int_0^t \lambda \exp(-(\rho+\lambda)s) a(s) \, \mathrm{d}s + \exp(-(\rho+\lambda)t)(a(t) - \phi). \tag{3}$$

The integral represents the discounted expected payoff to the agent from real arrivals that occur at all times $s < t$. If no real arrival occurs before t, then the agent submits a fake project which costs ϕ and is accepted with probability $a(t)$. In the limit where the agent *never* submits a fake project, he can secure a non-negative payoff of $u(\infty)$. Indeed, when the cost of submitting a fake project is sufficiently high, then the agent never submits one in equilibrium. This is formalized in the following lemma.

Lemma 4.2 (No Fakes)**.** *If $\phi > \widehat{\phi} \equiv \frac{\rho}{\rho+\lambda}$, then there is a unique equilibrium of the game and it involves the agent never submitting a fake project and the principal approving any submission she receives with probability 1.*

The intuition is straightforward. Suppose for the moment that the principal approves any submission she receives with probability 1. Given the stationarity of the environment, the strategic agent effectively faces two alternatives: he can submit a fake immediately and earn payoff $1-\phi$, or he can wait for a real arrival, earning payoff $\lambda/(\rho+\lambda) = 1-\widehat{\phi}$. Obviously, when $\phi > \widehat{\phi}$, he prefers the second alternative. In this case, the agent never submits a fake, and it is sequentially rational for the principal to approve any submission she receives with probability 1. Motivated by this observation, we maintain the following assumption (without restatement) for the rest of the paper.

Assumption 1. *The cost of faking is sufficiently low that the equilibrium in which all submissions are real does not exist: $\phi < \widehat{\phi}$.*

When $\phi < \widehat{\phi}$, the strategic agent must submit a fake project with positive probability in equilibrium. He cannot, however, submit a fake with positive probability at any specific point in time t because the probability of a real arrival at t is 0, implying that the principal's best response would be to reject with probability 1. This suggests that in equilibrium, the strategic type of agent must partially pool with the ethical type by submitting a fake project according to some probability density function, $f(\cdot)$. In order for randomization to be optimal for the strategic agent, he must be indifferent between all cheating times that he might select. Given the agent's impatience, he will be tempted to fake early, because an early approval is more valuable. To maintain indifference, it must be that early submissions are approved less often than later ones. This intuition is formalized in the following lemma.

Lemma 4.3 (Equilibrium Structure)**.** *In any equilibrium of the game*

(i) the time at which the agent submits a fake is drawn from a continuous mixed strategy with no mass points or gaps supported on an interval $[0, \bar{t}]$, where $\bar{t} \in (0, \infty)$.

(iii) for $t \in [\bar{t}, \infty)$, the principal always approves the project, $a(t) = 1$.

(ii) for $t \in [0, \bar{t})$, the principal's strategy $a(\cdot)$ is strictly increasing, continuous, and differentiable almost everywhere, with $\lim_{t \to \bar{t}} a(t) = 1$.

In equilibrium, the interval of arrival times is divided into two phases: an early *phase of doubt* $[0, \bar{t})$, in which the agent's submission is treated with skepticism, inducing the principal to reject with positive probability, and a late *phase of credibility*, $[\bar{t}, \infty)$ in which a submission originates only from the ethical type and is approved with certainty. The strategic agent's mixed strategy is supported continuously on the entire phase of doubt. It can have no mass points because this would induce rejection, and it can have no gaps because this would induce acceptance thereby creating a profitable deviation. Building on these observations, it is also possible to show that the phase of doubt must be finite (for $\sigma < 1$). Mathematically, it is simple to show that the cheating rate, $\mu(t)$, approaches zero as time passes, regardless of the agent's strategy.[14] It follows that at some finite time, the cheating rate becomes small enough that the principal strictly prefers to accept an arrival, and the agent never waits past this time to submit a fake, resulting in a finite phase of doubt.

Because the strategic agent mixes over the phase of doubt in equilibrium, his payoff $u(t)$ must be constant. Using this observation, we show that the approval probability $a(\cdot)$ is continuous and differentiable. Furthermore, because the agent is impatient, the approval probability must rise over time so as to maintain indifference between early fake submissions and later ones. Moreover, the acceptance probability approaches one at the end of the phase of doubt. Indeed, once the phase of doubt ends, the principal knows that the agent is ethical. By implication, for times near the end of the phase of doubt, the principal must accept with probability approaching one; otherwise, the agent could benefit by delaying his fake until the phase of doubt is over.

Lemma 4.3 implies that during the phase of doubt the mixed strategies for the principal and agent must obey a pair of first-order linear differential equations,

$$g(t) = \theta \;\Rightarrow\; \frac{\sigma f(t)}{1-\sigma F(t)} = \frac{\lambda(1-\theta)}{\theta}, \tag{4}$$

$$u'(t) = 0 \;\Rightarrow\; a'(t) - \rho a(t) + \phi(\rho+\lambda) = 0. \tag{5}$$

The first equation requires the principal to be indifferent between accepting and rejecting an arrival, while the second requires the agent to be indifferent about submitting a fake over all times inside the phase of doubt. Solving the first equation with boundary condition $F(0) = 0$ (which comes from the absence of a mass point at $t = 0$) yields the agent's equilibrium mixed strategy. Using the equilibrium mixed strategy, we find $\bar{t}$ by solving $F(\bar{t}) = 1$. Finally, solving the second differential equation with boundary condition $a(\bar{t}) = 1$ yields the principal's acceptance strategy. To characterize the equilibrium most succinctly, define

$$\mu \equiv \frac{\lambda(1-\theta)}{\theta}.$$

Note that the constant μ defined above is the *equilibrium* cheating rate and should not be confused with the function $\mu(t) = \sigma f(t)/(1-\sigma F(t))$, the cheating rate in general.

Proposition 4.1 (Equilibrium Fakes and Approvals.)**.** *The unique equilibrium of the game is characterized as follows.*

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

$$F(t) = \frac{1}{\sigma}(1-\exp(-\mu t)) \tag{6}$$

[14]Integrability requires $f(t) \to 0$ as $t \to \infty$, while $(1-\sigma F(t)) \geq 1-\sigma > 0$. Hence, $\mu(t) \to 0$.

supported on interval $[0, \bar{t})$, *where*

$$\bar{t} = -\frac{\ln(1-\sigma)}{\mu}. \tag{7}$$

If $t \in [0, \bar{t})$, *then the principal accepts with probability*

$$a(t) = \frac{\phi}{\widehat{\phi}} + \left(1 - \frac{\phi}{\widehat{\phi}}\right) \exp\{-\rho(\bar{t} - t)\}, \tag{8}$$

and with probability 1 otherwise.

