Make It 'Til You Fake It*

Raphael Boleslavsky
Indiana University
rabole@iu.edu

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

January 29, 2024

Abstract

We study the dynamics of fraud and trust in a continuous-time reputation game. The principal wishes to approve a real project and reject a fake. The agent is either an ethical type that produces a real project, or a strategic type that also can produce a fake. Producing a real project takes an uncertain amount of time, while a fake can be created instantaneously at some cost. The unique equilibrium features an initial phase of doubt, during which the strategic agent randomly fakes and the principal randomly approves. Only submissions that arrive after the phase of doubt are beneficial to the principal. We investigate three variants of the model that mitigate this problem. With full commitment, the principal incentivizes the strategic agent to fake at one specific time and commits to approve. Though the principal knows that earlier arrivals are real, she commits to reject them with positive probability. When she can delegate authority, the principal benefits by transferring it to a proxy who is more cautious than she is. When the principal can conduct a costly test prior to her approval decision, the principal also benefits. The equilibrium testing probability may be non-monotonic over time, first increasing, then decreasing. JEL Classifications: C73, D21, D82, L15, M42.

Keywords: fraud, trust, time pressure, delegation, audit.

---

*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, University of Texas Austin (Economics and Finance), and the 34th Stony Brook International Conference on Game Theory. Boleslavsky thanks the Weimer Faculty Fellowship at Indiana University for financial support. Taylor thanks the University of Colorado where he was on sabbatical when some of this work was performed.

1 Introduction

The pursuit of short term advantage can lead to misconduct in numerous settings where malfeasance is hard to detect; examples are ubiquitous.

- In order to close 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 rushed to produce the Pinto, skimping on safety testing, with disastrous results (Dowie, 1977).
- 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 risk losing the slot.

In this paper we present and analyze a dynamic principal-agent model of fraud and trust. The principal has limited commitment power and wishes to approve a *real* project and reject a *fake* one. 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 differ in their willingness or capacity to commit fraud. An ethical type is unwilling or unable to produce a fake, while a strategic type would do so immediately if he believed it would be approved. In our baseline model, real projects and fakes are indistinguishable by the principal. All she observes is the time at which a project is proffered when deciding whether to approve or reject it. Thus, the time of project arrival plays a critical role in the principal's approval decision, allowing her to update her beliefs about the agent's type and the project's authenticity.

In the unique equilibrium of the baseline model, 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 generates 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. Indeed, the strategic type submits a fake with probability 1 by some finite date, at which point the phase of doubt ends. Thereafter, the principal is fully confident that the agent is ethical and she approves any submission with probability 1.

Because the principal is indifferent between approving and rejecting projects that are submitted during the phase of doubt, her expected payoff equals 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 the agent actually is ethical and he does not produce a real project sooner, the principal's equilibrium payoff is relatively low in the baseline model. The rest of the paper, therefore, is dedicated to investigating common methods principals utilize to improve their welfare.

First, we analyze a benchmark setting in which the principal is endowed with commitment power. Specifically, we suppose the principal can optimize jointly over the agent's faking strategy and her own approval strategy, subject to incentive compatibility constraints that ensure the agent's compliance (in the absence of transfers). Although the analysis is somewhat involved, the solution turns out to be quite intuitive. Similar to the equilibrium structure of the baseline model, the horizon is partitioned into two intervals. During the initial "phase of doubt," the strategic agent is indifferent about submitting a fake, but refrains from doing so, and the principal approves submissions with increasing probability. At the end point of this interval, the strategic agent stops searching for a real breakthrough and submits a fake, which the principal commits to accept. Thereafter the phase of credibility ensues in which only the ethical type of agent is active and the principal accepts any submission with certainty. The solution, therefore, involves two strong types of commitment: the principal must reject with positive probability submissions during the initial phase that she *knows* to be real and she must accept with probability 1 a submission at the end of the initial phase that she *knows* to be fake. Although the principal benefits ex ante from commitment de facto, the degree of fidelity she is required to possess once projects of known quality are in front of her seems implausible in many – perhaps most – settings. For this reason, we explore two other potential improvements.

When committing to reject real projects and accept fakes is not feasible, the principal may nevertheless be able to commit to transfer her approval authority to another individual whose preferences differ from her own, whom we refer to as an "evaluator." In this context we present three main findings. First, when the principal can only delegate to an evaluator whose tolerance for fraud is close to her own, she benefits from delegation to a more cautious evaluator. Thus, even a very limited form of delegation is beneficial. Second, when the probability of a strategic agent is high, it is optimal for the principal to delegate to *the most* cautious evaluator that she can find. Third, we show that this finding is overturned when the probability of an ethical agent is high. In this case, it is optimal for the principal to delegate to an evaluator who is more cautious, but it is *not* optimal to delegate to the most cautious evaluator; in fact, delegating authority to an evaluator who is overly cautious is worse for the principal than keeping the authority herself.

Finally, we investigate a setting where the principal possesses a technology for auditing submissions, but she cannot commit to use it. The audit technology is costly and noisy: though it reduces the agent's informational advantage, it does not eliminate it. The dynamic structure of the equilibrium is especially intriguing when the auditing technology is relatively weak and the principal's prior that the agent is ethical is low. In this case, the phase of doubt is partitioned into two stages, and auditing plays a different role in each. In the initial stage, the principal randomizes between auditing and rejection. Here, an audit is used to identify the rare real project which merits approval. As time passes without a submission and the principal's trust in the agent correspondingly grows, she becomes more-inclined to audit his project. When trust grows sufficiently, the equilibrium transitions to the second stage where the principal randomizes between outright approval and auditing. Here, the audit serves to screen out fakes, and as the principal's trust in the agent continues to grow, she has less incentive to check his work. Thus, the equilibrium audit probability is non-monotonic over the phase of doubt. It begins at a low level, increases until it reaches one at the transition point between the two stages, and then decreases until it reaches zero at the end of the phase of doubt. Past this time, the principal is fully confident that the agent is ethical, and she approves any submission with no audit. We show that when the auditing technology is cheap enough to be used in equilibrium, the principal's payoff is higher than in the baseline model, and *both* types of agent can also be better off.[1]

The possibility of an ethical agent and the resulting growth of trust are crucial for our analysis. In the main model, these features generate a finite phase of doubt, which is necessary and sufficient for the principal to benefit from the relationship. In other words, if the agent were known to be strategic from the outset, the resulting equilibrium would be stationary, the phase of doubt infinite, and the principal's expected payoff zero. Thus, without some

[1] In an earlier draft we investigated two other methods for improving the principal's equilibrium payoff: opaque standards (i.e., uncertainty by the agent regarding the principal's preferences) and logjams (i.e., exogenous delays in the approval process). Analysis of these remedies has been omitted from the current draft to save space, but is available upon request from the authors. Opaque standards are related to Ederer, Holden, and Meyer (2018) who study opacity in a multi-task moral hazard model. The logjam 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.

probability of an ethical agent, the principal gains nothing from the production process (a related point is made by Strulovici (2020) in a different context). In the delegation extension, selecting a more cautious evaluator is costly only if trust evolves over time. Without this force, the principal always prefers to delegate to the most cautious evaluator. Finally, one of the most interesting findings of the auditing extension is the link between the growth of trust and the dynamics of auditing—the growth of trust increases the probability of auditing when trust is low and decreases it when trust is high. In a model with a known strategic agent the evolution of trust is absent, the equilibrium is stationary, and these novel dynamics do not obtain. In addition, when trust evolves, the principal always benefits when an auditing technology is available (provided it is not too expensive). This result also hinges on the evolution of trust; without it, the principal only gains if the test is sufficiently strong.

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. Perez-Richet and Skreta (2018) consider the design of an optimal test when the agent has the ability to manipulate the process by which the test determines his type.

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 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 (forthcoming) 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.

3 Model

A principal (she/her) interacts with an agent (he/him) over an indefinite horizon. Time is continuous and both parties discount future payoffs at rate $\rho > 0$. 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 but fast.

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. 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, we are referring to the strategic type.

Once a project (of either type) has been developed, it is instantly submitted to the principal for approval. 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 $\theta \in (0,1)$ thus represents the principal's tradeoff between type I and type II errors. The strategic agent would like his project to be approved regardless of its state, obtaining a gross benefit of 1 from approval, and 0 from rejection.

Discussion of Assumptions. Before moving to the analysis, we discuss some of our modeling choices. This discussion is self-contained and is not essential for understanding the rest of the paper.

Cost structure. Producing a fake is assumed to impose an immediate fixed cost on the agent, regardless of the principal's ultimate approval decision. This cost could be a resource (e.g. effort) or financial cost, but it could also incorporate the expectation of future punishment if fraud is eventually detected. In this vein, it is also possible to include a second "ex post" cost of submitting a fake, which is sustained only if the project is approved. If this ex post cost is not so large as to deter faking, then the results are similar to those presented below. In particular, only the closed form of the principal's approval strategy is affected.

On the other hand, employing the authentic technology is assumed to be free for either type of agent. If it involved a positive flow cost, then the strategic agent could produce a fake in order to avoid the cost of authentic project development. Our goal is to study the incentive to commit *fraud*, motivated by the desire for short term gain, which has received relatively little formal analysis, as opposed to the incentive to *shirk,* which has been researched extensively. Nevertheless, 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 formally as a reduction in ϕ (calculation in Appendix).

Common Discounting. Although we assume ρ is the same for principal and agent, nearly all results generalize easily to disparate rates of time preference. The most natural interpretation for a common rate is that the relationship ends exogenously according to a Poisson arrival; e.g., a technological innovation that renders the project obsolete.

Instant Arrivals. We focus on a setting in which projects are submitted instantly once they are developed, regardless of the underlying technology. This assumption rules out equilibria in which the agent is 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 a submission strategically.

Behavioral Type. The assumption that the ethical agent cannot use the fake technology is made solely for expositional convenience. It is straightforward to construct an outcome equivalent variant of the baseline model in which the ethical type can employ the fraudulent technology but chooses not to because – relative to the strategic type – he is more patient, has a smaller payoff when a fake is approved, or has a larger cost of producing a fake.

Poisson Arrival. Development of a real project is modeled as a Poisson arrival. This assumption simplifies some of the calculations and extensions, but it is not essential for the main characterization or welfare results. The equilibrium is qualitatively similar for alternative arrival processes, so long as the strategic agent prefers to fake immediately if he anticipates certain approval. For example, this holds if the real project arrival follows a distribution supported on $\mathbb{R}_+$ with decreasing hazard rate $H(\cdot)$ and $\phi < \rho/(\rho + H(0))$. In this case, the equilibrium structure is identical to that specified below in Lemma 4.3. Moreover, modified versions of equations (4,5) continue to hold resulting in an equilibrium characterization similar to Proposition 4.1 below.

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.[2] 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 beliefs are derived from Bayes' rule.[3]

[2]Multiple equilibria exist iff $\phi = \frac{\rho}{\rho+\lambda}$. To streamline presentation, we suppress this and all other non-generic knife-edge cases throughout.

[3]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.

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 produce a fake submission 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 $F(\cdot)$.[4] 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 submission at t given a submission 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*

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

where

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

The function $\nu(\cdot)$ 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 $\nu(\cdot)$ 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; for this reason we refer to $\nu(\cdot)$ as the "perceived rate of faking." 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 the ratio of the rate of a real arrival to the sum of the rate of a real arrival and the perceived rate of a 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

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

[4] 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.

Remark 1. *(Delayed Approval.) We have assumed that the principal makes a decision whether to accept or reject the agent's submission as soon as it arrives. Thus, the principal's strategy is a probability* $a(t)$ *with which she approves a project that is submitted at time* t*. We could allow for delays in the principal's approval decision with no changes to the results. In particular, suppose that the principal chooses a delay* $d(t) \geq 0$*, which specifies how long she will wait to approve a project that is submitted at time* t*, with* $d(t) = \infty$ *corresponding to rejection. Allowing the principal to randomize over such pure strategies, she chooses random variables* $D(t) \geq 0$ *for* $t \geq 0$*, representing the stochastic delay that is imposed before a project submitted at time* t *is approved. This is a direct extension of the current formulation, which considers strategies of the form* $\Pr(D(t) = 0) = a(t)$ *and* $\Pr(D(t) = \infty) = 1 - a(t)$*. Nevertheless, focusing on our formulation is without loss. To see the equivalence, consider random variable* $D(t)$*. Note that, following a submission at time* t*, the agent's expected continuation payoff is* $E[\exp(-\rho D(t))]$*, and the principal's is* $E[\exp(-\rho D(t))](g(t) - \theta)$*. Thus, all delay random variables* $D(t)$ *that generate the same* $E[\exp(-\rho D(t))]$ *are payoff equivalent for principal and agent. In particular, for any delay random variable* $D(t)$*, a strategy in our class with acceptance probability* $a(t) = E[\exp(-\rho D(t))]$ *is payoff equivalent for both players. Confining attention to pure delay strategies, the preceding argument implies delay function* $d(t) = -\ln(a(t))/\rho$ *and approval strategy* $a(t)$ *are payoff-equivalent for principal and agent. As we show below,* $a(t)$ *is increasing in equilibrium, so* $d(t)$ *is decreasing. In other words, the earlier a project is submitted, the longer the principal waits prior to acceptance.*

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 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 with probability 1. Motivated by Lemma 4.2, we maintain the following assumption (without restatement) in what follows.

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* $\bar{t} \in (0, \infty)$.

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

(iii) 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. If the agent's strategy had an atom, the principal would reject an arrival at the atom. If it had a gap, the principal would approve any arrival in the gap. In both cases, the agent has a profitable deviation.

Building on the preceding 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, $\nu(t)$, approaches 0 as time passes, regardless of the agent's strategy.[5] 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. In other words, because the probability of the ethical type is positive, the strategic type cannot keep the principal's belief at (or below) her point of indifference (θ) indefinitely. Intuitively, as time passes without an arrival, the principal's belief that the agent is ethical grows. In order to maintain belief $g(t) = \theta$ about a submitted project over time, the strategic type must fake more and more aggressively, reinforcing the inference that the agent is ethical in the absence of an arrival. Eventually, the agent must fake so aggressively that a non-arrival reveals him to be ethical, and $g(t) = 1$ for any subsequent arrival.

Because the strategic agent mixes over the phase of doubt in equilibrium, his payoff $u(\cdot)$ 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, $\bar{t}$. Indeed, once the phase of doubt ends, the principal knows that the agent is ethical. By implication, for times near $\bar{t}$, the principal must approve with probability approaching one; otherwise, the agent could benefit by waiting until just after $\bar{t}$ to fake.

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 succinctly, define

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

which is the perceived rate of cheating in equilibrium (see (4)).

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

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*

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

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 a submission 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 if $t \geq \bar{t}$*.*

Beliefs. *If* $t \in (0, \bar{t})$*, then* $g(t) = \theta$*, 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*

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

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 strategy of a particular functional form: a constant plus an exponential function with growth rate ρ, the discount rate.

To understand the shape of the approval strategy, it is helpful to consider the differential equation for $a(\cdot)$, given in (5). This equation admits a particular solution with a constant acceptance function $a(t) = \phi/\widehat{\phi}$ and a complementary solution $a(t) = \exp(\rho t)$. 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 (5) 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

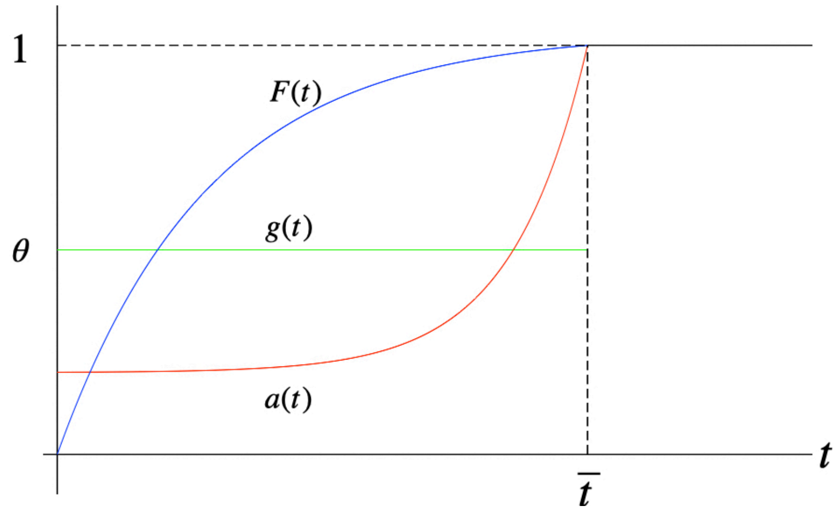

Figure 1: *Proposition 4.1.*

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. Furthermore, in order to satisfy the boundary condition $a(\bar{t}) = 1$, the weight on the complementary solution (the exponential) must be strictly positive. Intuitively, we find that the probability of acceptance is directly linked to the belief about the agent's type. If the agent were known to be strategic with probability 1, then the game would be stationary and $a(t) = \phi/\widehat{\phi}$ in the unique equilibrium. With a positive probability of an ethical agent, the acceptance probability exceeds $\phi/\widehat{\phi}$ at time 0, and it increases continuously as the principal's trust grows. At the boundary $\bar{t}$, the principal's belief that the agent is ethical and her acceptance probability attain 1 simultaneously; the weight on the complementary solution ensures this equality.

What about the agent's strategy? It is sequentially rational for the principal to mix over the phase of doubt if and only if her belief that a submission is real is $g(t) = \theta$, which implies that the perceived rate of faking is constant, as in (4). Indeed, because the authentic technology has a constant arrival rate, a constant perceived rate of faking ensures that the principal does not learn about the state of a project from observing the time at which it was submitted. Solving (4), we find that the only candidates 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. It is worth pointing out that the principal's perceived rate of faking is constant in the phase of doubt, $\nu(t) = \mu$, the strategic agent's actual faking rate $f(t)/(1 - F(t))$ increases over time. As the principal's trust in the agent grows, the strategic type takes advantage by committing fraud at a higher rate, approaching infinity at $\bar{t}$.

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 strategy 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.

To understand the principal's payoff and the normative implications of fraud, 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 ($t \geq \bar{t}$), 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 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 no 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 the principal.

5 Commitment Power

In this section we investigate a benchmark in which the principal can commit to her approval strategy $a(\cdot)$. As will become clear, implementation of the optimal approval strategy requires very strong commitment power on the part of the principal, which may be implausible in some settings. Nevertheless, it is instructive to highlight the differences between the commitment solution and the equilibrium strategy derived in the preceding section. Additionally, understanding the limitations associated with full commitment sets the stage for us to consider other methods for improving the principal's payoff vis-à-vis the equilibrium outcome.

To derive the commitment solution, we adopt the standard approach of the principal-agent paradigm without transfers, assuming that the principal designs her approval strategy $a(\cdot)$ and recommends a faking strategy to the agent. The agent complies with the principal's recommendation, provided that doing so is incentive compatible. In other words, (almost) all realizations τ in the support of the recommended faking strategy must be optimal; $u(\tau) \geq u(t)$ for all $t \geq 0$ ($u(t)$ is the agent's payoff of faking at t, given in (3)).

Remark 2. *(Implementation via Delayed Approval) Though we focus on the design of the principal's random approval strategy $a(\cdot)$, following the logic of Remark 1, our analysis applies equally to the design of an optimal deterministic delay function, $d(\cdot)$, which specifies the delay imposed on a project that is submitted at time t, before it is approved. This implementation is somewhat more plausible than random approval because it is easier to verify the deterministic delay $d(t)$ than the randomization $a(t)$.*

Next, we argue that it is without loss of generality to focus on deterministic faking strategies. Consider an optimal approval strategy for the principal, $a(\cdot)$, and define the set of optimal faking times for the strategic agent by $S = \{\tau | u(\tau) \geq u(t) \text{ for all } t \geq 0\}$. Note that any faking strategy supported on S is incentive compatible. Consider the subset of faking times $S' \subseteq S$ that are optimal for the principal. If S' is a singleton, τ, then it is optimal for the principal to recommend that the agent fakes at time τ. If S' is not a singleton, then both the principal and agent are evidently indifferent over all mixtures supported on S', including a degenerate one that puts probability 1 on any $\tau \in S'$. Thus, it is without loss of generality to focus on an equilibrium in which the principal recommends a single incentive compatible faking time to the agent, denoted τ.

Building on the preceding observations, the principal's design problem is

$$\max_{a(\cdot),\tau} (1-\theta) \int_0^\tau \lambda \exp\{-(\lambda+\rho)t\} a(t) dt + (1-\theta)(1-\sigma) \int_\tau^\infty \lambda \exp\{-(\lambda+\rho)t\} a(t) dt \\ -\theta\sigma \exp\{-(\lambda+\rho)\tau\} a(\tau)$$

subject to (IC), $u(t) \leq u(\tau)$ for all $t \geq 0$ and (F), $a(t) \in [0,1]$ for all $t \geq 0$. As is customary in the contracting literature (e.g. Guesnerie and Laffont (1984)) we focus on absolutely continuous $a(\cdot)$.[6]

The first term of the principal's objective represents the expected benefit generated by a real arrival (from either type of agent) that occurs before the recommended faking time,

[6]Such an assumption is standard in dynamic optimization problems. For example, one could formulate the principal's problem as a control problem. A natural formulation has $u(\cdot)$ as a state variable. The law of motion for $u(\cdot)$ involves both $a(\cdot)$ and $a'(\cdot)$, and thus it is natural to treat $a'(\cdot)$ as the control and $a(\cdot)$ as a state variable. Thus, $a'(\cdot)$ must integrate to $a(\cdot)$, and hence $a(\cdot)$ is absolutely continuous.

the second term represents the benefit generated by a real arrival that occurs after the recommended faking time (which is obtained only if the agent is ethical), and the third term is the cost of approving a fake submission with probability $a(\tau)$ at the recommended faking time. The principal also faces an incentive constraint (IC) which requires that the agent's payoff of faking at the recommended time is at least as large as his payoff of faking at any other time. The principal also faces a feasibility constraint (F), which requires that $a(t)$ is a probability.