Beliefs. *The principal's belief that she faces the ethical agent given no submission by* t *is* $\frac{1-\sigma}{1-\sigma F(t)}$. *On the other hand, if she receives a submission at* $t \in (0, \bar{t})$, *then she believes it is real with probability* $g(t) = \theta$, *and* $g(t) = 1$ *otherwise.*

Payoffs. *The strategic agent's equilibrium payoff is* $U^S = a(0) - \phi$, *and the ethical agent's payoff is* $U^E = U^S - (\widehat{\phi} - \phi)\exp(-(\rho + \lambda)\bar{t})$. *The principal's payoff is*

$$V = (1-\sigma)(1-\theta) \int_{\bar{t}}^{\infty} \lambda \exp\{-(\rho+\lambda)s\}\, \mathrm{d}s = (1-\sigma)(1-\theta)\frac{\lambda}{\rho+\lambda} \exp(-(\rho+\lambda)\bar{t}) \tag{9}$$

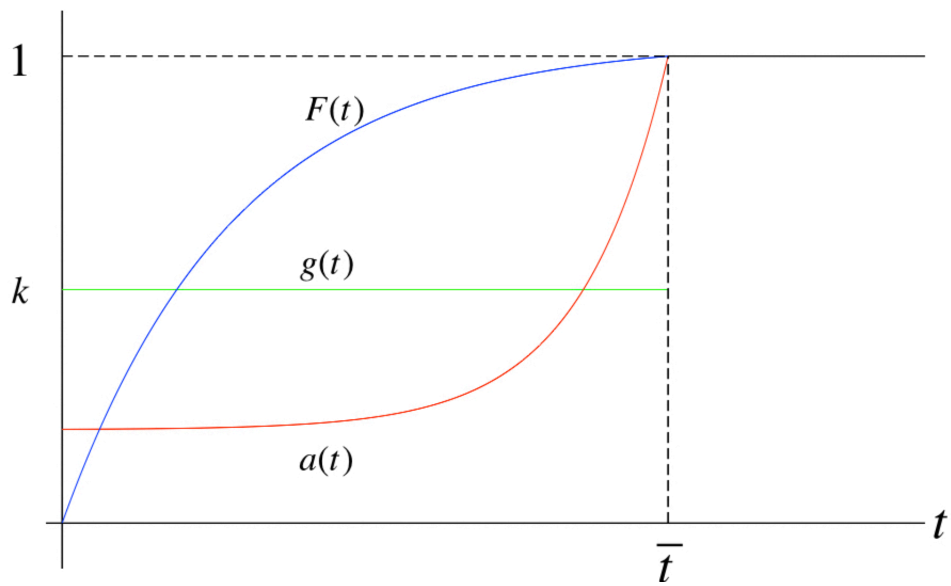

Figure 1: *Proposition 4.1*

We discuss each aspect of Proposition 4.1, beginning with the strategies. Formally, the agent's indifference condition, coupled with the finite phase of doubt implies that the principal must approve early submissions—that are more tempting to fake—with lower probability and late submissions with higher probability. The differential equation and boundary condition for the agent's indifference deliver an acceptance function of a particular functional form: a constant, plus an exponential function with growth rate ρ, the discount rate.

To understand the shape of the acceptance strategy it is helpful to consider the differential equation for $a(\cdot)$, given in (4). This equation admits a particular solution with a constant acceptance function $a(t) = \phi/\widehat{\phi}$ and a complementary solution $a(t) = \exp(\rho t)$.[15] To see where the particular solution comes from, note that by delaying, the strategic agent loses the opportunity to generate an instant acceptance with probability $a(t)$, losing $\rho a(t)$ at the margin. However, by delaying, the agent may save on the faking cost ϕ, either if the game ends due to an exogenous shock (at rate ρ) or a real project arrives (at rate λ), resulting in a marginal gain of $\phi(\rho + \lambda)$. Thus, the constant term equates the marginal loss and gain from delay, leaving the agent indifferent. To see where the exponential term appears, note that (4) also accounts for the increases in the acceptance probability over time, reflected in the term $a'(t)$. Thus, by marginally delaying, the agent also increases the probability that his fake will be approved. To maintain the agent's indifference, the increase in the acceptance probability must be exactly offset by discounting, resulting in an acceptance strategy that grows at rate ρ.

What about the agent's strategy? If the principal mixes over the phase of doubt, then her belief that a submission is real must be constant, $g(t) = \theta$. Thus, from the principal's perspective, the arrival rate of a fake must also be constant, as in (5). Because the authentic technology has a constant arrival rate, the fraudulent technology must appear to the principal to be deployed with a constant rate so that she does not learn about the state of a project from observing the time at which it was submitted. Solving (5), we find that the only distributions with constant arrival rate μ are truncated exponentials of the form $\frac{1}{\sigma}(1 - \kappa \exp(-\mu t))$, where κ is an integration constant. Finally, ruling out a mass point at zero yields, $\kappa = 1$, delivering the stated distribution.

We consider payoffs next. In equilibrium, the strategic agent is indifferent over submitting a fake project at any time within the phase of doubt. Thus, his payoff must equal the expected return to faking at $t = 0$, where it is accepted with the lowest probability, $a(0)$. Unlike the strategic agent, the ethical agent has no opportunity to fake. In other words, he must "wait forever" to fake, which yields payoff $u(\infty)$. By implication, the ethical agent's equilibrium payoff is strictly lower than the strategic agent's, who could always mimic the ethical type's strategic but strictly prefers not to (when $\phi < \widehat{\phi}$). Furthermore, compared to the case when $\sigma \approx 0$, both the ethical and strategic agents are worse off. If the principal believes the agent is very likely ethical, the phase of doubt collapses to zero, and both types' submissions are almost certain to be approved.

[15] A constant particular solution satisfies $-\rho a + \phi(\rho + \lambda) = 0$; dividing by ρ and substituting yields the claimed solution. The complementary solution satisfies $a'(t) - \rho a(t) = 0$ and is evidently exponential with growth rate ρ.

To understand the principal's payoff and the normative implications of faking, note first that for an arrival during the phase of doubt ($t \leq \bar{t}$), the principal mixes between approval and rejection, and therefore, she expects zero surplus from any arrival during this phase. Past the phase of doubt ($t > \bar{t}$), only the ethical agent is still active: in equilibrium, the strategic agent submits a fake before $\bar{t}$ with probability 1. Consequently, when a project is submitted past time $\bar{t}$, the principal is confident that it is real, and she approves it. Thus, an arrival after $\bar{t}$ generates expected surplus $1 - \theta$ for the principal.

Together, these observations imply that the principal expects a positive surplus from the project only when two conditions are met. First, the agent must be ethical: if the agent is strategic, then he will submit during the phase of doubt with probability 1 in equilibrium, and his arrival, whether real or fake, generates zero *expected* surplus for the principal. Second, the real technology must produce an arrival relatively late, after the phase of doubt is over. If the ethical agent is "lucky" and produces a real arrival quickly, it also generates no expected surplus for the principal. This normative implication is particularly pernicious: absent fraud, it is the early arrivals that are most valuable to her. In the next two sections we explore possible institutional remedies to this problem, opaque standards, and bureaucratic delays.

5 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.[16] The agent's expected payoff of selecting cheating time t is identical to his payoff in the main model, substituting the

[16] 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.

expected acceptance probability $a_U(\cdot)$ for the acceptance probability $a(\cdot)$ of the main model. Similar arguments to those in Lemma 4.3 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 mixes and the high standards principal rejects. In the second sub-phase, the low standards principal accepts, and the high standards principal mixes.

Lemma 5.1 (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 5.1 (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.*

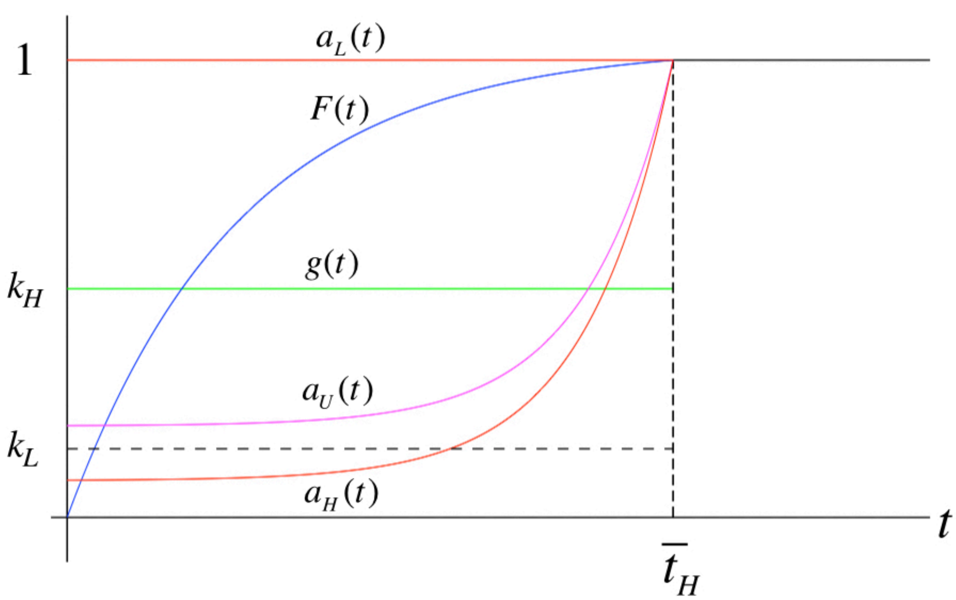

Figure 2: *Proposition 5.1*

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 5.2 (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 $\widetilde{t}_U, \overline{t}_U$ in the two stage equilibrium are*

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

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

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

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

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.*

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, $t \in [0, \widetilde{t}_U)$, 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.

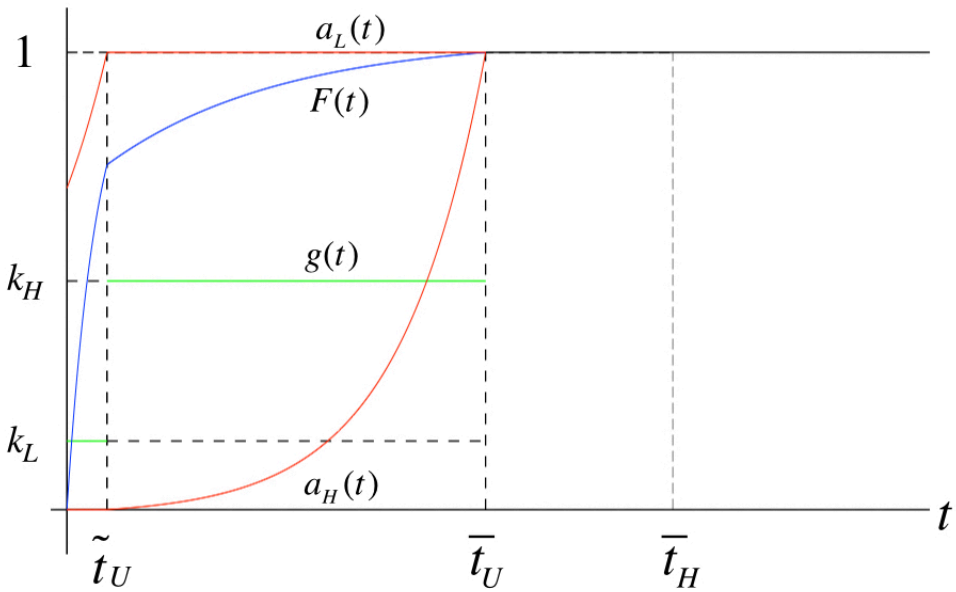

Figure 3: *Proposition 5.2. Note that $g(\cdot)$ jumps from θ_L to θ_H at $\widetilde{t}_U$.*

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.[17]

6 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.[18] 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

[17] 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.

[18] 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.

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, $a_J(t)$.[19] In other words, the unjammed type approves immediately with probability $a_J(t)$ and the jammed type approves with this same probability as soon as she is able, provided the game does not end first.[20]

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 2. *The rate at which the logjam clears is sufficiently low that $\bar{a}(t)$ is less than*

[19] 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.

[20] 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.

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.[21] Unlike the main model, where 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 6.1. *(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 $t \geq 0$.

[21] 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.

(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 6.1. *(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.*

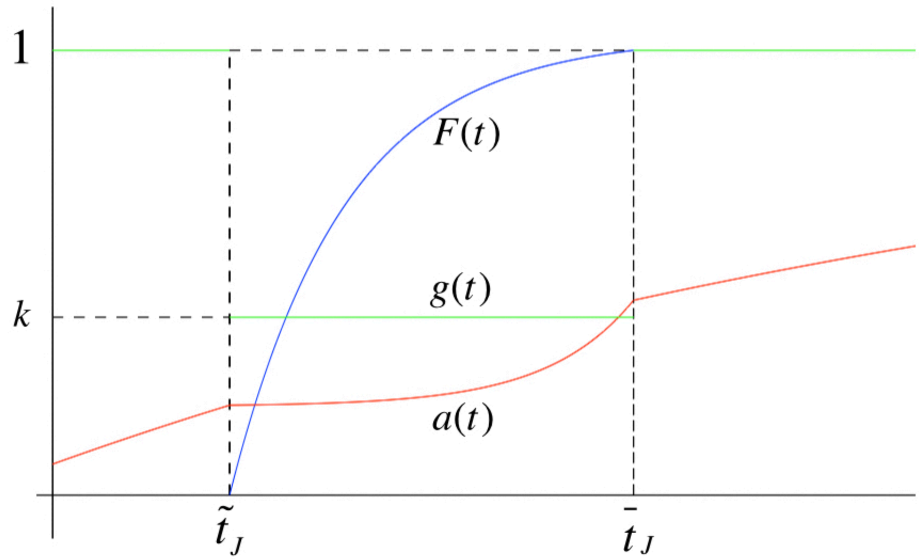

Figure 4: *Proposition 6.1. Note that $g(\cdot)$ jumps from 1 to θ at $\widetilde{t}_J$ and back up to 1 at $\bar{t}_J$.*