This formulation, in and of itself, hints at the value of commitment for the principal. At the recommended (incentive compatible) faking time, the principal commits to approve a fraudulent project with probability $a(\tau)$, which is costly if this time is reached and the agent is strategic. By providing incentives for the agent to fake only at this time, the principal ensures that projects submitted at all other times are real, and she would like to approve them with probability 1. However, incentive compatibility generally requires the principal to approve such projects with lower probability. Thus, the principal leverages commitment power in both directions—sometimes approving a project she knows is fraudulent, and sometimes rejecting a project that she knows is real. Her design problem optimally trades off these costs and benefits; qualitatively, an earlier faking time τ or a higher probability of approval $a(\tau)$ increases the direct cost, but it also relaxes the agent's incentive to fake at other times and allows real projects to be approved with higher probability.

We analyze this problem in detail in the online Appendix; here we focus on the key results. First, we find that with commitment, the structure of the principal's optimal approval strategy is similar to the equilibrium strategy without commitment. In particular, with commitment, the agent's incentive constraint binds on an interval of early times and is slack otherwise. In other words, the agent is indifferent between faking at all times in an initial interval, and strictly prefers any time in this interval to a time outside this interval; a $T \geq 0$ exists (possibly infinite) such that $u(t) = u^*$ for $t \in [0, T]$, and $u(t) < u^*$ for $t > T$, where u^* is the agent's payoff under principal commitment. One immediate implication is that the recommended faking time satisfies $\tau \in [0, T]$. Furthermore, for each T, the approval strategy is determined by the agent's indifference, as in (5). Thus, the optimal approval strategy is

$$a(t) = \frac{\phi}{\widehat{\phi}} + (1 - \frac{\phi}{\widehat{\phi}}) \exp(-\rho(T - t)),$$

for some T. Substituting into the principal's objective, it is straightforward to establish that among the incentive compatible faking times $\tau \in [0, T]$, it is optimal to recommend $\tau = T$, thereby delaying faking as long as possible. Incorporating this finding, we show that the principal's objective function is single-peaked in τ, the optimal τ is interior, and it is found in closed form.

Proposition 5.1 (Commitment)**.** *With commitment, the optimal faking time is*

$$\tau = -\frac{\ln\left(\frac{(1-\theta)(\widehat{\phi}-\phi)}{(1-\theta)(\widehat{\phi}-\phi)+\sigma(1-\widehat{\phi}+\widehat{\phi}\theta)}\right)}{\lambda}.$$

The optimal approval strategy is

$$a^*(t) = \frac{\phi}{\widehat{\phi}} + (1-\frac{\phi}{\widehat{\phi}})\exp(-\rho(\tau - t)),$$

for $t \leq \tau$, and $a^*(t) = 1$ for $t > \tau$. The strategic agent's payoff with principal commitment is $a^*(0) - \phi$. Incentive compatibility binds for $t < \tau$, and it is slack otherwise; $u(t) = u(\tau)$ for $t \in [0, \tau]$ and $u(t) < u(\tau)$ for $t > \tau$.

The proof of this result is rather involved and constitutes the entire online appendix.

The structure of the commitment solution closely resembles the equilibrium structure of the baseline model, with a "phase of doubt" $t \in [0, \tau)$ where $a(t) < 1$, followed by a phase of credibility, where $a(t) = 1$. However, two crucial differences emerge. First, all faking occurs at time τ—as described previously, the principal leverages her commitment power by approving a fake at time τ with probability 1, and rejecting real projects that arrive before τ with positive probability. Hence, there is no "doubt" in the phase of doubt under commitment because such submissions are never fakes. Second, the duration of the phase of doubt is different with commitment. Without commitment, the duration of this phase is pinned down by the evolution of the agent's credibility. Given that he produces fakes at perceived rate $\nu(t) = \mu$, it takes exactly time $\bar{t}$ for his credibility to be fully established. In contrast, with commitment, the duration of the phase of doubt, τ, is determined by the interaction of the IC constraint and the principal's objective. It therefore depends on parameters that are absent from $\bar{t}$: for example, ϕ, which determines how tempting cheating is for the agent and, indirectly, how big of a distortion is needed for incentive compatibility.

Building on the previous observations, τ can be either larger or smaller than $\bar{t}$. For example, when ϕ is sufficiently close to $\widehat{\phi}$, the optimal recommended faking time is arbitrarily large, but $\bar{t}$ is finite for $\sigma < 1$. In this case faking is not very tempting for the agent, so the distortions arising from IC are relatively small. Thus, the principal can maintain a high approval probability at times before τ, and it is worthwhile to delay faking as much as possible. Conversely, for $\phi < \widehat{\phi}$, the optimal recommended faking time τ is finite for all σ, but $\bar{t}$ is arbitrarily large for σ close to 1. Taken together, these observations imply that commitment may either extend or shorten the phase of doubt. Comparing each type of agent's payoff when the principal can commit to the case where she cannot, the only difference is the duration of the phase of doubt. Because commitment may extend or shorten this phase, the agent may benefit or suffer when the principal can commit.

As noted earlier, the principal's commitment power is quite demanding: not only does the principal commit to accept a project she knows is fake (at time τ), but she also commits to reject a project that she knows to be real with positive probability (before time τ). Because the agent (of either type) also prefers that such projects are accepted, a positive probability of rejection is not renegotiation proof. Even if she lacks such strong commitment power, the principal may yet possess other methods for improving her expected payoff relative to the equilibrium of the baseline model. We explore two such instruments in the ensuing sections, transfer of approval authority to a proxy with non-congruent preferences and randomly auditing submissions.

6 Delegation

When fully committing to the optimal approval policy is not feasible, project sponsors may still be able to exploit institutional features to generate some degree of commitment. In this section we explore one common practice, delegation of approval authority to a proxy. Thus, consider a setting where the principal has the capability to delegate the acceptance decision to an "evaluator" with different preferences than her own. In particular, the principal assigns decision rights to a proxy whose preference parameter is $\widetilde{\theta}$ rather than θ, and the transfer of authority is observed directly by the agent who also knows $\widetilde{\theta}$.

The equilibrium of the game between agent and evaluator is the same as the one characterized in Proposition 4.1, with $\widetilde{\theta}$ replacing θ. The principal's payoff from delegating authority at the beginning of the game is therefore

$$V_D(\widetilde{\theta}|\theta) = \int_0^{\bar{t}_D} \exp(-\rho t) w_D(t)(g_D(t) - \theta) a_D(t) dt + (1-\theta)(1-\sigma) \int_{\bar{t}_D}^{\infty} \lambda \exp(-(\rho+\lambda)t) dt,$$

where the subscript D denotes the equilibrium of the delegation subgame (note that $V_D(\theta|\theta) = V$, see (9)). From Proposition 4.1 we have $g_D(t) = \widetilde{\theta}$ and

$$(1-\sigma F_D(t)) = \exp(-\lambda \frac{1-\widetilde{\theta}}{\widetilde{\theta}} t), \quad \sigma f_D(t) = \lambda \frac{1-\widetilde{\theta}}{\widetilde{\theta}} \exp(-\lambda \frac{1-\widetilde{\theta}}{\widetilde{\theta}} t), \quad \bar{t}_D = -\frac{\ln(1-\sigma)\widetilde{\theta}}{\lambda(1-\widetilde{\theta})}.$$

Substitution then yields

$$V_D(\widetilde{\theta}|\theta) = \int_0^{\bar{t}_D} \exp(-(\rho + \frac{\lambda}{\widetilde{\theta}})t) \frac{\lambda}{\widetilde{\theta}} (\widetilde{\theta} - \theta) a_D(t) dt + (1-\theta)(1-\sigma) \int_{\bar{t}_D}^{\infty} \lambda \exp(-(\rho+\lambda)t) dt.$$

This expression allows us to identify the main tradeoffs associated with delegation. By delegating to an evaluator with different preferences than her own, the principal changes the equilibrium belief during the phase of doubt from $g(t) = \theta$ to $g_D(t) = \widetilde{\theta}$. Therefore, by delegating to an evaluator who is more cautious than she is ($\widetilde{\theta} > \theta$), the principal gains positive expected surplus ($\widetilde{\theta} - \theta$) from any arrival that is approved during the phase of doubt. At the same time, the agent cheats less aggressively, at rate $\lambda(1 - \widetilde{\theta})/\widetilde{\theta} < \lambda(1 - \theta)/\theta$. This reduction in the agent's cheating rate has three negative consequences for the principal. First, it extends the phase of doubt ($\bar{t}_D > \bar{t}$), delaying the phase of credibility when all submissions are authentic. Second, it slows down the overall arrival rate of projects during the phase of doubt ($\lambda/\widetilde{\theta} < \lambda/\theta$) so that the expected benefit from approval is likely to be discounted more heavily. Third, the evaluator's approval probability is itself strictly lower during the phase of doubt ($a_D(t) < a(t)$ for $t < \bar{t}_D$). Of course, delegating to a less cautious proxy ($\widetilde{\theta} < \theta$) reverses these effects: arrivals during the phase of doubt are harmful if approved, arrive faster, and are more likely to be approved, but the phase of doubt is shorter.

Because of these tradeoffs, it is not obvious whether the principal benefits from delegating to a more cautious or less cautious evaluator, or – in fact – from delegation at all. While answering these questions in full generality is analytically intractable, we can answer them for important regions of the parameter space.

First we consider a limited form of delegation in which the principal can only delegate "locally," to an evaluator whose preference parameter is not too different from her own. This setting is interesting as a theoretical benchmark, but it also captures a plausible feature of some organizations. Specifically, suppose that the organization is somewhat rigid, preventing the principal from transferring formal authority to the evaluator. A transfer of real authority may nevertheless be possible as long as the principal pays a sufficiently high cost c (e.g., loss of reputation) if she overrules the evaluator. For instance, if she delegates to a more (less) cautious surrogate then – as discussed above – the principal will strictly expect to benefit (suffer) from acceptance of any arrival during the phase of doubt. However, if $|\widetilde{\theta} - \theta| < c$, then the principal will not overrule the evaluator's decision in equilibrium.[7] Thus, when transfer of formal authority is not possible, the principal can successfully delegate to evaluators with $\widetilde{\theta} \in (\theta - c, \theta + c)$. In particular, when c is small, only local delegation is credible. The following proposition characterizes the optimal policy in this case.

Proposition 6.1. *(Local Delegation). Compared with keeping the decision authority herself, the principal strictly gains (loses) by delegating to an evaluator who is slightly more (less)*

[7]If the principal delegates to a more-cautious evaluator ($\widetilde{\theta} > \theta$) who rejects an arrival in the phase of doubt, it is sequentially rational for the principal not to overrule provided $\widetilde{\theta} - \theta - c < 0$, i.e., $\widetilde{\theta} < \theta + c$. Conversely, if she delegates to a less-cautious evaluator ($\widetilde{\theta} < \theta$) who approves an arrival during the phase of doubt, the principal will prefer not to overrule the evaluator if $\widetilde{\theta} - \theta > -c$, i.e. $\widetilde{\theta} > \theta - c$. If the principal delegates to an evaluator such that $|\widetilde{\theta} - \theta| > c$, then an equilibrium of the delegation subgame exists in which the principal sometimes overrules the evaluator but is worse off than if she had retained authority herself (details available upon request).

cautious than she is; an $\epsilon > 0$ exists such that (1) $V_D(\widetilde{\theta}|\theta) > V$ for $\widetilde{\theta} \in (\theta, \theta + \epsilon)$, (2) $V > V_D(\widetilde{\theta}|\theta)$ for $\widetilde{\theta} \in (\theta - \epsilon, \theta)$.

This proposition has two noteworthy implications. First, even the very limited form of local delegation considered is valuable for the principal. In this sense, delegation is a robust and potentially powerful instrument for addressing the incentive problem that we study. Second, when delegating locally, the principal benefits from an evaluator who is more cautious and is harmed by an evaluator who is less cautious.

Next, we consider the principal's globally optimal choice of evaluator $\widetilde{\theta} \in [0, 1]$ when the probability that the agent is strategic is either close to 1 or close to 0.

Proposition 6.2. *(Global Delegation). If σ is sufficiently close to 1, then the principal's payoff is increasing in $\widetilde{\theta}$, and it is therefore optimal for her to select the most cautious evaluator, $\widetilde{\theta} = 1$. If σ is sufficiently close to 0, then it is optimal for the principal to select an evaluator who is more cautious than she is, but it is not optimal to select the most cautious evaluator; the optimal $\widetilde{\theta} \in (\theta, 1)$.*

To appreciate the significance of this proposition, note first that the tradeoff associated with the evaluator's preference parameter is only substantive if trust evolves over time. In particular, if the agent is a known strategic type ($\sigma = 1$), then the equilibrium is stationary and the phase of doubt is infinite regardless of the evaluator's preference. By implication, two of the costs associated with increasing $\widetilde{\theta}$ are absent (extending the phase of doubt, reducing the approval probability), and the principal's payoff is strictly increasing in $\widetilde{\theta}$. Therefore, for $\sigma = 1$ it is globally optimal to delegate to the most cautious evaluator, $\widetilde{\theta} = 1$. Proposition 6.2 shows that these findings continue to hold when the agent is sufficiently likely to be strategic, but that they are overturned if the agent is sufficiently likely to be ethical. In particular, for σ sufficiently close to 0, the principal's payoff is not monotonically increasing in $\widetilde{\theta}$ so that the global optimum involves $\widetilde{\theta} \in (\theta, 1)$. In fact, as we show next, excessive caution on the part of the evaluator may completely undo the benefit of delegation.

Proposition 6.3. *(Extreme Caution). If σ is below a threshold, then delegating to an evaluator who is too cautious is strictly worse for the principal than keeping the authority herself; given $\sigma < \overline{\sigma}$ (characterized in closed form), there exists $\epsilon > 0$ such $V_D(\widetilde{\theta}|\theta) < V$ for $\widetilde{\theta} \in (1 - \epsilon, 1]$.*

In spite of the complex form of the principal's payoff function in the delegation subgame, Propositions 6.1, 6.2, and 6.3 shed light on the principal's optimal policy. She always benefits from employing a proxy that is slightly more cautious than she is, even if she is almost sure that the agent is ethical. Furthermore, she prefers the most cautious proxy if and only if she is sufficiently sure that the agent is strategic. In the parameter cases we have analyzed, the principal never prefers to delegate to a proxy who is less cautious than she is. On this point, it is worth noting that in our numerical simulations, the principal never benefits from employing an evaluator less cautious than herself.

7 Auditing

Another common method for mitigating fraud is to audit submissions prior to approval. We explore this possibility in this section. Throughout we assume that the principal cannot commit to a testing policy, and solve for an equilibrium of the resulting project submission and auditing game.

Suppose that the principal can pay a cost $k > 0$ to perform a test on a project that reveals fraud with probability $\alpha \in (0, 1]$. That is, if the project is fake, it fails the test with probability α and passes the test (a type II error) with probability $1 - \alpha$. Furthermore, a real project always passes the test.[8] An audit is called *strong* if $\alpha > \alpha^* \equiv \widehat{\phi} - \phi$, and *weak* if $\alpha < \alpha^*$.

Let $g(t)$ represent the principal's belief that a submission at time t is real before performing an audit. We begin with an observation that follows from basic reasoning about the value of information, namely that there is no point in paying for a costly signal unless different realizations lead to different optimal decisions.

Lemma 7.1. *If it is sequentially rational for the principal to test a submission at t, then it is optimal to approve the project if and only if it passes the test.*

Following an arrival at time t, the principal chooses between auditing, approving with no audit, and rejecting with no audit. She therefore compares the following three expected payoffs,

$$\begin{array}{ll} \text{Audit:} & g(t) - \theta + \alpha\theta(1 - g(t)) - k \\ \text{Approve:} & g(t) - \theta \\ \text{Reject:} & 0. \end{array} \tag{10}$$

If the test is too expensive, then the principal will not use it and the equilibrium of the main model, where $g(t) = \theta$ over the phase of doubt, will obtain. Plugging $g(t) = \theta$ into (10) reveals that the test will be deployed in equilibrium if and only if

$$k < \alpha\theta(1 - \theta) \equiv k^*,$$

[8] Allowing for real submissions to fail the test (a type I error) complicates expressions significantly without adding additional insight.

which we assume throughout the rest of this section.

Let $p(t)$ be the probability the principal tests a project that arrives at time t, $a(t)$ be the probability she approves the arrival without testing, and $r(t)$ be the probability she rejects without testing. Comparing the principal's payoffs from each of the three alternatives in (10), we find the following sequentially rational strategy.

$$\begin{aligned} a(t) = 1 &\quad \text{if} \quad g(t) \in (\theta_1, 1] \\ a(t) + p(t) = 1 &\quad \text{if} \quad g(t) = \theta_1 \\ p(t) = 1 &\quad \text{if} \quad g(t) \in (\theta_0, \theta_1) \\ p(t) + r(t) = 1 &\quad \text{if} \quad g(t) = \theta_0 \\ r(t) = 1 &\quad \text{if} \quad g(t) \in [0, \theta_0), \end{aligned} \tag{11}$$

where

$$\theta_1 \equiv 1 - \frac{k}{\alpha\theta} \quad \text{and} \quad \theta_0 \equiv \frac{(1-\alpha)\theta + k}{1 - \alpha\theta}.$$

Note that $k < k^*$ implies $\theta_0 < \theta < \theta_1$and $k = k^*$ implies $\theta_0 = \theta = \theta_1$. Thus, the principal approves when her belief is high, rejects when her belief is low, and audits for a (non-degenerate) interval of moderate beliefs surrounding θ. At the boundaries between these intervals, the principal is indifferent between the adjacent actions. In particular, if $g(t) = \theta_1$, then the principal is indifferent between approving and auditing. If the principal mixes between auditing and acceptance at time t, then we say that she "mixes with default approval." Similarly, if $g(t) = \theta_0$, the principal is indifferent between auditing and rejecting. If she mixes between auditing and rejecting at time t, we say that she "mixes with default rejection." Define the perceived faking rates associated with $\theta_i \in \{\theta_0, \theta_1\}$ as μ_i; i.e., $g(t) = \theta_i \Leftrightarrow \nu(t) = \mu_i \equiv \lambda(1-\theta_i)/\theta_i$; clearly, $\mu_0 > \mu_1$. We establish next that the principal audits randomly in equilibrium, unless she is certain that the project is real.

Lemma 7.2. *In any equilibrium the following must hold:*

(i) If $\nu(t) > 0$ for all $t \in (t_1, t_2)$, then $p(t) > 0$ for all such t.

(ii) No open interval (t_1, t_2) exists such that $p(t) = 1$ for all $t \in (t_1, t_2)$.

Part (i) is easily explained. If the principal does not audit in some interval where the agent fakes, then maintaining the agent's indifference requires the principal to mix between approval and rejection. Thus, the agent must mix as in the main model, with $\nu(t) = \mu$ and $g(t) = \theta$. But then auditing would be strictly beneficial (see (11)).

To appreciate part (ii), suppose to the contrary there is an equilibrium in which the principal audits every submission in some time interval (t_1, t_2). If the test is strong, then the agent expects a fake project to be detected with relatively high probability α. In this case, the agent would like to delay faking as much as possible in order to increase the chance that a real project arrives (his expected payoff of faking is strictly increasing in t). In contrast, if the test is weak, then the agent expects that a fake will pass the test with relatively high probability $1-\alpha$ and will be approved. Because he is impatient, the agent would like to fake as soon as possible (his expected payoff is strictly decreasing in t). In either case, no fakes are submitted in the interior of the interval, and the principal can profitably deviate by approving a submission at $t \in (t_1, t_2)$ without performing a costly audit.

For the following discussion, we stipulate that the equilibrium begins with a phase of doubt in which the agent fakes (this is proved formally in Lemma 7.3 with similar intuition to the main model). Lemma 7.2 shows that the principal must audit randomly at almost every time in the phase of doubt, but the question is whether she randomizes with default rejection or default acceptance. In other words, the phase of doubt can be sub-divided into stages, depending on whether the principal mixes with default approval or default rejection.

To understand the structure of these stages, note first that in order to offset the agent's impatience, the probability that a fake project is approved, $a_F(\cdot)$, must increase over the phase of doubt.[9] The logic is similar to the main model—the agent must mix over faking times in the phase of doubt to avoid detection, but delayed faking is costly due to impatience. Thus, to maintain indifference, a fake that is submitted later must be approved with higher probability. In a stage of mixing with default approval, a fake is approved unless it is tested and fails, probability $a_F(t) = 1 - \alpha p(t)$. Thus, in a default approval stage, $p(\cdot)$ must be decreasing. In contrast, in a stage of mixing with default rejection, a fake is approved if it is tested and it passes, $a_F(t) = p(t)(1-\alpha)$. Therefore $p(\cdot)$ must be increasing in a default rejection stage. That $a_F(\cdot)$ is increasing also implies the equilibrium can have at most one stage of mixing with default rejection, and if such a stage exists, it must occur before a stage of mixing with default approval. As described above, during a default rejection stage $a_F(t) = p(t)(1-\alpha) \leq 1-\alpha$; meanwhile during a default approval stage, $a_F(t) = 1-\alpha p(t) \geq 1-\alpha$. Thus, an increasing $a_F(\cdot)$ implies that a stage of default rejection occurs before default approval, if at all, and that there can be no more than one default rejection stage.

Additionally, continuity of $a_F(\cdot)$, which is necessary for the agent's indifference, determines boundary conditions for $p(\cdot)$ at the transitions between stages. In particular, at the transition time between the stages of default rejection and approval, we must have $(1-\alpha)p(t) =$

[9] Unlike the main model where the approval probability is identical for real and fake projects, auditing introduces a wedge, $a_R(t) > a_F(t)$. If it is tested, a real project always passes the test, but a fake passes with probability $1-\alpha$. The agent's strategic decision about the timing of his fake depends crucially on the probability that a *fake* submission is approved, and thus, on the properties of $a_F(\cdot)$. The probability that a real project is approved $a_R(\cdot)$ plays a role in the equilibrium characterization of the audit probability.