With a logjam, the equilibrium has *two* phases of credibility, an early one $(0, \widetilde{t}_J)$, and a late one $(\bar{t}_J, \infty)$, with a single phase of doubt $[\widetilde{t}_J, \bar{t}_J]$, 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, $\bar{t}_J - \widetilde{t}_J = \bar{t}$. 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 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.

7 Conclusion

In this paper we investigate a dynamic model of fraud and trust in which malfeasance is motivated by desire for a short-term gain. A principal with limited power of commitment faces an agent whose type — ethical or strategic — is private information. Both types of agent would like to complete the project quickly. However, producing a real project takes a positive and uncertain amount of time, whereas a fake project can be fabricated instantly. An ethical agent can only develop a real project, while a strategic agent chooses between developing a real project and fabricating a fake one.

In the unique equilibrium of the baseline model, the strategic agent randomizes about when to commit fraud and the principal randomizes about whether to accept a project as a function of its time of submission. As time passes without an arrival, the principal's belief that the agent is ethical rises until the point when she obtains full trust and accepts any subsequent submission with certainty. Indeed, the principal receives a positive expected payoff from a project submission if and only if it arrives after she obtains full trust in the agent. This generates a relatively low present expected equilibrium payoff for the principal. This leads us to explore two institutional remedies capable of improving her welfare: opaque standards, and the possibility of a bureaucratic logjam that constrains the principal.

In our analysis, the motivation to commit fraud is rooted in the agent's time-preference, deriving either from intrinsic impatience or a constant hazard that the game will end.[22] The

[22] Though we model time-preference as discounting, similar results hold if the agent's time preference is

agent commits fraud in order to accelerate the arrival of his reward from project approval. Of course, fraud can be motivated by other forces. For example, if there is a positive probability that the real technology may *never* deliver an arrival, then the agent's pessimism about the viability of the real technology grows over time. If enough time passes, the agent might resort to faking because he doubts that a real project can ever be produced. A related incentive to fake is generated by a deadline. As the deadline approaches, the probability of meeting it by honest means decreases, and the agent may resort to faking in order to do so. In a somewhat different vein, an agent may be tempted to fake in order to improve or maintain a reputation. For example, if there is uncertainty about the arrival rate of the agent's real technology, then an agent who would like to be perceived as "fast" might want to fake an arrival in order to affect the belief about his ability. With relatively straightforward modifications, each of These alternatives can be analyzed within the general framework we present here.

Relatedly, we also believe that our analysis can be extended to study fraud and trust building in a number of other settings. For instance, it would be edifying to know how the potential for repeated interactions influences the agent's equilibrium incentives to commit fraud and the principal's incentives to approve submissions. Yet another possibility would be to investigate the impact on incentives of competition between several agents engaged in an innovation race. We plan to address these and other related questions in future work.

References

Aghion, P., and J. Tirole, 1997, "Formal and real authority in organizations," *Journal of political economy*, 105, 1–29.

Bar-Isaac, H., 2003, "Reputation and Survival: learning in a dynamic signalling model," *The Review of Economic Studies*, 70, 231–251.

Barron, D., G. Georgiadis, and J. Swinkels, 2019, "Optimal Contracts with a Risk-Taking Agent," *Theoretical Economics, forthcoming.*

Bergemann, D., and U. Hege, 1998, "Venture capital financing, moral hazard, and learning," *Journal of Banking & Finance*, 22, 703–735.

Bergemann, D., and U. Hege, 2005, "The financing of innovation: Learning and stopping," *RAND Journal of Economics*, pp. 719–752.

Carreyrou, J., 2018, *Bad Blood: Secrets and Lies in a Silicon Valley Startup.*

modeled as a flow cost.

Deb, R., M. Mitchell, and M. Pai, 2019, "(Bad) Reputation in Relational Contracting," *Mimeo.*

Dowie, M., 1977, "Pinto Madness," *Mother Jones*, September/October.

Ederer, F., R. Holden, and M. Meyer, 2018, "Gaming and strategic opacity in incentive provision," *The RAND Journal of Economics*, 49, 819–854.

Ely, J. C., and J. Välimäki, 2003, "Bad reputation," *The Quarterly Journal of Economics*, 118, 785–814.

Frankel, A., and N. Kartik, 2019, "Muddled information," *Journal of Political Economy*, 127, 1739–1776.

Fuchs, W., and A. Skrzypacz, 2015, "Government interventions in a dynamic market with adverse selection," *Journal of Economic Theory*, 158, 371–406.

Fuchs, W., and A. Skrzypacz, 2019, "Costs and benefits of dynamic trading in a lemons market," *Review of Economic Dynamics*, 33, 105 – 127.

Glazer, J., H. Herrera, and M. Perry, 2019, "Fake Reviews," *Mimeo.*

Gratton, G., R. Holden, and A. Kolotilin, 2018, "When to drop a bombshell," *The Review of Economic Studies*, 85, 2139–2172.

Klein, N., 2016, "The importance of being honest," *Theoretical Economics*, 11, 773–811.

Kolb, A., and E. Madsen, 2020, "Traitors in our Midst: Trust Dynamics with Secret Undermining," *Mimeo.*

Lawson, G., 2011, "The Stoner Arms Dealers: How Two American Kids Became Big-Time Weapons Traders," *Rolling Stone.*

Levitt, S. D., and C. Syverson, 2008, "Market Distortions When Agents Are Better Informed: The Value of Information in Real Estate Transactions," *The Review of Economics and Statistics*, 90, 599–611.

Li, J., N. Matouschek, and M. Powell, 2017, "Power dynamics in organizations," *American Economic Journal: Microeconomics*, 9, 217–41.

Mason, R., and J. Välimäki, 2015, "Getting it done: dynamic incentives to complete a project," *Journal of the European Economic Association*, 13, 62–97.

Perez-Richet, E., and V. Skreta, 2018, "Test Design Under Falsification," .

Shadmehr, M., and D. Bernhardt, 2015, "State censorship," *American Economic Journal: Microeconomics*, 7, 280–307.

Sobel, J., 1985, "A theory of credibility," *The Review of Economic Studies*, 52, 557–573.

Sun, P., and F. Tian, 2018, "Optimal contract to induce continued effort," *Management Science*, 64, 4193–4217.

Sun, Y., 2018, "A Dynamic Model of Censorship," *Mimeo*.

Varas, F., 2018, "Managerial short-termism, turnover policy, and the dynamics of incentives," *The Review of Financial Studies*, 31, 3409–3451.

A Proofs

Calculations for Footnote 5. Suppose that the real technology imposes flow cost c on the agent and the agent pays this flow cost until either (i) he has a real arrival, at rate λ, (ii) the game exogenously ends, at rate ρ, (iii) he submits a fake. Equation 3 becomes,

$$\begin{aligned} u_c(t) &= \int_0^t \lambda \exp(-(\rho+\lambda)s)(a(s)-cs) - \rho \exp(-(\rho+\lambda)s)cs \, \mathrm{d}s + \exp(-(\rho+\lambda)t)(a(t)-ct-\phi) \\ &= u(t) - c\left(\int_0^t (\lambda+\rho)\exp(-(\rho+\lambda)s)s \, \mathrm{d}s + \exp(-(\rho+\lambda)t)t\right), \end{aligned}$$

where $u(t)$ (as in (3)) is the agent's payoff function when $c = 0$. Note that

$$\int_0^t (\lambda+\rho)\exp(-(\rho+\lambda)s)s \, \mathrm{d}s + \exp(-(\rho+\lambda)t)t = K - \frac{\exp(-(\rho+\lambda)t)}{\lambda+\rho},$$

where K does not depend on t. Thus,

$$u_c(t) = \int_0^t \lambda \exp(-(\rho+\lambda)s)a(s)ds + \exp(-(\rho+\lambda)t)(a(t) - (\phi - \frac{c}{\lambda+\rho})) - cK.$$

Thus, the model with flow cost is strategically equivalent to the model without flow cost, with a smaller value of ϕ.

A.1 Proofs for Baseline Model

Proof of Lemma 4.1. We derive the probability distributions for the time of submission and the time of submission for a real project. If only the authentic technology is used, then the submission time for a *real* project is $T_R \sim H(t) \equiv 1 - \exp(-\lambda t)$. The waiting time for a fake project is $T_F \sim F(t)$. The overall waiting time for a submission is

$$T = (1-\sigma)T_R + \sigma \min\{T_R, T_F\}.$$

An ethical agent only uses the real technology, while a strategic agent uses the minimum of the real and fake waiting time. Therefore, the CDF of the arrival time is

$$\begin{aligned} W(t) = (1-\sigma)H(t) + \sigma(H(t) + F(t) - F(t)H(t)) &= H(t) + \sigma F(t)(1-H(t)) \\ &= 1 - \exp(-\lambda t)(1-\sigma F(t)), \end{aligned}$$

with associated density $w(t) = \exp(-\lambda t)[\lambda(1-\sigma F(t)) + \sigma f(t)]$. Using similar reasoning, the waiting time for a real arrival is distributed according to

$$\begin{aligned} W_A(t) &= (1-\sigma)H(t) + \sigma \int_0^t h(s)(1-F(s))\,\mathrm{d}s \\ &= (1-\sigma)(1-\exp(-\lambda t)) + \sigma \int_0^t \lambda \exp(-\lambda s)(1-F(s))\,\mathrm{d}s. \end{aligned}$$

If agent is ethical, a real project arrives before time t with probability $H(t)$. If agent is strategic, then a real project arrives before time t if a real arrival occurs at any $s \leq t$ and the fake arrival takes longer than s. Integrating over $s \leq t$ yields the expression. The density of the waiting time for a real arrival is therefore,

$$w_A(t) = \lambda \exp(-\lambda t)(1-\sigma F(t)).$$

By Bayes' Rule, the probability that a submission at time t is real is

$$g(t) = \frac{w_A(t)}{w(t)} = \frac{\lambda(1-\sigma F(t))}{\lambda(1-\sigma F(t)) + \sigma f(t)}.$$

Dividing numerator and denominator by $1 - \sigma F(t)$ yields the desired result. □

Proof of Lemma 4.2. We verify that the prescribed behavior is an equilibrium when $\phi > \widehat{\phi}$. Suppose $a(t) = 1$ for all $t \geq 0$. Substituting into (3) and evaluating gives $u'(t) > 0 \iff \phi > \widehat{\phi}$. Thus, the agent's best response to $a(\cdot) = 1$ is never to submit a fake. Furthermore, because the agent never fakes, we have $g(t) = 1$ for all $t \geq 0$. From (2), we have $a(\cdot) = 1$.

We verify uniqueness with the following steps.

Step 1. We show that the agent's mixed strategy cannot have a mass point. If such mass point exists at t, then $g(t) = a(t) = 0$. An elementary calculation implies that $u(t) < u(\infty)$, and thus faking at t is not a best response.

Step 2. We show that there exists some finite t^* such that $a(t) = 1$ and $f(t) = 0$ for $t \geq t^*$. Integrability implies $\lim_{t\to\infty} f(t) = 0$. Furthermore, $1 - \sigma F(t) \geq 1 - \sigma$, and hence, $\lim_{t\to\infty} \mu(t) = 0$. It follows that there exists t^* such that, $\mu(t) < \theta$ for $t > t^*$. From (2), we have $a(t) = 1$ for $t > t^*$. Substituting into (3) and differentiating at $t > t^*$, we find that $u'(t) > 0 \iff \phi > \widehat{\phi}$. Hence, $u(t) < u(\infty)$ for $t > t^*$. By implication, $f(t) = 0$ for such t.

Let $\bar{t}$ be $\sup\{t : f(t) > 0\}$. From Step 2, we have $\bar{t} < \infty$.