$1-\alpha p(t) \Rightarrow p(t)=1$. Furthermore, because $a_F(\cdot) \leq 1-\alpha$ in a default rejection stage, but $a_F(\cdot)=1$ in the phase of credibility, the phase of doubt must end with a stage of default approval. Indeed, at the transition between the stage of default approval and the phase of credibility, we must have $1-\alpha p(t)=1 \Rightarrow p(t)=0$. The following result formalizes these observations.

Lemma 7.3 (Auditing Equilibrium Structure)**.** *In an equilibrium with auditing, there exists* $\bar{t}_A \in (0,\infty)$ *and* $\widetilde{t}_A \in [0,\bar{t}_A)$ *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 $[0,\bar{t}_A]$*.*

(ii) for $t \in [\bar{t}_A,\infty)$*, the principal approves the project with no audit,* $a(t)=1$*, and the agent never fakes,* $\nu(t)=0$*.*

(iii) for $t \in [\widetilde{t}_A,\bar{t}_A)$ *the principal mixes between auditing and approval,* $p(t) \in (0,1)$ *and* $a(t)+p(t)=1$*. The belief* $g(t)=\theta_1$*, and the perceived rate of faking is* $\nu(t)=\mu_1$*. The probability of auditing is strictly decreasing, continuous, and differentiable, with* $\lim_{t\to\bar{t}_A} p(t)=0$*.*

(iv) if $\widetilde{t}_A>0$*, then for* $t \in [0,\widetilde{t}_A)$ *the principal mixes between auditing and rejection,* $p(t) \in (0,1)$ *and* $p(t)+r(t)=1$*. The belief* $g(t)=\theta_0$*, and the perceived rate of faking is* $\nu(t)=\mu_0$*. The probability of auditing is strictly increasing, continuous, and differentiable, with* $\lim_{t\to\widetilde{t}_A} p(t)=1$*. Furthermore, for* $t \in (\widetilde{t}_A,\bar{t}_A)$*, we have* $\lim_{t\to\widetilde{t}_A} p(t)=p(\widetilde{t}_A)=1$*.*

We refer to the structure with $\widetilde{t}_A=0$ as the *one-stage structure*, (default approval only), and to $\widetilde{t}_A>0$ as the *two-stage structure* (mixing with default rejection, then mixing with default approval). In both cases, we refer to the interval $[0,\bar{t}_A]$ as the phase of doubt. The one-stage and two-stage structures are summarized concisely in Figure 2.

Lemma 7.3 suggests that the dynamics of auditing are closely tied to the evolution of trust. Consider the two-stage structure. In the initial stage of mixing with default rejection, trust is relatively low. The agent fakes aggressively and the principal behaves skeptically, occasionally testing a submission in hopes of fishing out a rare real project and rejecting the rest. As time passes and trust grows, the principal behaves less skeptically toward the project and is more likely to test it. Once trust reaches a certain point, the principal tests with probability 1. At this point, the equilibrium transitions to the second stage in which the principal mixes with default approval. Here trust is relatively high but not perfect: the agent fakes less aggressively and the principal gives the project the benefit of the doubt, approving it if she doesn't audit. As trust continues to grow, the principal is more inclined to approve

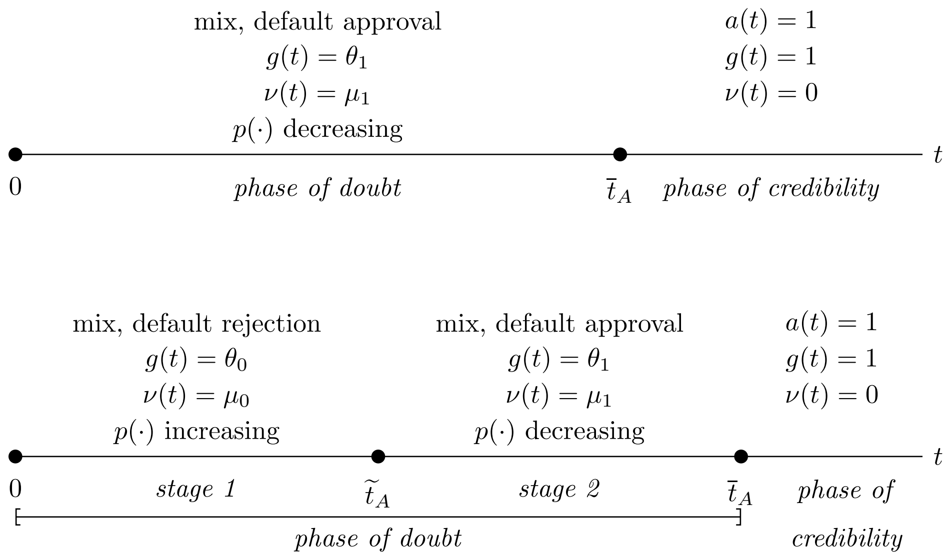

Figure 2: *The one-stage structure (top) and two-stage structure (bottom). Note:* $p(\cdot)$ *is continuous, with* $p(\widetilde{t}_A) = 1$ *and* $p(\bar{t}_A) = 0$*. Agent mixed strategy has no mass points or gaps.*

submissions without testing them. At the end of the second stage the audit probability becomes 0—the game transitions to the phase of credibility. The principal is confident that the agent is ethical, and all arrivals are approved. In other words, the principal audits with relatively low probability when trust is low and when trust is high, and audits with relatively high probability in between.

The following result summarizes the key properties of the equilibrium with auditing—a complete closed-form characterization is presented in the Appendix (Propositions A.1, A.2).

Proposition 7.1 (Auditing Equilibrium Summary)**.** *The equilibrium of the game with auditing has the following features.*

(i) The equilibrium of the game with auditing is unique.

(ii) If the test is strong ($\alpha > \alpha^*$) *then the equilibrium has the one-stage structure.*

(iii) If the test is weak ($\alpha < \alpha^*$)*, then a threshold* $\sigma^* \in (0, 1)$ *exists such that*

- *if the test is weak and* $\sigma < \sigma^*$*, then the equilibrium has the one-stage structure.*
- *if the test is weak and* $\sigma > \sigma^*$*, then the equilibrium has the two-stage structure. At the transition time between the first and second stage,* $\widetilde{t}_A > 0$*, the principal's belief that the agent is ethical is* σ^**.*

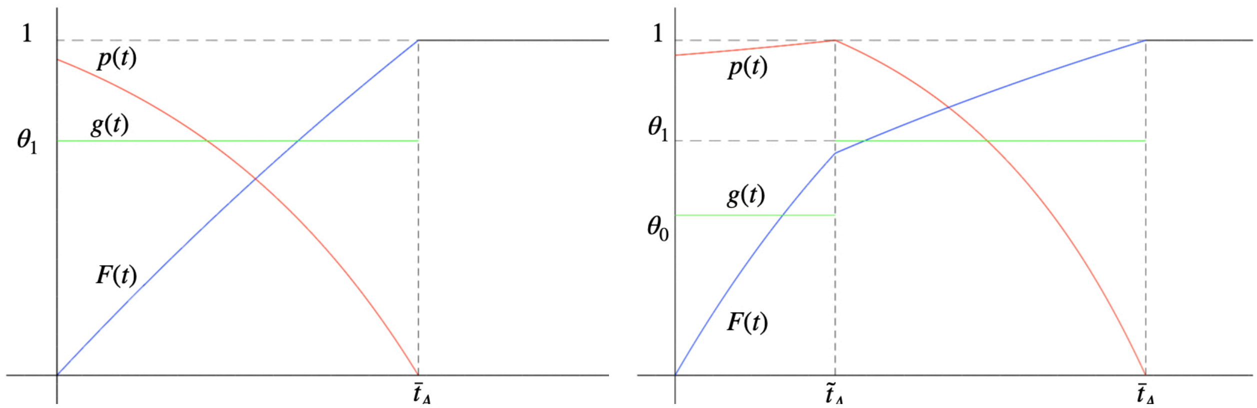

Figure 3: *Illustration of Proposition 7.1. Parameters are identical in both figures, except for σ. The test is weak. In the left panel, $\sigma < \sigma^*$ and equilibrium has a one-stage structure. In the right panel, $\sigma > \sigma^*$ and equilibrium has a two-stage structure. Graphs were generated from the closed-form characterization in the Appendix.*

While Lemma 7.3 suggests a link between the dynamics of auditing and trust, Proposition 7.1 characterizes this connection explicitly. In equilibrium, the only role of a strong test is to identify and reject fake projects. Recognizing that a fake is likely to be exposed by the test, the agent treads lightly, faking at a low rate. As trust grows, the principal gives the agent additional leeway, checking his work less often, and once she becomes confident that he is ethical, not at all. Similar dynamics can also arise with a weak test, but only if initial trust exceeds a critical threshold, $\sigma < \sigma^*$. At low levels of trust a weak test is used to identify rare real projects that would otherwise be rejected, and the growth of trust increases the probability of testing. At time $\widetilde{t}_A$, trust reaches the critical level σ^*, the testing probability reaches 1, and the "regime" changes: the principal gives the agent the benefit of the doubt, switching from default rejection to acceptance, the perceived faking rate drops, and the role of testing flips from identifying real projects for approval to identifying fakes for rejection.

To understand auditing's impact on the principal's payoff, first suppose the equilibrium has the one-stage structure. In this case, $g(t) = \theta_1$ and $\nu(t) = \mu_1$ over the entire phase of doubt. Thus, the agent cheats less aggressively than in the main model, at perceived rate $\mu_1 < \mu$ and the belief that an arrival is real is higher, $\theta_1 > \theta$. Furthermore, for any arrival during the phase of doubt, the principal is indifferent between approving and auditing, and thus her expected payoff is the same as if she approves. Thus, when the equilibrium has the one-stage structure, the phase of doubt is longer than in the main model, but any arrival during this phase generates an expected payoff to the principal of $\theta_1 - \theta$. This is identical to delegating to a more cautious proxy, $\widetilde{\theta} = \theta_1$, with one crucial difference. Under delegation, a submission during the phase of doubt yields the principal $a_D(t)(\theta_1 - \theta)$. Because the proxy approves with probability $a_D(t) < 1$ during the phase of doubt, the principal prefers auditing to delegating to such an evaluator.

In the two-stage equilibrium, an additional complication arises. In the second stage (mixing with default approval), the same basic tradeoff applies—the slowdown in cheating increases the principal's surplus from an arrival, but delays the onset of the phase of credibility. However, during the first stage (mixing with default rejection) the principal is indifferent between testing and rejection, and therefore obtains no expected surplus. At the same time, the agent cheats rapidly during the first stage (i.e., $\mu_0 > \mu > \mu_1$), reducing the duration of the phase of doubt. The tradeoff in the initial default rejection stage is similar to delegating to a less cautious evaluator with $\widetilde{\theta} = \theta_0$, except that the principal's expected payoff from a submission under auditing is 0 and under delegation is $a_D(t)(\theta_0 - \theta) < 0$.

Building on these observations, in the next proposition we show that the principal always benefits from the ability to audit. It is important to note that in equilibrium, the principal is indifferent between auditing and her default action at all times in the phase of doubt. Therefore, the value of information from conducting an audit is exactly equal to its cost. Thus, the normative gains for the principal do not come directly from her ability to acquire information about the project. Rather, these gains come indirectly, via the effect on the strategic agent's incentive to commit fraud.

Proposition 7.2 (Normative Analysis of Auditing.)**.** *In the equilibrium with auditing, the principal's and ethical agent's payoffs are strictly higher than in the equilibrium of the baseline model, and the strategic agent's payoff is higher if σ is sufficiently large.*

Both types of agent can benefit when the principal has the ability to audit a submission. Because the ethical agent only submits a real project, when the principal mixes with default acceptance his projects are always approved. Thus, when the equilibrium has the one-stage structure, the ethical agent gets his first best payoff. Similarly, in the two-stage equilibrium, the ethical agent's projects are always approved past time $\widetilde{t}_A$, but at $t < \widetilde{t}_A$ the ethical agent's project is only approved when it is tested. It is relatively straightforward to show that in the two-stage structure, the probability of testing for $t < \widetilde{t}_A$ is larger than the probability of acceptance in the baseline model. Thus, for all possible times, the probability that a real project is approved is higher in the equilibrium with auditing than in the main model. Consequently, the ethical agent's equilibrium payoff is also higher.

More remarkably, the increased probability that a real project is accepted under auditing may also benefit the strategic agent. To see this most clearly, suppose that $\sigma \to 1$, so that the phase of doubt becomes unbounded, both in the equilibrium of the baseline model and the auditing model. In this case, the strategic agent is indifferent in equilibrium over *all* possible cheating times in $\mathbb{R}_+$, and his equilibrium payoff is therefore the same as if he never cheats (and the principal follows her equilibrium strategy). But if the strategic agent never cheats, and the probability that a real arrival is accepted is higher with auditing, then the strategic type's real arrivals are also more likely to be approved. Thus, when σ is large, the strategic type also has a higher payoff with auditing than without, and the existence of the auditing technology generates a Pareto improvement.

8 Conclusion

In this paper we investigate a dynamic model of fraud and the evolution of 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 producing a fake one.

The key force in our analysis is the growth of trust over time. Because the strategic agent can use either the real or fake technology while the ethical agent can only use the real one, a strategic agent is more likely to generate an arrival at each time (and strictly so if the likelihood of fraud is non-zero). Thus, as time passes without an arrival, the principal believes that the agent is more likely ethical and trust grows. This effect interacts with the strategic agent's incentive to commit fraud and shapes the equilibrium in the main model, and the delegation and auditing extensions. In the main model, the growth of trust generates a finite phase of doubt, which allows the principal to benefit from the interaction. In the delegation extension, the evolution of trust generates costs for increasing caution which affect the principal's optimal evaluator. The results of the auditing section highlight the link between the principal's auditing strategy and the level of trust in the relationship, revealing novel dynamics. This force also introduces subtlety into the normative analysis, ensuring that even a weak test benefits 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 relationship will end exogenously. The 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 commit fraud is generated by a deadline. As a deadline approaches, the probability of meeting it by honest means decreases, and the agent may resort to fraud 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 "talented" 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, and we intend to do so in future work.

References

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.

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.

Guesnerie, R., and J.-J. Laffont, 1984, "A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-Managed Firm," *Journal of Public Economics*, 25, 329–369.

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.*

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.

Strulovici, B., 2020, "Can society function without ethical agents? an informational perspective," *arXiv preprint arXiv:2003.05441.*

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

Sun, Y., forthcoming, "A Dynamic Model of Censorship," *Theoretical Economics.*

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

A Proofs

Flow cost for real technology. 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,

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

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 any 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 for a submission 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, $\lim_{t\to\infty} \nu(t) = 0$. It follows that there exists t^* such that, $\nu(t) < \theta$ 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, $\nu(t) = 0$, 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 $f(t) = 0 \implies \nu(t) = 0 \implies g(t) = 1 \implies a(t) = 1$, 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 $\nu(t') = 0$ 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 $\nu(t) = \lambda(1-\theta)/\theta > 0$. 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 $\nu(t) = \infty$, 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, $\nu(t) = \lambda(1 - \theta)/\theta$ 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,

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

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 $t \in (0,\overline{t})$. So it is either constant, strictly increasing, or strictly decreasing. Suppose $a(t)$ is constant or strictly decreasing for all $t < \overline{t}$. It follows that $a(t) \le \phi/\widehat{\phi}$ for all $t < \overline{t}$. Because $\phi/\widehat{\phi} < 1$, we have $\lim_{t\to\overline{t}} a(t) < 1$, 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 $t \in (0,\overline{t})$ 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 $t \in [0,\overline{t}]$, 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 $t \in (0,\overline{t})$, we also have

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

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

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

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

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

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 $t \in [0,\overline{t})$, and hence, $U^S = a(0) - \phi$. Simplifying, we have

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

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,

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

□

A.2 Proofs for Delegation

Principal's Payoff Function. In the equilibrium with delegation, the agent's mixed strategy $F_D(\cdot)$ is such that

$$(1-\sigma F_D(t)) = \exp(-\lambda\frac{1-\widetilde{\theta}}{\widetilde{\theta}}t) \quad \text{and} \quad \sigma f_D(t) = \lambda\frac{1-\widetilde{\theta}}{\widetilde{\theta}}\exp(-\lambda\frac{1-\widetilde{\theta}}{\widetilde{\theta}}t),$$

with support on $[0, \bar{t}_D]$ where

$$\bar{t}_D \equiv -\frac{\widetilde{\theta}\ln(1-\sigma)}{\lambda(1-\widetilde{\theta})},$$

the evaluator's equilibrium acceptance strategy is

$$a_D(t) = \frac{\phi}{\widehat{\phi}} + (1-\frac{\phi}{\widehat{\phi}})\exp\{-\rho(\bar{t}_D - t)\},$$

and the belief that an arrival is authentic inside the phase of doubt is $g_D(t) = \widetilde{\theta}$. Substituting, we find that

$$w_D(t) = \exp(-\lambda t)[\lambda(1-\sigma F_D(t)) + \sigma f_D(t)] = \exp(-\lambda t)\lambda(1+\frac{1-\widetilde{\theta}}{\widetilde{\theta}})\exp(-\lambda\frac{1-\widetilde{\theta}}{\widetilde{\theta}}t) = \frac{\lambda}{\widetilde{\theta}}\exp(-\frac{\lambda}{\widetilde{\theta}}t).$$

Proof of Proposition 6.1. We show that $\partial V_D/\partial\widetilde{\theta}]_{\widetilde{\theta}=\theta} > 0$, which implies that $V_D(\cdot|\theta)$ is increasing at $\widetilde{\theta} = \theta$. Differentiate the principal's payoff with respect to $\widetilde{\theta}$,

$$\begin{aligned}\frac{\partial V_D}{\partial\widetilde{\theta}} &= \int_0^{\bar{t}_D} \frac{d}{d\widetilde{\theta}}\Big(\exp(-(\rho+\frac{\lambda}{\widetilde{\theta}})t)\frac{\lambda}{\widetilde{\theta}}a_D(t)\Big)(\widetilde{\theta}-\theta)dt \\ &+ \int_0^{\bar{t}_D} \exp(-(\rho+\frac{\lambda}{\widetilde{\theta}})t)\frac{\lambda}{\widetilde{\theta}}a_D(t)dt + \exp(-(\rho+\frac{\lambda}{\widetilde{\theta}})\bar{t}_D)\frac{\lambda}{\widetilde{\theta}}a_D(\bar{t}_D)(\widetilde{\theta}-\theta)\frac{d\bar{t}_D}{d\widetilde{\theta}} \\ &- (1-\theta)(1-\sigma)\lambda\exp(-(\rho+\lambda)\bar{t}_D)\frac{d\bar{t}_D}{d\widetilde{\theta}}\end{aligned}$$

Note that

$$\frac{d\bar{t}_D}{d\widetilde{\theta}} = \frac{-\ln(1-\sigma)}{\lambda}\frac{1}{(1-\widetilde{\theta})^2} = \frac{\bar{t}_D}{\widetilde{\theta}(1-\widetilde{\theta})}.$$

Next, evaluate this derivative at the principal's preference parameter, $\widetilde{\theta} = \theta$, which implies, $\bar{t}_D = \bar{t}$ and $a_D(t) = a(t)$. We have

$$\begin{aligned}\frac{\partial V_D}{\partial\widetilde{\theta}}\Big]_{\widetilde{\theta}=\theta} &= \int_0^{\bar{t}} \exp(-(\rho+\frac{\lambda}{\theta})t)\frac{\lambda}{\theta}a(t)dt - (1-\theta)(1-\sigma)\lambda\exp(-(\rho+\lambda)\bar{t})\frac{\bar{t}}{\theta(1-\theta)} \\ &= \int_0^{\bar{t}} \exp(-(\rho+\frac{\lambda}{\theta})t)\frac{\lambda}{\theta}a(t)dt - (1-\sigma)\frac{\lambda}{\theta}\exp(-(\rho+\lambda)\bar{t})\bar{t}.\end{aligned}$$

Note that

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

$$= \exp\{-(\frac{\lambda}{\theta}+\rho)\bar{t}\},$$

and hence,

$$\begin{aligned}\frac{\partial V_D}{\partial\widetilde{\theta}}\Big]_{\widetilde{\theta}=\theta} &= \int_0^{\bar{t}} \exp(-(\rho+\frac{\lambda}{\theta})t)\frac{\lambda}{\theta}a(t)dt - (1-\theta)(1-\sigma)\lambda\exp(-(\rho+\lambda)\bar{t})\frac{\bar{t}}{\theta(1-\theta)} \\ &= \int_0^{\bar{t}} \exp(-(\rho+\frac{\lambda}{\theta})t)\frac{\lambda}{\theta}a(t)dt - \frac{\lambda}{\theta}\exp(-(\rho+\frac{\lambda}{\theta})\bar{t})\bar{t}.\end{aligned}$$

Thus,

$$\frac{\partial V_D}{\partial\widetilde{\theta}}\Big]_{\widetilde{\theta}=\theta} > 0 \iff \frac{\int_0^{\bar{t}}\exp(-(\rho+\frac{\lambda}{\theta})t)\frac{\lambda}{\theta}a(t)dt}{\bar{t}} > \exp(-(\rho+\frac{\lambda}{\theta})\bar{t})\frac{\lambda}{\theta}a(\bar{t}).$$

Thus, the derivative in question is positive if the average value of $\Gamma(t) \equiv \exp(-(\rho+\frac{\lambda}{\theta})t)\frac{\lambda}{\theta}a(t)$ on interval $[0,\bar{t}]$ is larger than $\Gamma(\bar{t})$. To complete the proof, we show that $\Gamma(\cdot)$ is a decreasing function, which implies that the preceding inequality holds. Substituting $a(\cdot)$ and simplifying, we have

$$\begin{aligned}\Gamma(t) &= \frac{\lambda}{\theta}[\frac{\phi}{\widehat{\phi}}\exp(-(\rho+\frac{\lambda}{\theta})t)+(1-\frac{\phi}{\widehat{\phi}})\exp(-\rho(\bar{t}-t)-(\rho+\frac{\lambda}{\theta})t)] \\ &= \frac{\lambda}{\theta}[\frac{\phi}{\widehat{\phi}}\exp(-(\rho+\frac{\lambda}{\theta})t)+(1-\frac{\phi}{\widehat{\phi}})\exp(-\rho\bar{t}-\frac{\lambda}{\theta}t)].\end{aligned}$$