Step 3. We show that no equilibrium with faking exists. If $\bar{t} = 0$ and faking occurs in equilibrium, then $f(0) > 0$, contradicting Step 1. Therefore if such an equilibrium exists, then it must have $\bar{t} > 0$. Suppose this is the case. By definition of $\bar{t}$, for any small ϵ, there exists $t \in (\bar{t} - \epsilon, \bar{t}]$ such that $f(t) > 0$. For any such t we have that the agent's equilibrium payoff must be

$$u(t) = \int_0^t \lambda \exp(-(\rho + \lambda)s)a(s)\,\mathrm{d}s + \exp(-(\rho + \lambda)t)(a(t) - \phi).$$

This must be at least as large as

$$u(\infty) = \int_0^t \lambda \exp(-(\rho+\lambda)s)a(s)\,\mathrm{d}s + \int_t^{\bar{t}} \lambda \exp(-(\rho+\lambda)s)a(s)\,\mathrm{d}s + \int_{\bar{t}}^{\infty} \lambda \exp(-(\rho+\lambda)s)\,\mathrm{d}s.$$

Thus, for any choice of ϵ and a corresponding t, we have

$$\exp(-(\rho + \lambda)t)(a(t) - \phi) \geq \int_t^{\bar{t}} \lambda \exp(-(\rho + \lambda)s)a(s)\,\mathrm{d}s + \int_{\bar{t}}^{\infty} \lambda \exp(-(\rho + \lambda)s)\,\mathrm{d}s.$$

Since $\exp(-(\rho + \lambda)t)$ is decreasing, $a(\cdot) \in [0, 1]$, we have

$$\exp(-(\rho + \lambda)(\bar{t} - \epsilon))(1 - \phi) \geq \int_{\bar{t}}^{\infty} \lambda \exp(-(\rho + \lambda)s)\,\mathrm{d}s.$$

Integrating and simplifying, we have

$$\exp((\rho + \lambda)\epsilon))(1 - \phi) \geq (1 - \widehat{\phi}),$$

for any $\epsilon > 0$. Choosing $\epsilon \approx 0$ implies that $\phi < \widehat{\phi}$, contradicting the maintained hypothesis.

□

Proof of Lemma 4.3. We prove several steps which we then combine.

Step 1. We show that $a(t) \geq \phi$ for all $t \geq 0$. By way of contradiction, consider some $t \geq 0$ and suppose that $a(t) < \phi$. Because $a(t) - \phi < 0$, and $\int_t^\infty \lambda \exp\{-(\rho + \lambda)s\}a(s)ds \geq 0$, submitting a fake at time t is worse for the strategic agent than never submitting a fake, that is $u(t) < u(\infty)$. It follows that, $\mu(t) = 0$, which implies $g(t) = 1$, and hence $a(t) = 1$ by (2). Since we assumed $a(t) < \phi$, we find that $\phi > 1$, a contradiction.

Step 2. We show $f(t) = 0 \implies a(t) = 1$. This follows from $f(t) = 0 \implies \mu(t) = 0 \implies g(t) = 1 \implies a(t) = 1$, by (2).

Step 3. We show that if $a(t) = 1$, then $f(t') = 0$ and $a(t') = 1$ for all $t' > t$. Suppose $a(t) = 1$. Then for all $t' > t$ we have,

$$\begin{aligned} u(t') - u(t) &= \int_t^{t'} \lambda \exp(-(\rho+\lambda)s)a(s)\,\mathrm{d}s + \exp(-(\rho+\lambda)t')(a(t') - \phi) - \exp(-(\rho+\lambda)t)(1-\phi) \\ &\leq \int_t^{t'} \lambda \exp(-(\rho+\lambda)s)\,\mathrm{d}s + \exp(-(\rho+\lambda)t')(1-\phi) - \exp(-(\rho+\lambda)t)(1-\phi) \\ &= \left(\phi - \frac{\rho}{\rho+\lambda}\right)(\exp(-(\rho+\lambda)t) - \exp(-(\rho+\lambda)t') < 0, \end{aligned}$$

where the last inequality follows because $\phi < \widehat{\phi}$ and $t' > t$. Therefore, the agent receives a strictly higher payoff from submitting a fake at t than at any $t' > t$. This implies $f(t') = 0$ which implies $\mu(t') = 0$ which implies $g(t') = 1$ and by (2) $a(t') = 1$.

For the rest of the proof, let $\bar{t} \equiv \inf\{t : a(t) = 1\}$ or $\bar{t} = \infty$ if $a(t) < 1$ for all $t \geq 0$.

Step 4. We show that if $t < \bar{t}$, then $a(t) \in [\phi, 1)$ and $f(t) > 0$. That $a(t) < 1$ follows from Step 3 and the definition of $\bar{t}$. That $a(t) \geq \phi$ follows from Step 1. From the principal's indifference condition (2), we get $\mu(t) = \lambda(1-\theta)/\theta > 0$. Hence $f(t) > 0$.

Step 5. We show that $F(\cdot)$ has no mass point for any $t < \infty$. If $F(\cdot)$ has a mass point at t, then $\mu(t) = \infty$, and hence, $a(t) = 0$, which contradicts Step 1.

Step 6. We show that $\bar{t} \in (0, \infty)$. first suppose $\bar{t} = 0$. Then $a(t) = 1$ for all $t \geq 0$ by Step 3. Substituting into (3) and simplifying yields $u'(t) < 0 \iff \phi < \widehat{\phi}$. Hence, it is optimal for the agent to submit a fake with probability 1 at $t = 0$. From (2), $a(0) = 0$ is sequentially rational, a contradiction. Next, suppose that $\bar{t} = \infty$, i.e. for all t we have $f(t) > 0$. From Step 4, we have $a(t) \in (0, 1)$ for all t. Hence, $\mu(t) = \lambda(1-\theta)/\theta$ for all t. By implication, $f(t) = (1 - \sigma F(t))\lambda(1-\theta)/(\sigma\theta) \geq (1-\sigma)\lambda(1-\theta)/(\sigma\theta) > 0$. Thus the integral of $f(\cdot)$ is unbounded, a contradiction.

Proof of (i). This follows from Steps 3, 4, 5, and 6.
Proof of (ii). This follows from Steps 3 and 6.

Step 7. We show that $a(\cdot)$ is continuous at $\bar{t}$. Note that for $t > \bar{t}$ we have $\lim_{t \to \bar{t}} a(t) = a(\bar{t}) = 1$. We seek to show that for $t < \bar{t}$, we have $\lim_{t \to \bar{t}} a(t) = 1$. Consider $t < \bar{t}$. Because $f(t) > 0$, we must have $u(t) \geq u(\bar{t})$. Hence,

$$u(\bar{t}) - u(t) = \int_t^{\bar{t}} \lambda \exp(-(\rho+\lambda)s)a(s)\,\mathrm{d}s + \exp(-(\rho+\lambda)\bar{t})(1-\phi) - \exp(-(\rho+\lambda)t)(a(t) - \phi) \leq 0$$

Because $a(\cdot)$ is bounded, in the limit as $t \to \bar{t}$, we have $\lim_{t \to \bar{t}}\{u(\bar{t}) - u(t)\} = \exp(-(\rho + \lambda)\bar{t})(1 - \lim_{t \to \bar{t}} a(t)) \leq 0$. Because $a(t) \leq 1$ for all t, we have $\lim_{t \to \bar{t}} a(t) = 1$.

Step 8. We show that for $t < \bar{t}$, $a(t)$ is continuous, differentiable, and strictly increasing. Let $t, t' < \bar{t}$. Because $t, t' < \bar{t}$, from Claim (i) we have $f(t)\, f(t') > 0$. Hence, $u(t') = u(t)$. Therefore,

$$\begin{aligned}
0 &= u(t') - u(t) \\
&= \int_t^{t'} \lambda \exp(-(\rho+\lambda)s)a(s)\,\mathrm{d}s + \exp(-(\rho+\lambda)t')(a(t') - \phi) - \exp(-(\rho+\lambda)t)(a(t) - \phi) \\
&= \int_t^{t'} \lambda \exp(-(\rho+\lambda)s)a(s)\,\mathrm{d}t + [\exp(-(\rho+\lambda)t') - \exp(-(\rho+\lambda)t)](a(t') - \phi) \\
&\qquad + \exp(-(\rho+\lambda)t)(a(t') - a(t)).
\end{aligned}$$

Because the integrand above is bounded, taking the limit as $t' \to t$ yields $a(t') \to a(t)$. Hence, $a(t)$ is continuous.

To show that $a(t)$ is differentiable, divide the preceding equation by $t' - t$ to obtain,

$$\begin{aligned}
\frac{\int_t^{t'} \lambda \exp(-(\rho+\lambda)s)a(s)\,\mathrm{d}s}{t'-t} + \frac{\exp(-(\rho+\lambda)t') - \exp(-(\rho+\lambda)t)}{t'-t}(a(t') - \phi) \\
+ \exp(-(\rho+\lambda)t)\frac{a(t') - a(t)}{t'-t} = 0.
\end{aligned}$$

Because $a(\cdot)$ is continuous, the limit as $t' \to t$ gives

$$\lambda \exp(-(\rho+\lambda)t)a(t) - (\rho+\lambda)\exp(-(\rho+\lambda)t)(a(t) - \phi) + \exp(-(\rho+\lambda)t)\lim_{t'\to t}\frac{a(t') - a(t)}{t'-t} = 0.$$

Hence the derivative of $a(\cdot)$ exists at t. Furthermore, from this equation we have,

$$a'(t) = (\rho+\lambda)(a(t) - \phi) - \lambda a(t) = (\rho+\lambda)[a(t)\frac{\rho}{\rho+\lambda} - \phi] = (\rho+\lambda)[a(t)\widehat{\phi} - \phi].$$

It follows that $a(t)$ does not change monotonicity for any $t \in (0, \bar{t})$. It is either constant, strictly decreasing, or strictly decreasing. Suppose $a(t)$ is constant or strictly decreasing for all $t < \bar{t}$. It follows that $a(t) \le \phi/\widehat{\phi}$ for all $t < \bar{t}$. Because $\phi/\widehat{\phi} < 1$, we have $\lim_{t\to\bar{t}} a(t) < 1$, contradicting Step 7.

Proof of (iii). This follows from Steps 7 and 8. □

Proof of Proposition 4.1. Strategies. The agent must be indifferent about submitting at all times $t \in (0, \bar{t})$ and $a(\cdot)$ is differentiable. Therefore, for such t,

$$\begin{aligned}
u'(t) &= \exp(-(\rho+\lambda)t)\{\lambda a(t) - (\rho+\lambda)(a(t) - \phi) + a'(t)\} \\
&= \exp(-(\rho+\lambda)t)\{a'(t) - \rho a(t) + \phi(\rho+\lambda)\} = 0,
\end{aligned}$$

and hence, for $t \in [0, \bar{t}]$, we have $a(t) = \phi(1 + \frac{\lambda}{\rho}) + \kappa_1 \exp(\rho t)$ for some integration constant $\kappa_1 > 0$. Because $a(t) \in (0,1)$ for $t \in (0, \bar{t})$, we also have

$$g(t) = \theta \implies \mu(t) = \mu \implies F(t) = \frac{1}{\sigma}(1 - \kappa_2 \exp(-\mu t)),$$

where κ_2 is another integration constant. Note that the agent's mixed strategy cannot have a mass point, and hence $F(0) = 0$, which implies $\kappa_2 = 1$. It follows that

$$\bar{t} = -\ln(1-\sigma)\frac{\theta}{\lambda(1-\theta)}.$$

From the boundary condition $a(\bar{t}) = 1$ we find

$$\kappa_1 = (1 - \phi(1 + \frac{\lambda}{\rho}))\exp(-\rho\bar{t}) = (1 - \phi(1 + \frac{\lambda}{\rho}))(1-\sigma)^{\frac{\rho\theta}{\lambda(1-\theta)}}.$$

Observing that $\phi(1 + \frac{\lambda}{\rho}) = \frac{\phi}{\widehat{\phi}}$ completes the characterization of strategies.

Beliefs. Obvious.

Payoffs. **Strategic Agent.** The strategic agent's payoff is identical for all cheating times $t \in [0, \bar{t})$, and hence, $U^S = a(0) - \phi$. Simplifying, we have

$$U^S = \frac{\phi(\rho+\lambda)}{\rho} + (1 - \frac{\phi(\rho+\lambda)}{\rho})\exp\{-\rho\bar{t}\} - \phi = \frac{\phi\lambda}{\rho} + (1 - \phi - \frac{\phi\lambda}{\rho})(1-\sigma)^{\frac{\rho}{\mu}}.$$

Ethical Agent. Payoff of the ethical agent is

$$U^E = \int_0^\infty \lambda \exp(-(\rho+\lambda)t)a(t)\,\mathrm{d}t = \lim_{t\to\infty} u(t).$$

Recall that $u(t) = U^S$ on $[0, \bar{t}]$ and in particular $u(\bar{t}) = U^S$. Furthermore, since $a(t) = 1$ on $[\bar{t}, \infty)$, differentiation reveals that for $t \in (\bar{t}, \infty)$ we have

$$u'(t) = -(\widehat{\phi} - \phi)(\rho+\lambda)\exp(-(\rho+\lambda)t).$$

It follows that

$$U^E = \lim_{t\to\infty} u(t) = u(\bar{t}) + \int_{\bar{t}}^\infty u'(t)\,\mathrm{d}t = U^S - (\widehat{\phi} - \phi)\exp(-(\rho+\lambda)\bar{t}).$$

Principal. The principal is indifferent between accepting and rejecting for all $t < \bar{t}$, and therefore her expected payoff is 0 if an arrival occurs before time $\bar{t}$. Furthermore, if the agent is strategic, then an arrival will certainly occur before time $\bar{t}$. If the arrival occurs after $\bar{t}$, then it is real and will be accepted with probability 1. Hence,

$$V = (1-\theta)(1-\sigma)\int_{\bar{t}}^\infty \lambda\exp\{-(\rho+\lambda)t\}\,\mathrm{d}t = (1-\theta)(1-\sigma)\frac{\lambda}{\rho+\lambda}\exp\{-(\rho+\lambda)\bar{t}\}$$

$\square$

A.2 Proofs for Opaque Standards

Proof of Lemma 5.1. Proof of (i). This point follows exactly from the arguments in the proof of Lemma 4.3, 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, replacing $a(\cdot)$ with $a_U(\cdot)$.