Because $\Gamma(t)$ is the sum of two decreasing functions, it is itself decreasing. □

Proof of Proposition 6.2. Step 1. We evaluate the integrals in the principal's payoff function

$$V_D(\widetilde{\theta}|\theta) = \int_0^{\bar{t}_D}\exp(-(\rho+\frac{\lambda}{\widetilde{\theta}})t)\frac{\lambda}{\widetilde{\theta}}(\widetilde{\theta}-\theta)a_D(t)dt+(1-\theta)(1-\sigma)\int_{\bar{t}_D}^{\infty}\lambda\exp(-(\rho+\lambda)t)dt.$$

The first integral is somewhat long, but it is straightforward to verify that

$$\int\exp(-(\rho+\frac{\lambda}{\widetilde{\theta}})t)\frac{\lambda}{\widetilde{\theta}}(\widetilde{\theta}-\theta)a_D(t)dt = -(\widetilde{\theta}-\theta)\Big(\frac{\phi}{\widehat{\phi}}\frac{\lambda}{\lambda+\widetilde{\theta}\rho}\exp(-(\rho+\frac{\lambda}{\widetilde{\theta}})t)+(1-\frac{\phi}{\widehat{\phi}})\exp(-\frac{\lambda}{\widetilde{\theta}}t-\bar{t}_D\rho)\Big).$$

Evaluating and simplifying the definite integral, we have

$$\begin{aligned}&\int_0^{\bar{t}_D}\exp(-(\rho+\frac{\lambda}{\widetilde{\theta}})t)\frac{\lambda}{\widetilde{\theta}}(\widetilde{\theta}-\theta)a_D(t)dt = (\widetilde{\theta}-\theta)(\frac{\phi}{\widehat{\phi}})\frac{\lambda}{\lambda+\widetilde{\theta}\rho}+ \\ &(\widetilde{\theta}-\theta)(1-\frac{\phi}{\widehat{\phi}})\exp(-\rho\bar{t}_D)-(\widetilde{\theta}-\theta)(\frac{\lambda}{\lambda+\widetilde{\theta}\rho}+(1-\frac{\phi}{\widehat{\phi}})\frac{\widetilde{\theta}\rho}{\lambda+\widetilde{\theta}\rho})\exp(-(\rho+\frac{\lambda}{\widetilde{\theta}})\bar{t}_D).\end{aligned}$$

For the second integral, we have

$$(1-\theta)(1-\sigma)\int_{\bar{t}_D}^{\infty}\lambda\exp(-(\rho+\lambda)t)dt = (1-\theta)\frac{\lambda}{\lambda+\rho}(1-\sigma)\exp(-(\rho+\lambda)\bar{t}_D).$$

Note that

$$\begin{aligned}(1-\sigma)\exp(-(\rho+\lambda)\bar{t}_D) &= \exp\{(1+(\rho+\lambda)\frac{\widetilde{\theta}}{\lambda(1-\widetilde{\theta})})\ln(1-\sigma)\} \\ &= \exp\{-(\frac{\lambda}{\widetilde{\theta}}+\rho)(-\frac{\widetilde{\theta}}{\lambda(1-\widetilde{\theta})})\ln(1-\sigma)\} = \exp\{-(\frac{\lambda}{\widetilde{\theta}}+\rho)\bar{t}_D\}.\end{aligned}$$

Thus, we have

$$(1-\theta)(1-\sigma)\int_{\bar{t}_D}^{\infty}\lambda\exp(-(\rho+\lambda)t)dt = (1-\theta)\frac{\lambda}{\lambda+\rho}\exp(-(\rho+\frac{\lambda}{\widetilde{\theta}})\bar{t}_D).$$

Next, substitute $\bar{t}_D$. Noting that

$$\begin{aligned}\exp(-\rho\bar{t}_D) &= (1-\sigma)^{\frac{\rho\widetilde{\theta}}{\lambda(1-\widetilde{\theta})}} \\ \exp(-(\rho+\frac{\lambda}{\widetilde{\theta}})\bar{t}_D) &= (1-\sigma)^{\frac{\lambda+\rho\widetilde{\theta}}{\lambda(1-\widetilde{\theta})}}.\end{aligned}$$

Combining these calculations we have

$$\begin{aligned}V_D(\widetilde{\theta}|\theta) =&(\widetilde{\theta}-\theta)(\frac{\phi}{\widehat{\phi}})\frac{\lambda}{\lambda+\widetilde{\theta}\rho} + (\widetilde{\theta}-\theta)(1-\frac{\phi}{\widehat{\phi}})(1-\sigma)^{\frac{\rho\widetilde{\theta}}{\lambda(1-\widetilde{\theta})}} + \\ &\Big((1-\theta)\frac{\lambda}{\lambda+\rho} - (\widetilde{\theta}-\theta)(\frac{\lambda}{\lambda+\widetilde{\theta}\rho} + (1-\frac{\phi}{\widehat{\phi}})\frac{\widetilde{\theta}\rho}{\lambda+\widetilde{\theta}\rho})\Big)(1-\sigma)^{\frac{\lambda+\rho\widetilde{\theta}}{\lambda(1-\widetilde{\theta})}}\end{aligned}$$

To simplify notation, we write the payoff function as

$$V_D(\widetilde{\theta}|\theta) = K_1(\widetilde{\theta}) + K_2(\widetilde{\theta})(1-\sigma)^{\frac{\rho\widetilde{\theta}}{\lambda(1-\widetilde{\theta})}} + K_3(\widetilde{\theta})(1-\sigma)^{\frac{\lambda+\rho\widetilde{\theta}}{\lambda(1-\widetilde{\theta})}},$$

where

$$\begin{aligned}K_1(\widetilde{\theta}) &\equiv (\widetilde{\theta}-\theta)(\frac{\phi}{\widehat{\phi}})\frac{\lambda}{\lambda+\widetilde{\theta}\rho} \\ K_2(\widetilde{\theta}) &\equiv (\widetilde{\theta}-\theta)(1-\frac{\phi}{\widehat{\phi}}) \\ K_3(\widetilde{\theta}) &\equiv (1-\theta)\frac{\lambda}{\lambda+\rho} - (\widetilde{\theta}-\theta)(\frac{\lambda}{\lambda+\widetilde{\theta}\rho} + (1-\frac{\phi}{\widehat{\phi}})\frac{\widetilde{\theta}\rho}{\lambda+\widetilde{\theta}\rho}).\end{aligned}$$

Step 2. We show that $V_D(\cdot|\theta)$ is increasing in $\widetilde{\theta}$ if σ is sufficiently large. That it is optimal to delegate to the most cautious evaluator, $\widetilde{\theta}=1$, follows immediately. Differentiating with respect to $\widetilde{\theta}$, we have

$$\begin{aligned}\frac{\partial V_D}{\partial\widetilde{\theta}} =&K_1'(\widetilde{\theta}) + K_2'(\widetilde{\theta})(1-\sigma)^{\frac{\rho\widetilde{\theta}}{\lambda(1-\widetilde{\theta})}} + K_2(\widetilde{\theta})\frac{d}{d\widetilde{\theta}}(1-\sigma)^{\frac{\rho\widetilde{\theta}}{\lambda(1-\widetilde{\theta})}} + \\ &K_3'(\widetilde{\theta})(1-\sigma)^{\frac{\lambda+\rho\widetilde{\theta}}{\lambda(1-\widetilde{\theta})}} + K_3(\widetilde{\theta})\frac{d}{d\widetilde{\theta}}(1-\sigma)^{\frac{\lambda+\rho\widetilde{\theta}}{\lambda(1-\widetilde{\theta})}}.\end{aligned} \tag{A.1}$$

Note that

$$\begin{aligned}K_1'(\widetilde{\theta}) &= (\frac{\lambda\phi}{\widehat{\phi}})\frac{\lambda+\theta\rho}{(\lambda+\widetilde{\theta}\rho)^2} \\ \frac{d}{d\widetilde{\theta}}(1-\sigma)^{\frac{\rho\widetilde{\theta}}{\lambda(1-\widetilde{\theta})}} &= \frac{\rho}{\lambda(1-\widetilde{\theta})^2}(1-\sigma)^{\frac{\rho\widetilde{\theta}}{\lambda(1-\widetilde{\theta})}}\ln(1-\sigma) \\ \frac{d}{d\widetilde{\theta}}(1-\sigma)^{\frac{\lambda+\rho\widetilde{\theta}}{\lambda(1-\widetilde{\theta})}} &= \frac{\lambda+\rho}{\lambda(1-\widetilde{\theta})^2}(1-\sigma)^{\frac{\lambda+\rho\widetilde{\theta}}{\lambda(1-\widetilde{\theta})}}\ln(1-\sigma)\end{aligned} \tag{A.2}$$

Next, we take the limit as $\sigma\to 1^-$. Noting that $\lim_{\sigma\to1^-}(1-\sigma)^a\ln(1-\sigma)=0$ for any $a>0$, we have

$$\lim_{\sigma\to1^-}\frac{\partial V_D}{\partial\widetilde{\theta}} = K_1'(\widetilde{\theta}) > 0.$$

It follows that for all σ sufficiently close to 1, we have

$$\frac{\partial V_D}{\partial \widetilde{\theta}} > 0.$$

Step 3. We show that for σ below a threshold, $V_D(\theta|\theta) > V_D(1|\theta)$; that is, $\widetilde{\theta} = 1$ is suboptimal. Note that the exponents on $(1-\sigma)$ are unbounded at $\widetilde{\theta} = 1$, and thus the last two terms are zero at $\widetilde{\theta} = 1$. It follows that

$$V_D(1|\theta) = K_1(1) = (1-\theta)\frac{\phi}{\widehat{\phi}}\frac{\lambda}{\lambda+\rho}.$$

Meanwhile, via direct substitution, we have

$$V_D(\theta|\theta) = (1-\theta)\frac{\lambda}{\lambda+\rho}(1-\sigma)^{\frac{\lambda+\rho\theta}{\lambda(1-\theta)}}.$$

Therefore, we have the following equivalences

$$\sigma < 1 - (\frac{\phi}{\widehat{\phi}})^{\frac{\lambda(1-\theta)}{\lambda+\rho\theta}} \iff (1-\sigma)^{\frac{\lambda+\rho\theta}{\lambda(1-\theta)}} > \frac{\phi}{\widehat{\phi}} \iff V_D(\theta|\theta) > V_D(1|\theta).$$

Step 4. We show that if σ is sufficiently close to 0, then $V_D(\widetilde{\theta}|\theta) < V_D(\theta|\theta)$ for all $\widetilde{\theta} < \theta$. Consider the difference in payoffs,

$$\begin{aligned}
&V_D(\widetilde{\theta}|\theta) - V_D(\theta|\theta) = \\
&K_1(\widetilde{\theta}) + K_2(\widetilde{\theta})(1-\sigma)^{\frac{\rho\widetilde{\theta}}{\lambda(1-\widetilde{\theta})}} + K_3(\widetilde{\theta})(1-\sigma)^{\frac{\lambda+\rho\widetilde{\theta}}{\lambda(1-\widetilde{\theta})}} - (1-\theta)\frac{\lambda}{\lambda+\rho}(1-\sigma)^{\frac{\lambda+\rho\theta}{\lambda(1-\theta)}}.
\end{aligned}$$

For the rest of the proof, we treat the payoff difference as a function of σ, and therefore, the exponents are constants. To simplify notation, let

$$a_2 \equiv \frac{\rho\widetilde{\theta}}{\lambda(1-\widetilde{\theta})} \qquad a_3 \equiv \frac{\lambda+\rho\widetilde{\theta}}{\lambda(1-\widetilde{\theta})} \qquad a_4 \equiv \frac{\lambda+\rho\theta}{\lambda(1-\theta)}$$

Treating σ as the variable, perform a second order Taylor series expansion,

$$\begin{aligned}
&V_D(\widetilde{\theta}|\theta) - V_D(\theta|\theta) = \\
&K_1(\widetilde{\theta}) + K_2(\widetilde{\theta})(1 - a_2\sigma + \frac{a_2(a_2-1)}{2}\sigma^2) + K_3(\widetilde{\theta})(1 - a_3\sigma + \frac{a_3(a_3-1)}{2}\sigma^2) \\
&- (1-\theta)\frac{\lambda}{\lambda+\rho}(1 - a_4\sigma + \frac{a_4(a_4-1)}{2}\sigma^2).
\end{aligned}$$

With some algebraic manipulation, we have

$$K_1(\widetilde{\theta}) + K_2(\widetilde{\theta}) + K_3(\widetilde{\theta}) - (1-\theta)\frac{\lambda}{\lambda+\rho} = 0$$

$$K_2(\widetilde{\theta})a_2 + K_3(\widetilde{\theta})a_3 - (1-\theta)\frac{\lambda}{\lambda+\rho}a_4 = 0$$

$$K_2(\widetilde{\theta})a_2(a_2-1) + K_3(\widetilde{\theta})a_3(a_3-1) - (1-\theta)\frac{\lambda}{\lambda+\rho}a_4(a_4-1) = -\frac{\theta-\widetilde{\theta}}{\lambda(1-\widetilde{\theta})^2}\{\lambda+\theta\rho-(\rho(1-\frac{\phi}{\widehat{\phi}})+\lambda+\theta\frac{\phi}{\widehat{\phi}})\widetilde{\theta}\}.$$

Thus, the zero and first order terms of the Taylor expansion are 0. Therefore, up to second order

$$V_D(\widetilde{\theta}|\theta) - V_D(\theta|\theta) \approx -\frac{\theta-\widetilde{\theta}}{2\lambda(1-\widetilde{\theta})^2}\{\lambda+\theta\rho-(\rho(1-\frac{\phi}{\widehat{\phi}})+\lambda+\theta\frac{\phi}{\widehat{\phi}})\widetilde{\theta}\}\sigma^2.$$

Therefore, for small values of σ, the sign of the payoff difference is determined by the sign of the coefficient,

$$-\underbrace{\frac{\theta-\widetilde{\theta}}{2\lambda(1-\widetilde{\theta})^2}}_{\text{A}}\underbrace{\left(\lambda+\theta\rho-(\rho(1-\frac{\phi}{\widehat{\phi}})+\lambda+\theta\frac{\phi}{\widehat{\phi}})\widetilde{\theta}\right)}_{\text{B}}.$$

For $\widetilde{\theta} < \theta$, term A is positive. Consider term B. Note that for $\widetilde{\theta} < \theta$ we have

$$\begin{aligned}&\lambda+\theta\rho-(\rho(1-\frac{\phi}{\widehat{\phi}})+\lambda+\theta\frac{\phi}{\widehat{\phi}})\widetilde{\theta} >\\ &\lambda+\theta\rho-(\rho(1-\frac{\phi}{\widehat{\phi}})+\lambda+\theta\frac{\phi}{\widehat{\phi}})\theta =\\ &(1-\theta)(\lambda+\theta\rho\frac{\phi}{\widehat{\phi}}) > 0.\end{aligned}$$

Thus, term B is also positive. Thus, the coefficient in question is negative and the payoff difference is negative for σ sufficiently small.

The first claim of the proposition is proved in Step 2. The second claim follows from Steps 3 and 4, which show that the optimal $\widetilde{\theta} \in [\theta, 1)$ and Proposition A.1, which implies that $\widetilde{\theta} = \theta$ is suboptimal. The third claim is proved in Step 3. □

Proof of Proposition 6.3. . The result follows immediately from Step 3 in the proof of Proposition 6.2, coupled with the observation that $V_D(\cdot|\theta)$ is continuous in $\widetilde{\theta}$. □

A.3 Proofs for Auditing

Proof of Lemma 7.1. Suppose the principal audits a submission. She will obviously reject conditional on failing the test. If the principal weakly prefers rejection conditional on passing the test, then she should have rejected the project upon submission, without incurring the auditing cost k, because rejection is optimal regardless of the test outcome. □

Proof of Lemma 7.2. Suppose that there were some time interval (t_1, t_2) over which $p(t) = 1$. For such times, a real arrival is always tested and approved, while a fake is approved whenever it generates a false pass. Therefore, for $t \in (t_1, t_2)$, the agent's expected payoff of waiting until t to submit a fake is

$$u_A(t) = \int_0^t \exp\{-(\lambda+\rho)s\}\lambda ds + \exp\{-(\lambda+\rho)t\}(1-\alpha-\phi).$$

By implication,

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

Thus, for $\alpha \neq \widehat{\phi}-\phi$, the agent's payoff function is strictly monotonic on (t_1, t_2). By implication, times $t \in (t_1, t_2)$ are suboptimal cheating times, and hence $g(t) = 1$. Thus, $p(t) = 0$ at such times. □

Proof of Lemma 7.3. We prove several steps and then combine these to prove the claims in the lemma.

Let $a_R(t) \equiv a(t) + p(t)$, representing the probability that a real arrival is accepted given the principal's strategy, and let $a_F(t) \equiv a(t) + p(t)(1-\alpha)$, representing the probability that a fake arrival is accepted given the principal's strategy. The agent's expected payoff of waiting until time t to submit a fake is

$$u_A(t) \equiv \int_0^t \exp\{-(\rho+\lambda)s\}\lambda a_R(s)ds + \exp\{-(\rho+\lambda)t\}(a_F(t)-\phi).$$

Step 1. We show that $a_F(t) \geq \phi$ for all $t \geq 0$. By way of contradiction, consider some $t \geq 0$ and suppose that $a_F(t) < \phi$. Because $a_F(t) - \phi < 0$, and $\int_t^\infty \exp\{-(\rho+\lambda)s\}a_R(s)ds \geq 0$, submitting a fake at time t is worse for the strategic agent than never submitting a fake,

$$\int_0^t \exp\{-(\rho+\lambda)t\}\lambda a_R(s)ds + \exp\{-(\rho+\lambda)t\}(a_F(t)-\phi) < \int_0^\infty \exp\{-(\rho+\lambda)t\}\lambda a_R(s)ds.$$

Then, $\nu(t) = 0$ which implies $g(t) = 1$, and hence $a_F(t) = 1$ by (11), contradicting $\phi < 1$.

Step 2. We show $f(t) = 0 \implies a(t) = 1$. This follows from $f(t) = 0 \implies \nu(t) = 0 \implies g(t) = 1 \implies a(t) = 1$, by (11).

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

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

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 (11) $a(t') = 1$.

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

Step 4. We show that if $t < \bar{t}_A$, then $a_F(t) \in [\phi, 1)$ and $f(t) > 0$. Suppose $t < \bar{t}_A$. Note that $a_F(t) = a(t) + p(t)(1-\alpha) \leq a(t) + (1 - a(t))(1-\alpha) = 1 - (1-a(t))\alpha$. From Step 3 and the definition of $\bar{t}_A$, we have $a(t) < 1$. Hence, $a_F(t) < 1$. That $a_F(t) \geq \phi$ follows from Step 1. To show that $f(t) > 0$, note first that $a_F(t) < 1$ implies $a(t) < 1$. From the principal's indifference condition (11), we get $g(t) \leq \theta_1$, and hence, $\nu(t) \geq \lambda(1-\theta_1)/\theta_1 > 0$. 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 $\nu(t) = \infty$, and hence, $g(t) = 0$. By implication $a(t) = p(t) = 0$, yielding $a_F(t) = 0$, contradicting Step 1.

Step 6. We show that $\bar{t}_A \in (0, \infty)$. As in the main model, $\phi < \widehat{\phi}$ implies that if the principal approves all submissions, then the agent fakes at $t = 0$, but then $a(0) = 0$ is sequentially rational for the principal, a contradiction.

Next, suppose that $\bar{t}_A = \infty$. From Step 4, we have $a_F(t) \in (0, 1)$ for all t. From (11), we have $g(t) \leq \theta_1$ for all t. Hence, $\nu(t) \geq \lambda(1-\theta_1)/\theta_1$ for all t. By implication, $f(t) = (1 - \sigma F(t))\lambda(1-\theta_1)/(\sigma\theta_1) \geq (1-\sigma)\lambda(1-\theta_1)/(\sigma\theta_1) > 0$. 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 Step 3 and Step 6.

Step 7. We show that $a_F(\cdot)$ is continuous at $\bar{t}_A$ and further that $\lim_{t \to \bar{t}_A} a(t) = 1$ and

$\lim_{t\to\bar{t}_A} p(t) = 0$. Note that for $t > \bar{t}_A$ we have $\lim_{t\to\bar{t}_A} a_F(t) = a_F(\bar{t}_A) = 1$. We seek to show that for $t < \bar{t}_A$, we have $\lim_{t\to\bar{t}_A} a_F(t) = 1$. Consider $t < \bar{t}_A$. From Step 4 we have $f(t) > 0$, and thus $u(t) \geq u(\bar{t}_A)$. Hence,

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

Because $a_R(\cdot)$ is bounded, in the limit as $t \to \bar{t}_A$, we have

$$\lim_{t\to\bar{t}_A} \{u(\bar{t}_A) - u(t)\} = \exp(-(\rho+\lambda)\bar{t}_A)(1 - \lim_{t\to\bar{t}_A} a_F(t)) \leq 0.$$

Because $a_F(t) \leq 1$ for all t, we have $\lim_{t\to\bar{t}_A} a_F(t) = 1$. By implication, $\lim_{t\to\bar{t}_A}[a(t) + (1-\alpha)p(t)] = 1$. Because $a(\cdot)$ and $p(\cdot)$ are probabilities, we have $\lim_{t\to\bar{t}_A} a(t) = 1$ and $\lim_{t\to\bar{t}_A} p(t) = 0$.