By analogy with the proof of Lemma 4.3, 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.

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,

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

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

$$\begin{aligned} \nu a_H(t') + (1 - \nu) a_L(t') &> \nu a_H(t) + (1 - \nu) \\ \nu (a_H(t') - a_H(t)) &> (1 - \nu)(1 - a_L(t')) \geq 0, \end{aligned}$$

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, (A.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 $a_L(\cdot) \leq 1$ and $a_H(\cdot) \geq 0$ the result follows.

Proof of (iv). Follows from Steps 5-7. $\square$

Proof of Proposition 5.1. We construct a one phase equilibrium, consistent with Lemma 5.1, 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 5.1 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 5.1 to rule out a mass point in the agent's mixed strategy, thereby identifying an integration constant. Using point (i) of Lemma 5.1, 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 5.1 (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 4.1. 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

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

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 5.2. We construct a two stage equilibrium, consistent with Lemma 5.1, showing that such an equilibrium exists if and only if $\nu < (1-\frac{\phi}{\hat{\phi}})(1-\exp(-\rho\bar{t}_H))$, 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 5.1 with $\widetilde{t}_U > 0$.

Strategies and Phase Transitions. From Lemma 5.1, 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

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

From Lemma 5.1, 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 5.1. 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 5.1.

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

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

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

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 $\nu < (1 - \frac{\phi}{\hat{\phi}})(1 - \exp\{-\rho \bar{t}_H\})$, then $\bar{t}_U < \bar{t}_H$. From claim 1, for such ν we have

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

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

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

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

Beliefs. Follows immediately from Lemma 5.1 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

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

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

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

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 5.2, 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 5.1, 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,

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

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

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

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.

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

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

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

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.3 Proofs for Logjam

Proof of Lemma 6.1. 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 $t' \in (t - \delta', t + \delta')$,

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

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

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

It follows that $u(t) - u(t') \leq 3\epsilon - \overline{\epsilon} = -\frac{\overline{\epsilon}}{4}$, 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 $0 \le \widetilde{t}_J \le t^* \le \overline{t}_J \le \infty$ 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

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

Suppose $t' \in (\widetilde{t}_J, \overline{t}_J)$ and $a(t') = \overline{a}(t')$. We consider two cases. (1) If $\kappa \le 0$, then $a(t)$ is weakly decreasing. Thus, $a(t') = \overline{a}(t')$ implies $a(t'') \ge a(t')$ 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 $(\widetilde{t}_J, \overline{t}_J)$. Recall that $\overline{a}(\cdot)$ is a strictly increasing concave function. If $a'(t') < \overline{a}'(t')$, then the constraint $a(\cdot) \le \overline{a}(\cdot)$ 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 $t' \in (\widetilde{t}_J, \overline{t}_J) \Rightarrow a(t') < \overline{a}(t')$. 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 $\overline{t}_J < \infty$ then $a(t) = \overline{a}(t)$ for $t \in [\overline{t}_J, \infty)$. From Step 8, we know that $u(t) < u^*$ on intervals $[0, \widetilde{t}_J)$ and $(\overline{t}_J, \infty)$. Thus, $f(t) = 0$ on such intervals, and consequently, $a(\cdot) = \overline{a}(\cdot)$. By continuity of $a(\cdot)$ at $\widetilde{t}_J$, $\overline{t}_J$, we have $a(\overline{t}_i) = \overline{a}(\overline{t}_i)$ for $i \in \{H, L\}$.

Step 11. We show that $\overline{t}_J < \infty$. Suppose that $\overline{t}_J = \infty$. From Step 9, we have $a(t) < \overline{a}(t)$ for $t \in (\widetilde{t}_J, \overline{t}_J)$. 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 $\widetilde{t}_J < \overline{t}_J$. Suppose $\widetilde{t}_J = \overline{t}_J$. Combined with the conditions $\widetilde{t}_J \leq t^* \leq \overline{t}_J$, we have $\widetilde{t}_J = t^* = \overline{t}_J$. 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 $\widetilde{t}_J < \overline{t}_J$. 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 $\overline{t}_J > 0$ or $\overline{t}_J = 0$, we have $a(\overline{t}_J) \leq \overline{a}(\overline{t}_J) < \overline{a}(\overline{t}_J)$, 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 $\overline{a}(\overline{t}_J) = a(\overline{t}_J)$ for some $\overline{t}_J \in (\widetilde{t}_J, \infty)$. Thus, we have shown that $a(\widetilde{t}_J) < a(\overline{t}_J)$. Since the monotonicity of $a(\cdot)$ does not change on $(\widetilde{t}_J, \overline{t}_J)$, 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 6.1. Strategies. Based on Lemma 6.1, 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

$$
\begin{aligned}
\frac{\phi}{\hat{\phi}} + \kappa\exp(\rho\widetilde{t}_J) &= 1 - \frac{\rho}{\rho+\gamma}\exp(-\gamma\widetilde{t}_J) \\
\frac{\phi}{\hat{\phi}} + \kappa\exp(\rho(\widetilde{t}_J+\overline{t})) &= 1 - \frac{\rho}{\rho+\gamma}\exp(-\gamma(\widetilde{t}_J+\overline{t})).
\end{aligned}
$$

Solving, we have

$$
\begin{aligned}
\kappa &= \exp(-\rho\widetilde{t}_J)(\overline{a}(\widetilde{t}_J) - \frac{\phi}{\widehat{\phi}}), \\
\exp(-\gamma\widetilde{t}_J) &= (1-\frac{\phi}{\widehat{\phi}})(1+\frac{\gamma}{\rho})\frac{\exp(\rho\overline{t})-1}{\exp(\rho\overline{t})-\exp(-\gamma\overline{t})}.
\end{aligned}
$$

Obviously, $\widetilde{t}_J > 0$ and $\overline{t}_J < \infty$. That $\widetilde{t}_J < t^* < \overline{t}_J$ follows from a straightforward application of L'Hopital's rule. Note that, if we solve the second equation for κ, we find $\kappa = \exp(-\rho\overline{t}_J)(\overline{a}(\overline{t}_J) - \frac{\phi}{\widehat{\phi}})$, 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 $[\widetilde{t}_J, \overline{t}_J]$, 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 $(\widetilde{t}_J, \overline{t}_J)$ 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 $t > \overline{t}_J$ only if the agent is ethical.

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

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

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 $\overline{t} \to 0$. Using L'Hopital's rule,

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

which in turn implies that $\widetilde{t}_J \to t^*$, $\overline{t}_J \to t^*$. 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. $\square$