Step 8. We show that $a_F(t)$ is continuous for $t < \bar{t}_A$. Let $t, t' < \bar{t}_A$. Because $t, t' < \bar{t}_A$, from Claim (i) we have $f(t), f(t') > 0$. Hence, $u_A(t') = u_A(t)$. Therefore,

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

Because the integrand above is bounded, taking the limit as $t' \to t$ yields $a_F(t') \to a_F(t)$. Hence, $a_F(\cdot)$ is continuous at $t < \bar{t}_A$.

Step 9. We show that $p(t) > 0$ for $t < \bar{t}_A$. Consider $t < \bar{t}_A$. From Step 4, we have $a_F(t) \in [\phi, 1)$. By implication, $g(t) \in [\theta_0, \theta_1]$. If $g(t) \in (\theta_0, \theta_1)$, then $p(t) = 1$ from (11). Suppose $g(t) = \theta_1$. From (11), we have $a(t) + p(t) = 1$. Note that $a_F(t) < 1 \Rightarrow a(t) < 1$. Thus, $p(t) > 0$. Finally, suppose $g(t) = \theta_0$. From (11), we have $r(t) + p(t) = 1$. Thus, $a_F(t) \geq \phi \Rightarrow p(t) \geq \phi$.

Step 10. We show that for $t < \bar{t}_A$, (i) $a_F(t) > 1-\alpha$ if and only if $p(t) \in (0,1)$ and $a(t) + p(t) = 1$, (ii) $a_F(t) = 1-\alpha$ if and only if $p(t) = 1$, (iii) $a_F(t) < 1-\alpha$, if and only if $a(t) = 0$ and $p(t) \in (0,1)$. From the definition of $\bar{t}_A$, we have $a(t) < 1$ for $t < \bar{t}_A$ and from Step 9, we have $p(t) > 0$. From (11), when $k < k^*$ three sequentially rational strategies are possible for the principal: $a(t) + p(t) = 1$ with $0 < p(t) < 1$, $p(t) = 1$, or $p(t) + r(t) = 1$ with $0 < p(t) < 1$. If the principal's best response is $a(t) + p(t) = 1$ with $0 < p(t) < 1$, then $a_F(t) = 1 - p(t) + p(t)(1-\alpha) = 1 - p(t)\alpha$. Because $0 < p(t) < 1$, we have $1 - \alpha < a_F(t) < 1$. If the principal's best response is $p(t) = 1$, then $a_F(t) = 1 - \alpha$. If the principal's best response is $p(t) + r(t) = 1$ with $0 < p(t) < 1$, then $a_F(t) = p(t)(1 - \alpha)$. Because $p(t) < 1$, we have $a_F(t) < 1 - \alpha$. Note further that from the preceding calculations, (i) $a_F(t) \in (1 - \alpha, 1)$ is possible only if the principal's best response has $a(t) + p(t) = 1$ and $p(t) < 1$, (ii) $a_F(t) = 1$ is possible only if the principal's best response has $p(t) = 1$, (iii) $a_F(t) < 1 - \alpha$ is possible only if the principal's best response has $0 < p(t) < 1$ and $a(t) = 0$.

Step 11. We show that for $t < \bar{t}_A$, functions $a(\cdot)$ and $p(\cdot)$ are continuous. This follows from Steps 8 and 10.

Step 12. Consider $t_1 < t_2 < \bar{t}_A$, such that $a_F(t) \geq 1 - \alpha$ for all $t \in (t_1, t_2)$. We show that $a_F(\cdot)$ is differentiable and is either strictly increasing, strictly decreasing, or constant on (t_1, t_2). From Step 10, we have that $a(t) + p(t) = 1$ for all $t \in (t_1, t_2)$. By implication, $a_R(t) = 1$ and $a_F(t) = 1 - \alpha p(t)$ for all such t. Consider $t, t' \in (t_1, t_2)$. We have

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

Divide by $t' - t$, and consider $t \to t'$. We have

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

It follows that $a_F(\cdot)$ is differentiable at t'. In addition,

$$\begin{aligned}
& a_F'(t') = (\rho+\lambda)(a_F(t') - \phi) - \lambda \\
\iff \quad & a_F'(t') = (\rho+\lambda)[a_F(t') - \phi - (1 - \widehat{\phi})] \\
\iff \quad & a_F'(t') = (\rho+\lambda)[a_F(t') - (1 - \widehat{\phi} + \phi)].
\end{aligned}$$

Thus, whenever $a_F(t) > 1 - \widehat{\phi} - \phi$, we have $a_F'(t) > 0$, and hence, $a_F(\cdot)$ is increasing at t. Conversely, whenever $a_F(t) < 1 - \widehat{\phi} - \phi$, we have $a_F'(t) < 0$, and hence, $a_F(\cdot)$ is decreasing at t. Furthermore, whenever $a_F(t) = 1 - \widehat{\phi} - \phi$, we have $a_F'(t) = 0$, and hence, $a_F(\cdot)$ is constant at t. It follows that the derivative of $a_F(\cdot)$ cannot change sign on (t_1, t_2), and therefore $a_F(\cdot)$ is monotone.

Step 13. Consider $t_1 < t_2 < \bar{t}_A$, such that $a_F(t) \leq 1 - \alpha$ for all $t \in [t_1, t_2]$. We show that $a_F(\cdot)$ is differentiable and either strictly increasing, strictly decreasing, or constant on $[t_1, t_2]$. Proof is analogous to Step 12.

Step 14. We show that $a_F(\cdot)$ is strictly increasing on $[0, \bar{t}_A)$. First, we show that there exists a subinterval of $[0, \bar{t}_A)$ over which $a_F(\cdot)$ is strictly increasing. From Step 4, we have that $a_F(t) < 1$ for $t < \bar{t}_A$. Furthermore, from Step 7, we have $a_F(t) \to 1$ as $t \to \bar{t}_A$. By implication, there exists an interval (t_1, t_2) such that $a_F(t) > 1-\alpha$ for all (t_1, t_2) and $a_F(t_2) > a_F(t_1)$. From Step 12, $a_F(\cdot)$ must be strictly increasing on (t_1, t_2). Next we show that $a_F(\cdot)$ cannot switch from being strictly increasing to either strictly decreasing or constant. Suppose that there exists some t at which $a_F(\cdot)$ switches from strictly increasing to strictly decreasing or constant. Suppose that $a_F(t) > 1-\alpha$ (the other cases are analogous). By continuity of $a_F(\cdot)$, there exists an interval (t_1', t_2') with $t_1' < t < t_2'$, such that $a_F(t) \geq 1-\alpha$ for all $t \in (t_1', t_2')$. By Step 12, $a_F(\cdot)$ is either strictly increasing, strictly decreasing, or constant on (t_1', t_2'), and thus no switch is possible. A similar argument rules out a switch from $a_F(\cdot)$ constant or strictly decreasing to strictly increasing. Thus, we have shown that $a_F(\cdot)$ is strictly increasing on a subinterval of $[0, \bar{t}_A)$, and that it cannot switch from strictly increasing to decreasing or constant at any $t \in [0, \bar{t}_A)$.

Suppose $a_F(t) < 1 - \alpha$ for some $t \geq 0$. Let $\widetilde{t}_A$ denote the unique value of t for which $a_F(\widetilde{t}_A) = 1 - \alpha$ (existence and uniqueness of $\widetilde{t}_A$ is guaranteed because $a_F(\cdot)$ is continuous and strictly increasing). If $a_F(t) \geq 1 - \alpha$ for all t, then let $\widetilde{t}_A = 0$.

Step 15. We show that (i) $a(t) > 0$ for all $t > \widetilde{t}_A$ and (ii) if $\widetilde{t}_A > 0$, then $a(t) = 0$ for all $t \leq \widetilde{t}_A$. Claim (i): From the definition of $\widetilde{t}_A$ and Step 14, we have $t > \widetilde{t}_A \Rightarrow a_F(t) > 1-\alpha$. From Step 10, $a_F(t) > 1-\alpha \Rightarrow a(t)+p(t) = 1$ and $p(t) \in (0,1)$. Hence, $a(t) > 0$. Claim (ii): Suppose $\widetilde{t}_A > 0$. From the definition of $\widetilde{t}_A$ and Step 14, we have $t \in [0, \widetilde{t}_A) \Rightarrow a_F(t) < 1-\alpha$. From Step 10, $a_F(t) < 1 - \alpha \Rightarrow a(t) = 0$.

Step 16. We show that $\widetilde{t}_A < \bar{t}_A$. If $\widetilde{t}_A = 0$, then the claim follows from Step 6. Suppose $\widetilde{t}_A > 0$. From Step 7, we have $\lim_{t \to \bar{t}_A} a(t) = 1$. By implication, there exists ϵ_1 such that $a(\bar{t}_A - \epsilon) > 0$ for all $\epsilon \in (0, \epsilon_1)$. From Step 15 it follows that $\bar{t}_A - \epsilon_1 > \widetilde{t}_A$, and hence, $\bar{t}_A > \widetilde{t}_A$.

Step 17. We show that $p(t)$ is differentiable and decreasing for $t \in (\widetilde{t}_A, \bar{t}_A)$ and that for such t, $a(t) + p(t) = 1$ and $a(t) > 0$. For such t, Step 15 implies $a(t) > 0$, and Step 4 implies $a_F(t) < 1$, which in turn implies $a(t) < 1$. Thus, we have $a(t) \in (0,1)$. From Step 9, we have $p(t) > 0$ for such t. From (11), for $k < k^*$, $a(t) \in (0,1)$ and $p(t) \in (0,1)$ together imply $a(t) + p(t) = 1$. From Step 10, we have $a_F(t) > 1 - \alpha$ for such t. From Steps 12 and 14, we have that $a_F(\cdot)$ is differentiable and strictly increasing for such t. Observing that $a_F(t) = 1 - p(t) + p(t)(1 - \alpha) = 1 - \alpha p(t)$ for such t, we have that $p(t)$ is also differentiable and strictly decreasing for $t \in (\widetilde{t}_A, \bar{t}_A)$.

Proof of (iii). From Step 11 and 17, we have that $p(\cdot)$ is continuous, strictly decreasing and differentiable for $t \in (\widetilde{t}_A, \bar{t}_A)$. From Step 7, we have that $p(t) \to 0$ as $t \to \bar{t}_A$. Furthermore, from Step 17, we have that $a(t) + p(t) = 1$, with $a(t) > 0$ for $t \in (\widetilde{t}_A, \bar{t}_A)$.

Step 18. Suppose $\widetilde{t}_A > 0$. We show that $\lim_{t \to \widetilde{t}_A^+} p(t) = p(\widetilde{t}_A) = 1$. From Step 15 (ii), we have that $a(t) = 0$ for $t \in [0, \widetilde{t}_A)$. From Step 12, we have $a_F(t) \leq 1 - \alpha$ for $t \in [0, \widetilde{t}_A)$, and thus $\lim_{t \to \widetilde{t}_A^-} a_F(t) \leq 1 - \alpha$. From Step 15 (i) we have that $a(t) > 0$ for $t > \widetilde{t}_A$. From Step 12, we have $a_F(t) > 1 - \alpha$ for $t > \widetilde{t}_A$, and thus, $\lim_{t \to \widetilde{t}_A^+} a_F(t) \geq 1 - \alpha$. Because $a_F(\cdot)$ is continuous (Step 8), we have $1 - \alpha \leq \lim_{t \to \widetilde{t}_A^+} a_F(t) = a_F(\widetilde{t}_A) = \lim_{t \to \widetilde{t}_A^-} a_F(t) \leq 1 - \alpha$, and thus $\lim_{t \to \widetilde{t}_A^+} a_F(t) = a_F(\widetilde{t}_A) = 1 - \alpha$. From Step 10, we have $\lim_{t \to \widetilde{t}_A^+} p(t) = p(\widetilde{t}_A) = 1$.

Step 19. Suppose $\widetilde{t}_A > 0$. We show that $p(\cdot)$ is differentiable and strictly increasing for $t \in [0, \widetilde{t}_A)$. From Step 15 (ii), we have that $a(t) = 0$ for $t \in [0, \widetilde{t}_A)$. From Step 12, we have $a_F(t) \leq 1 - \alpha$ for $t \in [0, \widetilde{t}_A)$. From Steps 13 and 14, we have that $a_F(\cdot)$ is differentiable and strictly increasing on $[0, \widetilde{t}_A)$. Because $a_F(t) < 1 - \alpha$ for such t, Step 12 implies $a(t) = 0$ and $p(t) \in (0, 1)$. By implication, $a_F(t) = p(t)(1 - \alpha)$ for such t. Because $a_F(\cdot)$ is strictly increasing and differentiable on $[0, \widetilde{t}_A)$, so is $p(\cdot)$.

Proof of (iv). Suppose $\widetilde{t}_A > 0$. From Step 19, we have that $p(\cdot)$ is differentiable and strictly increasing for $t \in [0, \widetilde{t}_A)$. From Step 18, we have that $\lim_{t \to \widetilde{t}_A^+} p(t) = p(\widetilde{t}_A) = 1$. By implication, $t < \widetilde{t}_A \Rightarrow p(t) < 1$. Furthermore, from Step 15, we have $a(t) = 0$ for such t. Hence, for $t \in [0, \widetilde{t}_A)$ we have $p(t) + r(t) = 1$, with $r(t) > 0$. $\square$

Next, we provide a complete characterization of the two possible equilibrium structures. We begin by analyzing the one-stage equilibrium in which the entire phase of doubt involves default acceptance, corresponding to $\widetilde{t}_A = 0$ in Lemma 7.3. We then analyze the two-stage equilibrium, in which the phase of doubt consists of a stage of default rejection followed by a phase of default approval, corresponding to $\widetilde{t}_A > 0$ in Lemma 7.3.

To facilitate analysis, for $\theta_i \in \{\theta_0, \theta_1\}$ (as defined in (11)) let

$$\mu_i \equiv \lambda \frac{1 - \theta_i}{\theta_i} \qquad \bar{t}_i \equiv -\frac{\ln(1 - \sigma)}{\mu_i}.$$

Note that μ_i is the equilibrium cheating rate in the equilibrium of the baseline model when the principal's preference parameter is θ_i, and $\bar{t}_i$ is the duration of the corresponding phase of doubt. Furthermore, for a weak test let

$$\delta_A \equiv -\frac{\ln(1 - \frac{\alpha}{\widehat{\phi} - \phi})}{\lambda + \rho} \qquad \widetilde{t} \equiv \max\left\{0, \frac{\mu_1}{\mu_0}\left(\bar{t}_1 - \delta_A\right)\right\}.$$

Note that $\delta_A > 0$ if and only if the test is weak.

Proposition A.1 (One-stage auditing equilibrium.)**.** *A one stage auditing equilibrium exists if and only if (a) the test is strong or (b) the test is weak and* $\widetilde{t} = 0$*, and it is characterized below. Furthermore, no other one stage auditing equilibrium exists.*

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

$$F(t) = \frac{1}{\sigma}(1 - \exp(-\mu_1 t))$$

supported on interval $[0, \bar{t}_1]$*. If* $t \in [0, \bar{t}_1]$*, then the principal mixes between auditing and outright approval, and the probability of audit is*

$$p(t) = \frac{\widehat{\phi} - \phi}{\alpha}(1 - \exp\{-(\rho + \lambda)(\bar{t}_1 - t)\}).$$

If $t \geq \bar{t}_1$*, then the principal does not audit and accepts with probability 1.*

Beliefs. *If* $t \in (0, \bar{t}_1)$*, then* $g(t) = \theta_1$*, and* $g(t) = 1$ *otherwise.*

Payoffs. *The strategic agent's equilibrium payoff is* $U_A^S = 1 - \alpha p(0) - \phi$*. The principal's payoff is*

$$V_A = \int_0^{\bar{t}_1} \exp\{-(\rho + \frac{\lambda}{\theta_1})t\}\frac{\lambda}{\theta_1}(\theta_1 - \theta)dt + (1 - \theta)(1 - \sigma)\int_{\bar{t}_1}^{\infty} \lambda \exp\{-(\rho + \lambda)t\}dt.$$

Proof of Proposition A.1. Consider a single auditing stage in which the principal randomizes between performing the test and default approval, i.e., the scenario in which $\widetilde{t}_A = 0$ in Lemma 7.3.

Strategies. The principal's indifference requires the agent to randomize such that $g(t) = \theta_1$ for all $t \in [0, \bar{t}_A)$. This gives

$$F(t) = \frac{1}{\sigma}(1 - \kappa_1 \exp(-\mu_1 t)). \tag{A.3}$$

Because $F(0) = 0$ it follows that $\kappa_1 = 1$, and hence

$$\bar{t}_A = -\frac{\ln(1 - \sigma)}{\mu_1} = \bar{t}_1.$$

Under default approval, a real submission is never rejected. Therefore, the expected utility to the strategic agent from submitting a fake at time t if a real project does not arrive before then is

$$u_A(t) = \int_0^t \lambda \exp(-(\rho + \lambda)s)\, \mathrm{d}s + \exp(-(\rho + \lambda)t)(1 - \alpha p(t) - \phi).$$

Ergo, indifference requires

$$u'(t) = \exp(-(\rho + \lambda)t)(\lambda - (\rho + \lambda)(1 - \alpha p(t) - \phi) - \alpha p'(t)) = 0,$$

or

$$(\rho+\lambda)\alpha p(t) - \alpha p'(t) = \rho - (\rho+\lambda)\phi.$$

Solving yields

$$p(t) = \frac{\widehat{\phi} - \phi}{\alpha} + \kappa_3 \exp((\rho+\lambda)t). \tag{A.4}$$

Boundary condition, $p(\bar{t}_1) = 0$ gives

$$p(t) = \frac{\widehat{\phi} - \phi}{\alpha} \left(1 - \exp(-(\rho+\lambda)(\bar{t}_1 - t))\right).$$

This is a decreasing function as required. Moreover $p(0) \leq 1$ If $\alpha > \widehat{\phi} - \phi$, or if $\alpha < \widehat{\phi} - \phi$ and $\bar{t}_1 \leq \delta_A$, as stipulated.

Payoffs. The strategic agent is indifferent about submitting a fake project at all times $t \in [0, \bar{t}_1]$, including time zero. Thus, the strategic agent's payoff is $u_A^* = 1 - \alpha p(0) - \phi$.

Given the agent's mixed strategy $F(\cdot)$, the principal's payoff is

$$V_A = \int_0^{\bar{t}_1} \exp(-(\rho+\lambda)t)(\lambda(1-\sigma F(t))(1-\theta) - \sigma f(t)\theta)\,\mathrm{d}t + (1-\sigma)(1-\theta)\int_{\bar{t}_1}^{\infty} \lambda \exp(-(\rho+\lambda)t)\,\mathrm{d}t.$$

To see this, note first that for an arrival during $[0, \bar{t}_1]$, the principal is indifferent between accepting and auditing, and thus her payoff is as if she accepts any such arrival. Furthermore, the likelihood of a real arrival at time $t \in [0, \bar{t}_1]$ given the agent's mixed strategy is $\exp(-\lambda t)\lambda(1 - \sigma F(t))$, and the likelihood of a fraudulent arrival at such a time is $\sigma \exp(-\lambda t) f(t)$. Exploiting the principal's indifference, accounting for discounting, and using the principal's payoffs from accepting a real and fake arrival, yields the previous expression. Note that in the audit equilibrium,

$$(1 - \sigma F(t)) = \exp(-\mu_1 t) \quad \text{and} \quad \sigma f(t) = \mu_1 \exp(-\mu_1 t).$$

Thus, we have

$$\begin{aligned}
V_A &= \int_0^{\bar{t}_1} \exp(-(\rho+\lambda+\mu_1)t)(\lambda(1-\theta) - \theta\mu_1)\,\mathrm{d}t + (1-\sigma)(1-\theta)\int_{\bar{t}_1}^{\infty} \lambda \exp(-(\rho+\lambda)t)\,\mathrm{d}t \\
&= \int_0^{\bar{t}_1} \exp(-(\rho + \frac{\lambda}{\theta_1})t)(\lambda(1-\theta) - \theta\lambda\frac{1-\theta_1}{\theta_1})\,\mathrm{d}t + (1-\sigma)(1-\theta)\int_{\bar{t}_1}^{\infty} \lambda \exp(-(\rho+\lambda)t)\,\mathrm{d}t \\
&= \lambda(1 - \frac{\theta}{\theta_1})\int_0^{\bar{t}_1} \exp(-(\rho + \frac{\lambda}{\theta_1})t)\,\mathrm{d}t + (1-\sigma)(1-\theta)\int_{\bar{t}_1}^{\infty} \lambda \exp(-(\rho+\lambda)t)\,\mathrm{d}t.
\end{aligned}$$

□

Proposition A.2 (Two-stage auditing equilibrium.)**.** *A two stage auditing equilibrium exists if and only if the test is weak and* $\widetilde{t} > 0$*, and it is characterized below. Furthermore, no other two stage auditing equilibrium exists.*

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

$$F(t) = \begin{cases} \frac{1}{\sigma}(1 - \exp(-\mu_0 t)) & \text{for } t \in [0, \widetilde{t}_A) \\ \frac{1}{\sigma}(1 - \exp(-\mu_1 t - (\mu_0 - \mu_1)\widetilde{t}_A)) & \text{for } t \in [\widetilde{t}_A, \bar{t}_A] \end{cases}$$

supported on $[0, \bar{t}_A]$*, where* $\widetilde{t}_A = \widetilde{t}$ *and* $\bar{t}_A = \widetilde{t} + \delta_A$*. If* $t \in [0, \widetilde{t}_A]$*, then the principal randomizes between rejecting and auditing, and the audit probability is*

$$p(t) = \frac{\phi}{\widehat{\phi} - \alpha} + \left(1 - \frac{\phi}{\widehat{\phi} - \alpha}\right) \exp\left(-(\frac{\widehat{\phi} - \alpha}{1 - \alpha})(\rho + \lambda)(\widetilde{t}_A - t)\right).$$

For $t \in [\widetilde{t}_A, \bar{t}_A)$ *the principal randomizes between approval and auditing, and the audit probability is*

$$p(t) = \frac{\widehat{\phi} - \phi}{\alpha}(1 - \exp(-(\rho + \lambda)(\bar{t}_A - t)).$$

For $t \geq \bar{t}_A$ *the principal approves all submissions without performing an audit.*

Beliefs. *If* $t \in (0, \widetilde{t}_A)$*, then* $g(t) = \theta_0$*. If* $t \in (\widetilde{t}_A, \bar{t}_A)$*, then* $g(t) = \theta_1$*. Otherwise* $g(t) = 1$*.*

Payoffs. *The agent's equilibrium payoff is* $U_A^S = (1 - \alpha)p(0) - \phi$*. The principal's payoff is*

$$V_A = \exp(-(\mu_0 - \mu_1)\widetilde{t}_A) \int_{\widetilde{t}_A}^{\bar{t}_A} \exp(-(\rho + \frac{\lambda}{\theta_1})t)\frac{\lambda}{\theta_1}(\theta_1 - \theta)dt + (1 - \theta)(1 - \sigma) \int_{\bar{t}_A}^{\infty} \lambda \exp(-(\rho + \lambda)t)dt.$$

Proof of Proposition A.2. Consider an equilibrium in which the principal first randomizes between performing the test and default rejection and switches at time $\widetilde{t}_A$ to randomizing between performing the test and default approval, as in Lemma 7.3 with $\widetilde{t}_A > 0$. Expressions (A.3) and (A.4) remain valid for the default approval stage. So it remains to derive the agent and principal mixing functions during the default rejection stage. We will then paste the two stages together by ensuring continuity of the agent's CDF at the transition times, $\widetilde{t}_A$ and $\bar{t}_A$, combined with the boundary conditions, $p(\widetilde{t}_A) = 1$ and $p(\bar{t}_A) = 0$, established in Lemma 7.3.

Strategies and Phase Transitions. For the principal to be indifferent between auditing and rejecting we must have $g(t) = \theta_0$ for all $t \in [0, \widetilde{t}_A)$. The agent therefore mixes with constant arrival rate μ_0. Furthermore, since the equilibrium begins with this phase, we cannot have a mass point at zero. It follows that for $t \in [0, \widetilde{t}_A)$,

$$F(t) = \frac{1}{\sigma}(1 - \exp(-\mu_0 t)).$$

To find the audit probability during the default rejection stage, note that a real submission is approved iff it is tested, and hence, for $t \in [0, \widetilde{t}_A)$,

$$u_A(t) = \int_0^t \lambda \exp(-(\rho+\lambda)s)p(s)\,\mathrm{d}s + \exp(-(\rho+\lambda)t)((1-\alpha)p(t) - \phi).$$

The agent's indifference requires $u_A'(t) = 0$, or

$$\lambda p(t) - (\rho+\lambda)((1-\alpha)p(t) - \phi) + (1-\alpha)p'(t) = 0.$$

Dividing by $1-\alpha$ and combining terms renders this as

$$-\frac{(\rho+\lambda)(\widehat{\phi}-\alpha)}{1-\alpha}p(t) + p'(t) = -\frac{(\rho+\lambda)\phi}{1-\alpha}.$$

Solving yields

$$p(t) = \frac{\phi}{\widehat{\phi}-\alpha} + \kappa_2 \exp(\frac{(\rho+\lambda)(\widehat{\phi}-\alpha)}{1-\alpha}t).$$

To construct the equilibrium we must paste the two phases together ensuring continuity of the CDF, and the boundary conditions $p(\bar{t}_A) = 0$ and $p(\widetilde{t}_A) = 1$. We have the following five equations in the five unknowns κ_1, κ_2, κ_3, $\widetilde{t}$ and $\bar{t}_A$:

$$F(\bar{t}_A) = 1$$

$$\implies \qquad \frac{1}{\sigma}(1 - \kappa_1 \exp(-\mu_1 \bar{t}_A)) = 1, \tag{A.5}$$

$$\text{continuity of } F(\cdot) \text{ at } \widetilde{t}_A$$

$$\implies \qquad \frac{1}{\sigma}(1 - \exp(-\mu_0 \widetilde{t}_A)) = \frac{1}{\sigma}(1 - \kappa_1 \exp(-\mu_1 \widetilde{t}_A)), \tag{A.6}$$

$$p(\widetilde{t}_A^-) = 1$$

$$\implies \qquad \frac{\phi}{\widehat{\phi}-\alpha} + \kappa_2 \exp(\frac{(\rho+\lambda)(\widehat{\phi}-\alpha)}{1-\alpha}\widetilde{t}_A) = 1, \tag{A.7}$$

$$p(\widetilde{t}_A^+) = 1$$

$$\implies \qquad \frac{\widehat{\phi}-\phi}{\alpha} + \kappa_3 \exp((\rho+\lambda)\widetilde{t}_A) = 1, \tag{A.8}$$

$$p(\bar{t}_A) = 0$$

$$\implies \qquad \frac{\widehat{\phi}-\phi}{\alpha} + \kappa_3 \exp((\rho+\lambda)\bar{t}_A) = 0. \tag{A.9}$$

We will work with these equations sequentially, starting from the last one and working our way up. In this way we will reduce the system to one with just two linear equations and unknowns. From (A.9) we get

$$\kappa_3 = -\frac{\widehat{\phi} - \phi}{\alpha} \exp(-(\rho + \lambda)\bar{t}_A).$$

Substituting this into (A.8) gives

$$\frac{\widehat{\phi} - \phi}{\alpha} \left(1 - \exp(-(\rho + \lambda)(\bar{t}_A - \widetilde{t}_A))\right) = 1$$

$$\implies \qquad \exp(-(\rho + \lambda)(\bar{t}_A - \widetilde{t}_A)) = 1 - \frac{\alpha}{\widehat{\phi} - \phi}$$

An immediate implication is $\bar{t}_A > \widetilde{t}_A \iff \alpha < \widehat{\phi} - \phi$. Because $\bar{t}_A > \widetilde{t}_A$ is necessary for the equilibrium, we maintain the assumption that $\alpha < \widehat{\phi} - \phi$. Taking the natural log, we get

$$\bar{t}_A - \widetilde{t}_A = -\frac{\ln\left(1 - \frac{\alpha}{\widehat{\phi} - \phi}\right)}{\rho + \lambda} = \delta_A. \tag{A.10}$$

We will return to this equation later. From (A.7) we get

$$\kappa_2 = \left(1 - \frac{\phi}{\widehat{\phi} - \alpha}\right) \exp\left(-\frac{(\rho + \lambda)(\widehat{\phi} - \alpha)}{1 - \alpha}\widetilde{t}_A\right).$$

Observe that $\alpha < \widehat{\phi} - \phi$ implies $\kappa_2 > 0$. From (A.6) we get

$$\kappa_1 = \exp(-(\mu_0 - \mu_1)\widetilde{t}_A).$$

Note that $\kappa_1 > 0$. Furthermore, $\mu_0 > \mu_1$ implies $\kappa_1 < 1$ whenever $\widetilde{t}_A > 0$. Substituting κ_1 into (A.5), rearranging, and taking logs gives

$$(\mu_0 - \mu_1)\widetilde{t}_A + \mu_1 \bar{t}_A = -\ln(1 - \sigma). \tag{A.11}$$

Solving (A.10) and (A.11) for $\widetilde{t}_A$ and $\bar{t}_A$ yields the desired expressions.

We show that $\widetilde{t}_A > 0 \iff \bar{t}_1 > \delta_A$.

$$\widetilde{t}_A > 0 \iff \bar{t}_0 - \frac{\mu_1}{\mu_0}\delta_A > 0 \iff \frac{\mu_1}{\mu_0}(\bar{t}_1 - \delta_A) > 0 \iff \bar{t}_1 > \delta_A,$$

and thus the preceding is necessary and sufficient for the existence of the two stage auditing equilibrium.

We show that $\widetilde{t}_A > 0 \Rightarrow \bar{t}_A < \bar{t}_1$. Note that

$$\widetilde{t}_A > 0 \Rightarrow \bar{t}_0 > \frac{\mu_1}{\mu_0}\delta_A \Rightarrow \bar{t}_1 > \delta_A \Rightarrow \bar{t}_1(1 - \frac{\mu_1}{\mu_0}) > (1 - \frac{\mu_1}{\mu_0})\delta_A \Rightarrow$$
$$-\ln(1 - \sigma)\frac{\mu_0 - \mu_1}{\mu_1\mu_0} > (1 - \frac{\mu_1}{\mu_0})\delta_A \Rightarrow -\ln(1 - \sigma)(\frac{1}{\mu_1} - \frac{1}{\mu_0}) > (1 - \frac{\mu_1}{\mu_0})\delta_A \Rightarrow$$
$$\bar{t}_1 - \bar{t}_0 > (1 - \frac{\mu_1}{\mu_0})\delta_A \Rightarrow \bar{t}_1 > \bar{t}_0 + (1 - \frac{\mu_1}{\mu_0})\delta_A = \bar{t}_A.$$

Payoffs. The strategic agent is indifferent about submitting a fake project at all times $t \in [0, \bar{t}_A]$, including time zero. Thus, the strategic agent's payoff is $u_A^* = (1-\alpha)p(0) - \phi$.

Following similar logic to Proposition A.1, given the agent's mixed strategy $F(\cdot)$, the principal's payoff is

$$V_A = \int_{\widetilde{t}_A}^{\bar{t}_A} \exp(-(\rho+\lambda)t)(\lambda(1-\sigma F(t))(1-\theta) - \sigma f(t)\theta)\,\mathrm{d}t + (1-\sigma)(1-\theta)\int_{\widetilde{t}_A}^{\infty} \lambda \exp(-(\rho+\lambda)t)\,\mathrm{d}t.$$

Note that in the two stage equilibrium, the principal is indifferent between auditing and rejecting for arrivals at $t \in [0, \widetilde{t}_A)$. In the two stage audit equilibrium, for $t \in [\widetilde{t}_A, \bar{t}_A)$,

$$(1-\sigma F(t)) = \exp(-(\mu_0-\mu_1)\widetilde{t}_A)\exp(-\mu_1 t) \quad \text{and} \quad \sigma f(t) = \exp(-(\mu_0-\mu_1)\widetilde{t}_A)\mu_1 \exp(-\mu_1 t),$$

which implies that the principal's payoff is

$$\exp(-(\mu_0-\mu_1)\widetilde{t}_A)\int_{\widetilde{t}_A}^{\bar{t}_A} \exp(-(\rho+\lambda+\mu_1)t)(\lambda(1-\theta)-\theta\mu_1)\,\mathrm{d}t + (1-\theta)(1-\sigma)\int_{\bar{t}_A}^{\infty} \lambda\exp(-(\rho+\lambda)t)\,\mathrm{d}t.$$

Substituting μ_1 inside the integrand and simplifying yields

$$V_A = \exp(-(\mu_0-\mu_1)\widetilde{t}_A)(1-\frac{\theta}{\theta_1})\int_{\widetilde{t}_A}^{\bar{t}_A} \exp(-(\rho+\frac{\lambda}{\theta_1})t)\,\mathrm{d}t + (1-\sigma)(1-\theta)\int_{\bar{t}_A}^{\infty} \lambda\exp(-(\rho+\lambda)t)\,\mathrm{d}t$$

□

Proof of Proposition 7.1. We use the characterization of all possible equilibria with the one-stage structure (Proposition A.1) and all equilibria with the two-stage structure (Proposition A.2) throughout the proof.

Proof of (i). From Proposition A.1, the one-stage equilibrium exists if either (a) the test is strong or (b) the test is weak and $\widetilde{t} = 0$. From Proposition A.2, the two-stage equilibrium exists if the test is weak and $\widetilde{t} > 0$. Given our focus on tests that are either strong or weak, these conditions are mutually exclusive and exhaustive. Thus, for all parameters under consideration an equilibrium exists and is unique.

Step 1. We show that the condition $\widetilde{t} > 0 \iff \sigma > \sigma^*$, characterizing σ^* exactly. Note that

$$\widetilde{t} > 0 \iff \bar{t}_1 > \delta_A \Leftrightarrow -\ln(1-\sigma) > \mu_1\delta_A \iff \sigma > \sigma^* \equiv 1 - \exp(-\mu_1\delta_A).$$

Proof of (ii). From Proposition A.1, the one-stage equilibrium exists if and only if either (a) the test is strong or (b) the test is weak and $\widetilde{t} = 0$. Using Step 1, the result follows.

Proof of (iii). To prove the first part of (iii), from Proposition A.2, the two-stage equilibrium exists if and only if the test is weak and $\widetilde{t} > 0$. Using Step 1, the result follows. To prove the second part of (iii), note that the principal's belief that the agent is ethical, given no submission at time or before time t is

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

Indeed, the denominator is the probability of no submission at or before t, while the numerator is the probability of no submission at or before t and an ethical agent. From Proposition A.2 and continuity $F(\cdot)$ (Lemma 7.3) at the transition time $\widetilde{t}_A = \widetilde{t}$, we have

$$F(\widetilde{t}_A) = \frac{1}{\sigma}(1-\exp(-\mu_0\widetilde{t})) \Rightarrow 1-\sigma F(\widetilde{t}_A) = \exp(-\mu_0\widetilde{t}).$$

By assumption, $\widetilde{t} > 0$. Substituting, we have

$$\exp(-\mu_0\widetilde{t}) = \exp(-\mu_1(\bar{t}_1-\delta_A)) = \exp(\mu_1\delta_A)\exp(-\mu_1\bar{t}_1) = \exp(\mu_1\delta_A)(1-\sigma).$$

By implication, the belief that the agent is ethical at $\widetilde{t}$ is $\exp(-\mu_1\delta_A) = 1-\sigma^*$. The claim follows. □

Proof of Proposition 7.2. We consider the principal's, ethical agent's, and strategic agent's payoffs in the one-stage and two-stage auditing equilibrium.

Principal, one-stage equilibrium. From Proposition A.1, we have that the principal's payoff in the one stage auditing equilibrium is

$$V_A = \lambda(1-\frac{\theta}{\theta_1})\int_0^{\bar{t}_1}\exp(-(\rho+\frac{\lambda}{\theta_1})t)\,\mathrm{d}t + (1-\sigma)(1-\theta)\int_{\bar{t}_1}^{\infty}\lambda\exp(-(\rho+\lambda)t)\,\mathrm{d}t.$$

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 $\theta_1 = \theta$, the two expressions are equal, i.e. $V_A = V$. Note further that the one stage auditing equilibrium exists only if $k < k^* \iff \theta_1 > \theta$. Differentiating with respect to θ_1, we have

$$\frac{dV_A}{d\theta_1} = \frac{\lambda\theta}{\theta_1^2}\int_0^{\bar{t}_1}\exp(-(\rho+\frac{\lambda}{\theta_1})t)\,\mathrm{d}t$$
$$+\lambda(1-\frac{\theta}{\theta_1})[\exp(-(\rho+\frac{\lambda}{\theta_1})\bar{t}_1)\frac{d\bar{t}_1}{d\theta_1} + \int_0^{\bar{t}_1}\exp(-(\rho+\frac{\lambda}{\theta_1})t)\frac{\lambda}{\theta_1^2}tdt] - (1-\sigma)(1-\theta)\lambda\exp(-(\rho+\lambda)\bar{t}_1)\frac{d\bar{t}_1}{d\theta_1}.$$

Note that xxx

$$(1-\sigma)\exp(-(\rho+\lambda)\bar{t}_1) = \exp((1+(\rho+\lambda)\frac{\theta_1}{\lambda(1-\theta_1)})\ln(1-\sigma))$$
$$= \exp(-(\rho+\frac{\lambda}{\theta_1})(-\frac{\theta_1}{\lambda(1-\theta_1)})\ln(1-\sigma)) = \exp(-(\rho+\frac{\lambda}{\theta_1})\bar{t}_1),$$

and

$$\frac{d\bar{t}_1}{d\theta_1} = -\frac{\lambda(1-\theta_1)+\lambda\theta_1}{\lambda^2(1-\theta_1)^2}\ln(1-\sigma) = -\frac{1}{\lambda(1-\theta_1)^2}\ln(1-\sigma) = \frac{\bar{t}_1}{\theta_1(1-\theta_1)}.$$

Substituting and simplifying, we have

$$\frac{dV_A}{d\theta_1} = \frac{\lambda\theta}{\theta_1^2}\int_0^{\bar{t}_1}\exp(-(\rho+\frac{\lambda}{\theta_1})t)\,\mathrm{d}t - \frac{\lambda\theta}{\theta_1^2}\exp(-(\rho+\frac{\lambda}{\theta_1})\bar{t})\bar{t} + \lambda(\frac{\theta_1-\theta}{\theta_1})\int_0^{\bar{t}_1}\exp(-(\rho+\frac{\lambda}{\theta_1})t)\frac{\lambda}{\theta_1^2}tdt.$$

Noting that the last integral is strictly positive, we have

$$\frac{dV_A}{d\theta_1} > \frac{\lambda\theta}{\theta_1^2}[\int_0^{\bar{t}_1}\exp(-(\rho+\frac{\lambda}{\theta_1})t)\,\mathrm{d}t - \exp(-(\rho+\frac{\lambda}{\theta_1})\bar{t})\bar{t}] > 0,$$

where the last inequality follows because $\exp(-(\rho+\lambda/\theta_1)t)$ is a decreasing function, and thus, its average value on interval $[0,\bar{t}_1]$ is larger than its value at the right endpoint. Because (i) V_A is increasing in θ_1, (ii) $\theta_1 > \theta$ whenever the one-stage auditing equilibrium exists, and (iii) $\theta_1 = \theta \Rightarrow V_A = V$, we have that in the one-stage auditing equilibrium $V_A > V$.

Principal, two-stage equilibrium. We prove that the principal's payoff is higher in the two stage auditing equilibrium than in the baseline model in three steps.

In Step 1 we show that the principal's payoff in the two-stage equilibrium approaches her payoff in the one-stage equilibrium as $\sigma \to \sigma^{*+}$. Because we showed above that the principal's payoff is strictly higher in the one stage auditing 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 $\sigma \in (\sigma^*, \sigma^* + \epsilon)$.

In Step 2 we show that if the principal's payoff is higher in the two-stage auditing equilibrium than in the baseline model for some 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 auditing equilibrium than in the baseline model.

Step 1: We show that for any σ such that the two-stage auditing equilibrium exists, there exists $\sigma' < \sigma$ such that (i) the two-stage auditing equilibrium exists at σ', and (ii) at σ' the principal's payoff in the two-stage auditing equilibrium is higher than her payoff in the baseline model.

Consider parameters at which the two-stage auditing equilibrium exists; by Proposition A.2, we have $\alpha < \widehat{\phi} - \phi$ and $\sigma > \sigma^*$. By implication, if $\sigma > \sigma^*$, then the two stage auditing equilibrium exists for all $\sigma \in (\widetilde{\sigma}, \sigma)$.

Next, we argue that as $\sigma \to \sigma^{*+}$, the principal's payoff in the two-stage auditing equilibrium approaches her payoff in the one-stage auditing equilibrium. Note that in the two-stage equilibrium, the principal's payoff is

$$V_A = \exp(-(\mu_0-\mu_1)\widetilde{t}_A)(1-\frac{\theta}{\theta_1})\int_{\widetilde{t}_A}^{\bar{t}_A}\lambda\exp(-(\rho+\frac{\lambda}{\theta_1})t)dt + (1-\theta)(1-\sigma)\int_{\bar{t}_A}^{\infty}\lambda\exp(-(\rho+\lambda)t)dt.$$

As $\sigma \to \sigma^{*+}$, we have $\bar{t}_1 \to \delta_A$. By implication,

$$\sigma \to \sigma^{*+} \Rightarrow \widetilde{t}_A = \bar{t}_0 - \frac{\mu_1}{\mu_0}\delta_A = \frac{\mu_1}{\mu_0}(\bar{t}_1 - \delta_A) \to 0.$$

By routine simplification, as $\sigma \to \sigma^{*+}$, the principal's payoff in the two-stage equilibrium approaches

$$(1-\frac{\theta}{\theta_1})\int_0^{\delta_A} \lambda \exp(-(\rho+\frac{\lambda}{\theta_1})t)dt + (1-\theta)(1-\sigma)\int_{\delta_A}^{\infty} \lambda \exp(-(\rho+\lambda)t)dt,$$

which is the principal's payoff in the one-stage equilibrium that obtains at σ^*. From the previous part, this payoff strictly exceeds the principal's payoff in the baseline model. By implication, there exists $\epsilon > 0$ such that for an $\sigma' \in (\sigma^*, \sigma^* + \epsilon)$, the principal's payoff in the two-stage equilibrium at σ' strictly exceeds her payoff in the baseline model. Hence, for any $\sigma > \sigma^*$, there exists $\sigma' \in (\sigma^*, \sigma)$ at which the principal's payoff in the two stage equilibrium is higher than her payoff in the baseline model.

Step 2: Consider $\sigma > \sigma^*$. 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 auditing equilibrium and the baseline model,

$$\begin{aligned}
&\exp(-(\mu_0-\mu_1)\widetilde{t}_A)(1-\frac{\theta}{\theta_1})\int_{\widetilde{t}_A}^{\bar{t}_A} \exp(-(\rho+\frac{\lambda}{\theta_1})t)\,\mathrm{d}t \\
&+ (1-\theta)(1-\sigma)\left[\int_{\bar{t}_A}^{\infty} \lambda \exp(-(\rho+\lambda)t)\,\mathrm{d}t - \int_{\bar{t}}^{\infty} \lambda \exp(-(\rho+\lambda)t)\,\mathrm{d}t\right].
\end{aligned}$$

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.

$$\exp(-(\mu_0-\mu_1)\widetilde{t}_A)(1-\frac{\theta}{\theta_1})(\frac{\theta_1}{\theta_1\rho+\lambda})[\exp(-(\rho+\frac{\lambda}{\theta_1})\widetilde{t}_A) - \exp(-(\rho+\frac{\lambda}{\theta_1})\bar{t}_A)].$$

Note that

$$-(\mu_0-\mu_1) = -\lambda\frac{\theta_1(1-\theta_0)-\theta_0(1-\theta_1)}{\theta_1\theta_0} = \lambda\frac{\theta_1-\theta_0}{\theta_1\theta_0} = \frac{\lambda}{\theta_1} - \frac{\lambda}{\theta_0}.$$

Substituting, and using $\bar{t}_A = \widetilde{t}_A + \delta_A$, we have

$$(\frac{\theta_1}{\theta_1\rho+\lambda})\exp(-(\rho+\frac{\lambda}{\theta_0})\widetilde{t}_A)(1-\frac{\theta}{\theta_1})[1-\exp(-(\rho+\frac{\lambda}{\theta_1})\delta_A)].$$

Using $\widetilde{t}_A = \bar{t}_0 - \frac{\mu_1}{\mu_0}\delta_A$, we have

$$(\frac{\theta_1}{\theta_1\rho+\lambda})\exp(-(\rho+\frac{\lambda}{\theta_0})\bar{t}_0)\exp((\rho+\frac{\lambda}{\theta_0})\frac{\mu_1}{\mu_0}\delta_A)(1-\frac{\theta}{\theta_1})[1-\exp(-(\rho+\frac{\lambda}{\theta_1})\delta_A)].$$

Substituting the definition of $\bar{t}_0$, the first line is

$$\begin{aligned}
&(\frac{\theta_1}{\theta_1\rho+\lambda})(1-\sigma)^{(\rho+\frac{\lambda}{\theta_0})\frac{\theta_0}{\lambda(1-\theta_0)}}\exp((\rho+\frac{\lambda}{\theta_0})\frac{\mu_1}{\mu_0}\delta_A)(1-\frac{\theta}{\theta_1})[1-\exp(-(\rho+\frac{\lambda}{\theta_1})\delta_A)] = \\
&(\frac{\theta_1}{\theta_1\rho+\lambda})(1-\sigma)^{\frac{\rho\theta_0+\lambda}{\lambda(1-\theta_0)}}\exp((\rho+\frac{\lambda}{\theta_0})\frac{\mu_1}{\mu_0}\delta_A)(1-\frac{\theta}{\theta_1})[1-\exp(-(\rho+\frac{\lambda}{\theta_1})\delta_A)] = \\
&\kappa_1(1-\sigma)^{\frac{\rho\theta_0+\lambda}{\lambda(1-\theta_0)}},
\end{aligned}$$

where $\kappa_1 \equiv (\frac{\theta_1}{\theta_1\rho+\lambda})\exp((\rho+\frac{\lambda}{\theta_0})\frac{\mu_1}{\mu_0}\delta_A)(1-\frac{\theta}{\theta_1})[1-\exp(-(\rho+\frac{\lambda}{\theta_1})\delta_A)]$ is independent of σ. Next, we simplify the second line.

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

Substituting $\bar{t}_A = \bar{t}_0 + (1-\frac{\mu_1}{\mu_0})\delta_A$, we have

$$(1-\theta)(1-\sigma)\frac{\lambda}{\lambda+\rho}[\exp(-(\rho+\lambda)(1-\frac{\mu_1}{\mu_0})\delta_A)\exp(-(\rho+\lambda)\bar{t}_0)-\exp(-(\rho+\lambda)\bar{t})].$$

Note that

$$\begin{aligned}
(1-\sigma)\exp(-(\rho+\lambda)\bar{t}_0) &= \exp((1-(\rho+\lambda)\frac{\theta_0}{\lambda(1-\theta_0)})\ln(1-\sigma)) = \\
&\exp(-(\rho+\frac{\lambda}{\theta_0})(-\frac{\theta_0}{\lambda(1-\theta_0)})\ln(1-\sigma)) = \exp(-(\rho+\frac{\lambda}{\theta_0})\bar{t}_0),
\end{aligned}$$

and similarly, $(1-\sigma)\exp(-(\rho+\lambda)\bar{t}) = \exp(-(\rho+\frac{\lambda}{\theta})\bar{t})$. Continuing the simplification,

$$\begin{aligned}
&(1-\theta)\frac{\lambda}{\lambda+\rho}[\exp(-(\rho+\lambda)(1-\frac{\mu_1}{\mu_0})\delta_A)\exp(-(\rho+\frac{\lambda}{\theta_0})\bar{t}_0)-\exp(-(\rho+\frac{\lambda}{\theta})\bar{t})] = \\
&(1-\theta)\frac{\lambda}{\lambda+\rho}[\exp(-(\rho+\lambda)(1-\frac{\mu_1}{\mu_0})\delta_A)(1-\sigma)^{\frac{\rho\theta_0+\lambda}{\lambda(1-\theta_0)}}-(1-\sigma)^{\frac{\rho\theta+\lambda}{\lambda(1-\theta)}}] = \\
&\kappa_2(1-\sigma)^{\frac{\rho\theta_0+\lambda}{\lambda(1-\theta_0)}}-\kappa_3(1-\sigma)^{\frac{\rho\theta+\lambda}{\lambda(1-\theta)}},
\end{aligned}$$

where $\kappa_2 \equiv (1-\theta)\frac{\lambda}{\lambda+\rho}\exp(-(\rho+\lambda)(1-\frac{\mu_1}{\mu_0})\delta_A)$ and $\kappa_3 \equiv (1-\theta)\frac{\lambda}{\lambda+\rho}$ are independent of σ. Combining terms, the payoff difference as a function of σ is simply

$$(\kappa_1+\kappa_2)(1-\sigma)^{\frac{\rho\theta_0+\lambda}{\lambda(1-\theta_0)}}-\kappa_3(1-\sigma)^{\frac{\rho\theta+\lambda}{\lambda(1-\theta)}}.$$

Therefore,

$$(\kappa_1+\kappa_2)(1-\sigma)^{\frac{\rho\theta_0+\lambda}{\lambda(1-\theta_0)}} - \kappa_3(1-\sigma)^{\frac{\rho\theta+\lambda}{\lambda(1-\theta)}} > 0 \iff$$
$$(\kappa_1+\kappa_2) - \kappa_3(1-\sigma)^{\frac{\rho\theta+\lambda}{\lambda(1-\theta)}-\frac{\rho\theta_0+\lambda}{\lambda(1-\theta_0)}} > 0 \iff$$
$$\frac{\kappa_1+\kappa_2}{\kappa_3} > (1-\sigma)^{\frac{(\lambda+\rho)(\theta-\theta_0)}{\lambda(1-\theta)(1-\theta_0)}}.$$

Because $\theta_0 < \theta$ for $k \in (0, k^*)$, the right hand side is a decreasing function of σ, while the left hand side does not depend on σ. Thus, if the payoff difference is positive for some value of $\sigma > \sigma^*$, then it is also positive for $\sigma'' \in (\sigma, 1]$.

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

Ethical Agent, one-stage equilibrium. The ethical agent never submits a fake project and all real projects are approved in the one-stage equilibrium. Therefore he receives his first-best payoff

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

Because the baseline model has a positive probability of project rejection, the payoff in the baseline model is strictly smaller.

Ethical Agent, two-stage equilibrium. We claim that the ethical agent's payoff is higher in the two stage auditing equilibrium than in the baseline model,

$$\int_0^{\widetilde{t}_A} \lambda \exp(-(\rho+\lambda)t)p(t)\,\mathrm{d}t + \int_{\widetilde{t}_A}^\infty \lambda \exp(-(\rho+\lambda)t)dt >$$
$$\int_0^{\bar{t}} \lambda \exp(-(\rho+\lambda)t)a(t)\,\mathrm{d}t + \int_{\bar{t}}^\infty \lambda \exp(-(\rho+\lambda)t)dt \iff$$
$$\int_0^{\widetilde{t}_A} \lambda \exp(-(\rho+\lambda)t)[p(t)-a(t)]\,\mathrm{d}t + \int_{\widetilde{t}_A}^{\bar{t}} \lambda \exp(-(\rho+\lambda)t)dt > 0.$$

We establish this by showing (i) $\widetilde{t}_A < \bar{t}$ and (ii) $p(t) > a(t)$ for all $t \leq \widetilde{t}_A$. First note that

$$\widetilde{t}_A < \bar{t} \iff \bar{t}_0 - \frac{\mu_1}{\mu_0}\delta_A < \bar{t} \iff -\frac{\mu_1}{\mu_0}\delta_A < -\ln(1-\sigma)\left(\frac{1}{\mu} - \frac{1}{\mu_0}\right).$$

The left side of the last line is evidently negative and the right side is positive because $k < k^* \Rightarrow \mu < \mu_0$, which establishes (i). To establish (ii), consider the following string of implications for $t \leq \widetilde{t}_A$:

$$
\begin{aligned}
1 > \exp(-\rho(\bar{t}-t)) \quad \text{and} \quad \alpha > 0 \Rightarrow \\
\frac{\phi}{\widehat{\phi}-\alpha} - \frac{\phi}{\widehat{\phi}} > \left(\frac{\phi}{\widehat{\phi}-\alpha} - \frac{\phi}{\widehat{\phi}}\right)\exp(-\rho(\bar{t}-t) \Rightarrow \\
\frac{\phi}{\widehat{\phi}-\alpha} - \frac{\phi}{\widehat{\phi}} + \exp(-\rho(\bar{t}-t)) > \left(\frac{\phi}{\widehat{\phi}-\alpha} - \frac{\phi}{\widehat{\phi}}\right)\exp(-\rho(\bar{t}-t) + \exp(-\rho(\bar{t}-t)) \Rightarrow \\
\frac{\phi}{\widehat{\phi}-\alpha} + \left(1-\frac{\phi}{\widehat{\phi}-\alpha}\right)\exp(-\rho(\bar{t}-t)) > \frac{\phi}{\widehat{\phi}} + \left(1-\frac{\phi}{\widehat{\phi}}\right)\exp(-\rho(\bar{t}-t)).
\end{aligned}
$$

Now, because $\widetilde{t}_A < \bar{t}$ and $\frac{\widehat{\phi}-\alpha}{\widehat{\phi}(1-\alpha)} < 1$, we have

$$
\exp\left(-\frac{(\widehat{\phi}-\alpha)}{\widehat{\phi}(1-\alpha)}\rho(\widetilde{t}_A - t)\right) > \exp(-\rho(\bar{t}-t)),
$$

and hence,

$$
\exp\left(-\frac{\widehat{\phi}-\alpha}{1-\alpha}(\rho+\lambda)(\widetilde{t}_A - t)\right) > \exp(-\rho(\bar{t}-t)).
$$

Combining this with the last line above gives

$$
\frac{\phi}{\widehat{\phi}-\alpha} + \left(1-\frac{\phi}{\widehat{\phi}-\alpha}\right)\exp\left(-\frac{\widehat{\phi}-\alpha}{1-\alpha}(\rho+\lambda)(\widetilde{t}_A - t)\right) > \frac{\phi}{\widehat{\phi}} + \left(1-\frac{\phi}{\widehat{\phi}}\right)\exp(-\rho(\bar{t}-t)),
$$

or $p(t) > a(t)$ as desired.

Strategic Agent, one-stage equilibrium. The one-stage equilibrium exists when $\alpha > \widehat{\phi} - \phi$, for all $\sigma \in [0,1]$. In the one-stage equilibrium, the strategic agent's payoff is $1 - \alpha p(0) - \phi$. As $\sigma \to 1$ both $\bar{t}_A$ and $\bar{t}$ diverge so that

$$
\lim_{\sigma \to 1} 1 - \alpha p(0) = 1 - (\widehat{\phi} - \phi),
$$

and

$$
\lim_{\sigma \to 1} a(0) = \frac{\phi}{\widehat{\phi}}.
$$

Note that

$$
1 - (\widehat{\phi} - \phi) > \frac{\phi}{\widehat{\phi}} \iff (\widehat{\phi} - \phi)(1 - \widehat{\phi}) > 0 \iff \phi < \widehat{\phi} < 1.
$$

Strategic Agent, two-stage equilibrium. Note that for $\alpha < \widehat{\phi} - \phi$ and $\sigma > \widetilde{\sigma}$, the auditing equilibrium is in two phases. Since we assume that σ is large, the strategic agent's payoff is $(1-\alpha)p(0)-\phi$. As $\sigma \to 1$ both $\widetilde{t}_A$ and $\bar{t}_A$ diverge so that

$$\lim_{\sigma\to 1}(1-\alpha)p(0) = \frac{(1-\alpha)\phi}{\widehat{\phi}-\alpha},$$

and

$$\lim_{\sigma\to 1} a(0) = \frac{\phi}{\widehat{\phi}}.$$

Thus, the two stage auditing equilibrium payoff is higher whenever

$$\frac{(1-\alpha)\phi}{\widehat{\phi}-\alpha} > \frac{\phi}{\widehat{\phi}} \iff (1-\alpha)\phi\widehat{\phi} - \phi\widehat{\phi} + \phi\alpha > 0 \iff \phi\alpha(1-\widehat{\phi}) > 0,$$

and, thus, $\widehat{\phi} < 1$ implies that the first limit is larger than the second. □

B Online Appendix: Analysis of Commitment

Here we prove Proposition 5.1 by formally analyzing the setting in which the principal can commit to her acceptance strategy $a(\cdot)$. As noted in the text, we adopt the standard approach of the principal-agent literature sans transfers, assuming that the principal selects her strategy $a(\cdot)$, and recommends a strategy to the agent. The agent complies with the principal's recommendation, provided that doing so is incentive compatible. In what follows, we refer to $a(\cdot)$ and a recommendation as the "design." For the problem to be interesting, we must have $\phi < \widehat{\phi}$, which we assume throughout.

Suppose first that at the solution of the principal's design problem, multiple times are incentive compatible for the agent to select. That is, at the optimum more than one time t satisfies $u(t) \geq u(t')$ for all possible times t'. In this case, the principal recommends her most preferred cheating time among those that are incentive compatible, recommending a mixture only if she is indifferent between them. Thus, it is without loss of generality to focus on an equilibrium in which the principal recommends a single cheating time to the agent, denoted τ.

So the principal's objective function is

$$(1-\theta)\int_0^\tau \lambda\exp\{-(\lambda+\rho)t\}a(t)dt + (1-\theta)(1-\sigma)\int_\tau^\infty \lambda\exp\{-(\lambda+\rho)t\}a(t)dt \\ -\theta\sigma\exp\{-(\lambda+\rho)\tau\}a(\tau).$$

The first term captures the expected benefit of accepting a real arrival generated from either type of agent before the recommended cheating time, the second represents the benefit of accepting a real arrival generated after the recommended cheating time, which only is realized if the agent is ethical, and the third term is the cost of accepting a fraudulent submission at the recommended cheating time. The principal faces an incentive constraint which requires that the agent's payoff of using the recommended cheating time be at least as large as his payoff of using some other cheating time, $u(\tau) \geq u(t)$ for all t, where

$$u(t) = \int_0^t \lambda \exp(-(\rho+\lambda)s)a(s)dt + \exp(-(\rho+\lambda)t)(a(t) - \phi).$$

The principal also faces a feasibility constraint, $a(t) \leq 1$.

Approach. We separately consider the case $\tau = \infty$ and $\tau < \infty$. In the case of $\tau = \infty$, we show that the optimal acceptance strategy that is consistent with such a recommendation is stationary $a(t) = \phi/\widehat{\phi}$.

We then consider the case $\tau < \infty$. We first show that we only need to focus on acceptance strategies for which $a(\tau) > \phi/\widehat{\phi}$. Next, we solve for the optimal acceptance strategy, restricting $(\tau, a(\tau))$ to specific values. Finally, we jointly determine $(\tau, a(\tau))$.

Some of our arguments (in both parts) use the result of a simpler optimization problem, which we refer to as the *auxiliary* problem. In this problem, we fix a time domain $[0, T]$, a "target time" $\tau_A \in [0, T]$, and an acceptance probability at the target time, $a(\tau_A) = \alpha$. The principal's goal is to choose a function $a(\cdot)$ to optimize her payoff on time domain $[0, T]$, subject to IC and $a(\tau_A) = \alpha$. Relative to the original problem, the time domain is restricted to interval $[0, T]$ and we do not require $a(t) \leq 1$. We show that the optimal acceptance strategy is such that IC binds on $[0, T]$ and provide a characterization of $a(\cdot)$. We reference the optimal acceptance strategy in the auxiliary problem several times in the subsequent analysis.

B.1 Auxiliary Problem

Consider an auxiliary optimization problem that is related to the principal's original problem. There is an exogenous time horizon $T < \infty$, an exogenous faking time $\tau_A \in [0, T]$, and an exogenous value α. In the auxiliary problem, the principal's goal is to choose a function $a(\cdot)$ to maximize her objective function in the original problem confined to $[0, T]$, subject to IC and $a(\tau_A) = \alpha$. Relative to the original problem, feasibility is relaxed, we consider only times $[0, T]$, and both τ_A and $a(\tau_A)$ are given. Thus, the auxiliary problem is

$$\max_{a(\cdot)}(1-\theta)\int_0^{\tau_A}\lambda\exp\{-(\lambda+\rho)t\}a(t)dt+(1-\theta)(1-\sigma)\int_{\tau_A}^T\lambda\exp\{-(\lambda+\rho)t\}a(t)dt$$
$$-\theta\sigma\exp\{-(\lambda+\rho)\tau_A\}a(\tau_A),$$

subject to (IC), $a(\tau_A)=\alpha$. Note that, since $(\tau_A, a(\tau_A)=\alpha)$ are exogenous, we can ignore the third term in the objective.

Integral Representation. Given that $a(\cdot)$ is absolutely continuous, so is $u(\cdot)$. It follows that

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

We have

$$u'(t)\exp\{(\rho+\lambda)t\}-\phi(\rho+\lambda)=a'(t)-\rho a(t)$$
$$u'(t)\exp\{\lambda t\}-\phi(\rho+\lambda)\exp(-\rho t)=(a'(t)-\rho a(t))\exp(-\rho t)$$
$$u'(t)\exp\{\lambda t\}-\phi(\rho+\lambda)\exp(-\rho t)=\frac{d}{dt}[\exp\{-\rho t\}a(t)].$$

integrating from τ_A to t and using absolute continuity of $a(\cdot)$, we have

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

Lemma B.1. *In the solution of the auxiliary problem, incentive compatibility binds at all times, i.e., $u(t)=u(\tau_A)$ for all $t\in[0,T]$.*

Proof of Lemma B.1. Using the integral representation and ignoring the coefficient $(1-\theta)$, the first term of the objective function $(t<\tau_A)$ can be written,

$$\int_0^{\tau_A}\lambda\exp\{-(\lambda+\rho)t\}[\frac{\phi}{\hat{\phi}}+(a(\tau_A)-\frac{\phi}{\hat{\phi}})\exp\{-\rho(\tau_A-t)\}]dt+$$
$$\int_0^{\tau_A}\lambda u(t)-\lambda\exp\{-\lambda(t-\tau_A)\}u(\tau_A)\}dt+\int_0^{\tau_A}\int_t^{\tau_A}\lambda^2\exp\{-\lambda t\}\exp\{\lambda s\}u(s)dsdt$$

The first line is exogenous, since $(\tau_A, a(\tau_A)=\alpha)$ are given. Consider the double integral,

$$\int_0^{\tau_A}\int_t^{\tau_A}\lambda^2\exp\{-\lambda t\}\exp\{\lambda s\}u(s)dsdt=\int_0^{\tau_A}\int_0^{s}\lambda^2\exp\{-\lambda t\}\exp\{\lambda s\}u(s)dtds=$$
$$\int_0^{\tau_A}\lambda[1-\exp\{-\lambda s\}]\exp\{\lambda s\}u(s)ds=\int_0^{\tau_A}\lambda[\exp\{\lambda s\}-1]u(s)ds$$

Thus, the first term of the auxiliary objective function is

$$(1-\theta)\int_0^{\tau_A} \lambda \exp\{-(\lambda+\rho)t\}[\frac{\phi}{\hat{\phi}} + (a(\tau_A) - \frac{\phi}{\hat{\phi}})\exp\{-\rho(\tau_A - t)\} \\ + \Big(\lambda \exp(\lambda t)u(t) - \lambda \exp\{-\lambda(t-\tau_A)\}u(\tau_A)\Big)dt.$$

Ignoring the coefficient $(1-\theta)(1-\sigma)$, the second term of the auxiliary objective $(t > \tau_A)$ is

$$\int_{\tau_A}^{T} \lambda \exp\{-(\lambda+\rho)t\}[\frac{\phi}{\hat{\phi}} + (a(\tau_A) - \frac{\phi}{\hat{\phi}})\exp\{-\rho(\tau_A - t)\}]dt + \\ \int_{\tau_A}^{T} \lambda u(t) - \lambda \exp\{-\lambda(t-\tau_A)\}u(\tau_A)dt - \int_{\tau_A}^{T}\int_{\tau_A}^{t} \lambda^2 \exp\{-\lambda t\}\exp\{\lambda s\}u(s)dsdt$$

Focusing on the double integral. We have

$$\int_{\tau_A}^{T}\int_{\tau_A}^{t} \lambda^2 \exp\{-\lambda t\}\exp\{\lambda s\}u(s)dsdt = \int_{\tau_A}^{T}\int_{s}^{T} \lambda^2 \exp\{-\lambda t\}\exp\{\lambda s\}u(s)dtdt = \\ \int_{\tau_A}^{T} \lambda[\exp\{-\lambda s\} - \exp\{-\lambda T\}]\exp\{\lambda s\}u(s)ds = \int_{\tau_A}^{T} \lambda[1 - \exp\{-\lambda(T-s)\}]u(s)ds.$$

Therefore, up to the coefficient, the second term of the auxiliary objective function is

$$\int_{\tau_A}^{T} \lambda \exp\{-(\lambda+\rho)t\}[\frac{\phi}{\hat{\phi}} + (a(\tau_A) - \frac{\phi}{\hat{\phi}})\exp(-\rho(\tau_A - t))] + \\ \int_{\tau_A}^{T} \lambda u(t) - \lambda \exp\{-\lambda(t-\tau_A)\}u(\tau_A) - \lambda[1-\exp\{-\lambda(T-t)\}]u(t)dt = \\ \int_{\tau_A}^{T} \lambda \exp\{-(\lambda+\rho)t\}[\frac{\phi}{\hat{\phi}} + (a(\tau_A) - \frac{\phi}{\hat{\phi}})\exp(-\rho(\tau_A - t))] + \\ \Big(\lambda \exp\{-\lambda(T-t)\}u(t) - \lambda \exp\{-\lambda(t-\tau_A)\}u(\tau_A)\Big)dt$$

In both parts of the auxiliary objective ($t < \tau_A$ and $t > \tau_A$), the integrand is strictly increasing in $u(t)$, and hence, in the solution of the auxiliary problem $u(t) = u(\tau_A)$. □

Proposition B.1. *In the optimal auxiliary design, $u(t) = u(\tau_A)$ for all $t \in [0,T]$, and $a(t) = a_X(t)$ where*

$$a_X(t) \equiv \frac{\phi}{\hat{\phi}} + (\alpha - \frac{\phi}{\hat{\phi}})\exp\{-\rho(\tau_A - t)\}$$

Proof of Proposition B.1. Follows immediately from $u(t) = u(\tau_A)$ on $[0,T]$ with boundary condition at $t = \tau_A$. □

B.2 Optimal Design

In this section we use the results on the auxiliary problem to solve the principal's actual problem. We proceed in three steps. 1) We analyze the best possible acceptance strategy that incentivizes the agent to never cheat, i.e., $\tau = \infty$. 2) We consider the best possible design that induces cheating at some time $\tau < \infty$ and has a specific value $a(\tau) = \alpha$. We then show that the limit in this problem as $\tau \to \infty$ converges to the best payoff for $\tau = \infty$, and thus, the case of $\tau = \infty$ can be analyzed as a limiting case of τ finite. 3) We jointly optimize (τ, α).

Never faking. We first characterize the acceptance strategy that delivers the highest possible payoff, assuming $\tau = \infty$ is incentive compatible. In this case, incentive compatibility requires that for any t, there exists $t' > t$ such that $u(t') \geq u(t)$.

Lemma B.2. *Consider all acceptance strategies for which $\tau = \infty$ is incentive compatible. Of these, the optimal one for the principal is $a(t) = \phi/\widehat{\phi}$ for all $t \geq 0$.*

Proof of Lemma B.2. Step 1. We argue that the proposed design is incentive compatible. With $a(t) = \frac{\phi}{\widehat{\phi}}$, we have that $u(t) = \phi\lambda/\rho$. Thus, the agent's payoff of faking is constant and $\tau = \infty$ is incentive compatible.

Next, We will show that, among all acceptance strategies for which $\tau = \infty$ is IC, acceptance strategy $a(t) = \phi/\widehat{\phi}$ is best. Suppose that one of the best acceptance strategies with this property is denoted $\widetilde{a}(\cdot)$.

Step 2. We construct another acceptance strategy $\widehat{a}(\cdot)$, based on acceptance strategy $\widetilde{a}(\cdot)$. In particular, first select any time T. Consider the solution of the following auxiliary problem,

$$\max_{a(\cdot)} \int_0^T \lambda \exp(-(\rho + \lambda)t) a(t) dt$$

subject to (IC) and $a(T) = \widetilde{a}(T)$. That is, T arbitrary, $\tau_A = T$ and $\alpha = \widetilde{a}(T)$. From our analysis of the auxiliary problem in Proposition B.1, we know that the solution of this auxiliary problem is

$$a_X(t) = \frac{\phi}{\widehat{\phi}} + (\widetilde{a}(T) - \frac{\phi}{\widehat{\phi}}) \exp(-\rho(T - t)).$$

Using this function, we construct a new acceptance strategy, $\widehat{a}$, whose performance we will compare against the original strategy $\widetilde{a}$. In particular, consider

$$\widehat{a}(t) = \begin{cases} a_X(t) & t < T \\ \widetilde{a}(t) & t \geq T. \end{cases}$$

That is, to the left of T, the constructed acceptance strategy is the optimal acceptance strategy from the auxiliary problem, chosen to match target value $\widetilde{a}(T)$. To the right of T we use the original acceptance strategy. In addition, $\widehat{a}(\cdot)$ is continuous at T by construction.

Step 3. We show that (i) $\widehat{a}(\cdot) \in [0,1]$ and (ii) if the principal uses $\widetilde{a}(\cdot)$, then $\tau = \infty$ is incentive compatible. Note first that for $t > T$, we have $\widehat{a}(t) = \widetilde{a}(t) \in [0,1]$. Next, note that for $t \leq T$, we have $\widehat{a}(t) \in [\phi/\widehat{\phi}, \widetilde{a}(T)]$, and therefore $\widehat{a}(t) \in (0,1]$.

Next, we show that when the principal commits to $\widehat{a}(\cdot)$, $\tau = \infty$ is incentive compatible. In particular, we show that for any t, there exists $t' > t$ such that $\widehat{u}(t') \geq \widehat{u}(t)$, where $\widehat{u}(t)$ is the agent's payoff of selecting faking time t with acceptance strategy $\widehat{a}(\cdot)$.

First consider $t' > t \geq T$. Note that for all $t' > t \geq T$, the alternative design is identical to the original, and hence,

$$\begin{aligned}
&\widehat{u}(t') - \widehat{u}(t) = \\
&\int_t^{t'} \lambda \exp(-(\rho+\lambda)s)\widetilde{a}(s)ds + \exp(-(\rho+\lambda)t')(\widetilde{a}(t') - \phi) - \exp(-(\rho+\lambda)t)(\widetilde{a}(t) - \phi) = \\
&u(t') - u(t).
\end{aligned}$$

It follows that if $u(t') > u(t)$ in the original design, then it is also larger in the modified design. Because $\tau = \infty$ is IC in the original design, for any t, some $t' > t$ can be found such that $u(t') > u(t)$. Hence, $\widehat{u}(t') > \widehat{u}(t)$ in the modified design as well.

Next, consider $t < T$. It is elementary to show by direct verification that for all $t \in [0,T]$, we have $\widehat{u}(t) = \widehat{u}(T)$ (equivalently, recall that in the auxiliary problem, IC binds in the entire interval). Furthermore, from our previous argument, we have that there exists a $t' > T$ such that $\widehat{u}(t') > \widehat{u}(T)$, and hence $\widehat{u}(t') > \widehat{u}(T) = \widehat{u}(t)$.

Step 4. We show that the principal's payoff of the alternative acceptance strategy $\widehat{a}(\cdot)$ must be the same as the payoff of the optimal strategy $\widetilde{a}(\cdot)$, i.e.

$$\int_0^\infty \lambda \exp(-(\rho+\lambda)t)\widehat{a}(t)dt = \int_0^\infty \lambda \exp(-(\rho+\lambda)t)\widetilde{a}(t)dt.$$

Because $\widetilde{a}(\cdot)$ is optimal, it is enough to show that

$$\int_0^\infty \lambda \exp(-(\rho+\lambda)t)\widehat{a}(t)dt \geq \int_0^\infty \lambda \exp(-(\rho+\lambda)t)\widetilde{a}(t)dt.$$

First, recall that $a_X(\cdot)$ solves the auxiliary problem, and hence

$$\begin{aligned}
&\int_0^T \lambda \exp(-(\rho+\lambda)t)a_X(t)dt \geq \int_0^T \lambda \exp(-(\rho+\lambda)t)\widetilde{a}(t)dt \Rightarrow \\
&\int_0^T \lambda \exp(-(\rho+\lambda)t)a_X(t)dt + \int_T^\infty \lambda \exp(-(\rho+\lambda)t)\widetilde{a}(t)dt \geq \int_0^\infty \lambda \exp(-(\rho+\lambda)t)\widetilde{a}(t)dt.
\end{aligned}$$

Note that by construction, the left hand side is the principal's payoff of using $\widehat{a}(\cdot)$, so that

$$\int_0^\infty \lambda \exp(-(\rho+\lambda)t)\widehat{a}(t)dt \geq \int_0^\infty \lambda \exp(-(\rho+\lambda)t)\widetilde{a}(t)dt,$$

as we sought to show.

Step 5. We show that $\widetilde{a}(t) = \frac{\phi}{\widehat{\phi}}$. From Step 4, that alternative acceptance strategy $\widehat{a}(\cdot)$ achieves the same value as $\widetilde{a}(\cdot)$, and this is true for all $T > 0$ at which the alternative design is constructed. Hence,

$$\int_0^\infty \lambda \exp(-(\rho+\lambda)t)\widehat{a}(t)dt = \int_0^T \lambda \exp(-(\rho+\lambda)t)a_X(t)dt + \int_T^\infty \lambda \exp(-(\rho+\lambda)t)\widetilde{a}(t)dt$$

does not depend on T. By implication, its derivative with respect to T is 0, i.e.

$$\int_0^T \lambda \exp(-(\rho+\lambda)t)\frac{da_X(t)}{dT}dt = 0.$$

It therefore follows that for all t,

$$\frac{da_X(t)}{dT} = 0 \Rightarrow -\rho(\widetilde{a}(T) - \frac{\phi}{\widehat{\phi}})\exp(-\rho(T-t)) + \widetilde{a}'(T)\exp(-\rho(T-t)) = 0 \Rightarrow \widetilde{a}(T) - \frac{\phi}{\widehat{\phi}} = \kappa \exp(\rho T),$$

for some constant κ. Note that if $\kappa \neq 0$, then for sufficiently large T, $\widetilde{a}(T)$ will either exceed 1, or become negative. Hence, it must be that $\kappa = 0$. By implication $\widetilde{a}(T) = \phi/\widehat{\phi}$. Since T is chosen arbitrarily, we have the result. □

Finite Cheating. Now we consider acceptance strategies that recommend incentive compatible cheating at some finite time $\tau < \infty$. We will eventually consider acceptance strategies such that cheating at τ is IC and have a fixed value $a(\tau)$. Before we do that, however, we establish a bound on $a(\tau)$ that applies to the optimal design, i.e., the combination of faking time and acceptance strategy that are optimal in the principal's original problem (Lemmas B.3, B.4, B.5). In Lemma B.5 we consider acceptance strategies such that the recommended IC faking time is τ, and $a(\tau) = \alpha$, where α is consistent with the bound we established. We then optimize over (τ, α).

Lemma B.3. *If $\tau < \infty$ in the optimal design, then $u(\tau) > \frac{\phi\lambda}{\rho}$.*

Proof. Step 1. We show that in the optimal design, the principal's payoff must be weakly larger than $(1-\theta)\phi\lambda/\rho$. Consider a particular acceptance strategy for the principal, $a(t) = \phi/\widehat{\phi} = \phi(1+\frac{\lambda}{\rho})$. Obviously, $a(t) \in (0,1)$. Next, we show that for all $t \geq 0$, in this design $u(t) = \frac{\phi\lambda}{\rho}$, and thus any $\tau \geq 0$ is an incentive compatible recommendation. We have

$$u(t) = \frac{\lambda}{\lambda+\rho}(1-\exp\{-(\lambda+\rho)t\})\phi(1+\frac{\lambda}{\rho}) + \exp\{-(\lambda+\rho)t\}\frac{\phi\lambda}{\rho} = \frac{\phi\lambda}{\rho}.$$

Finally, we show that by recommending $\tau = \infty$ and $a(\tau) = \phi(1 + \frac{\lambda}{\rho})$, the principal can achieve a payoff $(1-\theta)\phi\lambda/\rho$. To see this, note that the principal's payoff for the proposed acceptance strategy as $\tau \to \infty$ is

$$(1-\theta)\int_0^\infty \lambda \exp\{-(\lambda+\rho)t\}\phi(1+\frac{\lambda}{\rho})dt = (1-\theta)\frac{\phi\lambda}{\rho}.$$

Thus, an incentive compatible design allows the principal to achieve payoff $(1-\theta)\phi\lambda/\rho$, and therefore her payoff in the optimal design must be weakly larger.

Step 2. We show that if the optimal design for the principal has $\tau < \infty$, then it cannot be the case that both $a(\tau) = 0$ and

$$\int_\tau^\infty \exp\{-(\lambda+\rho)t\}a(t)dt = 0.$$

That is, if $\tau < \infty$, then at least one of these is strictly positive. Suppose both are zero. For $t > \tau$ we have

$$\begin{aligned}
&u(t) - u(\tau) = \\
&\int_\tau^t \lambda \exp(-(\rho+\lambda)s)a(s)dt + \exp(-(\rho+\lambda)t)(a(t)-\phi) - \exp(-(\rho+\lambda)\tau)(a(\tau)-\phi) = \\
&\exp(-(\rho+\lambda)t)(-\phi) - \exp(-(\rho+\lambda)\tau)(-\phi) = \phi(\exp(-(\rho+\lambda)\tau) - \exp(-(\rho+\lambda)t)) > 0,
\end{aligned}$$

contradicting incentive compatibility.

Step 3. We show that if the optimal design has $\tau < \infty$, then $u(\tau) > \phi\lambda/\rho$. From Step 1, we know that in the optimal design, the principal's payoff must exceed $(1-\theta)\phi\lambda/\rho$,

$$\begin{aligned}
(1-\theta)\int_0^\tau \lambda \exp\{-(\lambda+\rho)t\}a(t)dt + (1-\theta)(1-\sigma)\int_\tau^\infty \lambda \exp\{-(\lambda+\rho)t\}a(t)dt \\
-\theta\sigma \exp\{-(\lambda+\rho)\tau\}a(\tau) \geq (1-\theta)\frac{\phi\lambda}{\rho}.
\end{aligned}$$

Next, from Step 2 we have the following inequality,

$$\theta\sigma \exp\{-(\rho+\lambda)\tau\}a(\tau) + (1-\theta)\sigma \int_\tau^\infty lambdaexp\{-(\rho+\lambda)t\}a(t)dt > 0.$$

Adding this quantity to the left hand side yields,

$$\begin{aligned}
(1-\theta)\int_0^\tau \lambda \exp\{-(\lambda+\rho)t\}a(t)dt + [(1-\theta)(1-\sigma)+\theta\sigma]\int_\tau^\infty \lambda \exp\{-(\lambda+\rho)t\}a(t)dt > (1-\theta)\frac{\phi\lambda}{\rho} \Rightarrow \\
(1-\theta)\int_0^\infty \lambda \exp\{-(\lambda+\rho)t\}a(t)dt > (1-\theta)\frac{\phi\lambda}{\rho} \\
\int_0^\infty \lambda \exp\{-(\lambda+\rho)t\}a(t)dt > \frac{\phi\lambda}{\rho}.
\end{aligned}$$

Next, note that $u(t) \leq u(\tau)$, and hence,

$$\lim_{t\to\infty} u(t) \leq u(\tau) \Rightarrow \int_0^\infty \lambda \exp\{-(\lambda+\rho)t\}a(t)dt \leq u(\tau).$$

Thus, we have shown

$$\frac{\phi\lambda}{\rho} < \int_0^\infty \lambda \exp\{-(\lambda+\rho)t\}a(t)dt \leq u(\tau).$$

□

Lemma B.4. *If a design satisfies the constraints of the principal's problem and $\tau < \infty$, then*

$$\int_0^\tau \lambda \exp\{-(\rho+\lambda)t\}a(t)dt \leq \frac{\phi\lambda}{\rho} + (a(\tau) - \frac{\phi}{\widehat{\phi}}) \exp\{-\rho(\tau-t)\} - \exp\{-(\rho+\lambda)\tau\}(a(\tau)-\phi).$$

Proof of Lemma B.4. The principal's problem requires incentive compatibility on interval $[0,\tau]$. Therefore, the optimal strategy of the principal's problem restricted to interval $[0,\tau]$ is within the feasible set of the auxiliary problem with $T=\tau$. It follows that

$$\int_0^\tau \lambda \exp\{-(\rho+\lambda)t\}a(t)dt \leq \int_0^\tau \lambda \exp\{-(\rho+\lambda)t\}[\frac{\phi}{\widehat{\phi}} + (a(\tau)-\phi)\exp(-\rho(\tau-t))]dt,$$

where the bracketed term is the solution of the auxiliary problem, derived in Lemma B.1. The rest follows by routine calculation. □

Lemma B.5. *If $\tau < \infty$ in the optimal design, then $a(\tau) > \frac{\phi}{\widehat{\phi}}$.*

Proof of Lemma B.5. Consider any design that satisfies the constraints of the principal's problem. We have,

$$u(\tau) = \int_0^\tau \lambda \exp\{-(\rho+\lambda)t\}a(t)dt + \exp\{-(\rho+\lambda)\tau\}(a(\tau)-\phi).$$

Applying the bound in Lemma B.4, we have

$$u(\tau) \leq \frac{\phi\lambda}{\rho} + (a(\tau) - \frac{\phi}{\widehat{\phi}})\exp\{-\rho(\tau-t)\}.$$

By Lemma B.3, we must have $a(\tau) > \phi/\widehat{\phi}$. □

We wish to derive the optimal design by first solving for the best acceptance strategy for exogenous finite cheating time τ and exogenous acceptance at the cheating time, $a(\tau)$, and then optimizing over these two. From Lemma B.5, we need only consider $a(\tau) > \phi/\widehat{\phi}$, since this is necessary if the optimal design has $\tau < \infty$.

Lemma B.6. *Consider all designs $(\tau, \tilde{a}(\cdot))$ that satisfy (i) incentive compatibility, (ii) feasibility, (iii) $\tau < \infty$, and (iv) $\tilde{a}(\tau) = \alpha \in (\phi/\widehat{\phi}, 1]$. Among such designs, an optimal acceptance strategy is*

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

for $t \in [0, T]$ and $a(t) = 1$ for $t \geq T$, where

$$T \equiv \tau + \frac{\ln(\frac{1-\frac{\phi}{\widehat{\phi}}}{\alpha - \frac{\phi}{\widehat{\phi}}})}{\rho}$$

In addition, $u(t) = u(\tau)$ for all $t \leq T$ and $u(t) < u(\tau)$ for $t > T$.

Proof Lemma B.6. First, note that the proposed design satisfies all of the conditions stated in the lemma. This can be directly verified by routine calculation (note, in particular, that the proposed acceptance strategy is continuous at T). Next, we will show that this design performs better than any other design that meets the conditions of the lemma.

The acceptance strategy described in the lemma for $t \leq T$ solves the auxiliary problem on time domain $[0, T]$, with $a(\tau) = \alpha$ (see Proposition B.1). By implication, the principal's objective on this interval is weakly higher than for any other acceptance strategy, $\tilde{a}(\cdot)$ that satisfies the assumptions of the lemma. Recalling that $T \geq \tau$ by construction, we have,

$$(1-\theta)\int_0^\tau \lambda \exp(-(\lambda+\rho)t)a(t)dt + (1-\theta)(1-\sigma)\int_\tau^T \lambda \exp(-(\lambda+\rho)t)a(t)dt - \theta\sigma \exp(-(\lambda+\rho)\tau)a(\tau) \geq$$
$$(1-\theta)\int_0^\tau \lambda \exp(-(\lambda+\rho)t)\tilde{a}(t)dt + (1-\theta)(1-\sigma)\int_\tau^T \lambda \exp(-(\lambda+\rho)t)\tilde{a}(t)dt - \theta\sigma \exp(-(\lambda+\rho)\tau)\tilde{a}(\tau).$$

Next, recall that $a(t) = 1$ for $t \geq T$. Feasibility implies that

$$(1-\theta)(1-\sigma)\int_T^\infty \lambda \exp(-(\lambda+\rho)t)a(t)dt \geq \ (1-\theta)(1-\sigma)\int_T^\infty \lambda \exp(-(\lambda+\rho)t)\widetilde{a}(t)dt$$

The optimality of this acceptance strategy follows by adding these inequalities.

The last statement of the lemma can be verified directly by integration or by recalling that IC binds on $[0, T]$ in the auxiliary problem, and that $u'(\cdot) < 0$ on any interval in which $a(\cdot) = 1$. □

Determining (τ, α). In Lemma B.6 we identify the optimal acceptance strategy that is incentive compatible, has finite faking time, and has a target value $a(\tau) = \alpha$. In addition, $\alpha > \phi/\widehat{\phi}$, which is a necessary condition for $\tau < \infty$ to be optimal overall (Lemma B.5).

By routine calculation, we can verify that as $\tau \to \infty$ the design of lemma B.6 and its payoff approach the design and payoff of the one in Lemma B.2, which characterizes the optimal acceptance strategy that satisfies IC and has $\tau = \infty$. Thus, as long as we allow $\tau \to \infty$, we can focus on the class of designs identified in Lemma B.5. Identifying the best (τ, α) in this class will give the optimal overall design.

To facilitate this calculation, it is helpful to re-parameterize the acceptance strategy according to the time, T, at which they hit 1. In particular, note that for any such strategy,

$$1 = \frac{\phi}{\widehat{\phi}} + (a(\tau) - \frac{\phi}{\widehat{\phi}}) \exp(-\rho(\tau - T))$$
$$a(\tau) = (1 - \frac{\phi}{\widehat{\phi}}) \exp(\rho(\tau - T)) + \frac{\phi}{\widehat{\phi}}$$

and hence, the acceptance strategy for $t \leq T$ can be written

$$a(t) = \frac{\phi}{\widehat{\phi}} + (1 - \frac{\phi}{\widehat{\phi}}) \exp(-\rho(T - t)).$$

Lemma B.7. *Consider a design within the class defined by Proposition B.6. For fixed T, the optimal recommended cheating time is $\tau = T$.*

Proof of Lemma B.7. Consider a design in the class defined by Proposition B.6, and fix a value of T. Note that the acceptance strategy does not depend on the recommended cheating time τ. Furthermore, from Proposition B.6 we know that the only candidates for the recommended cheating time are $\tau \in [0, T]$ (recommending a time after T is not incentive compatible). Consider a possible $\tau < T$, and differentiate the principal's payoff with respect to τ,

$$\begin{aligned}
\lambda \exp\{-(\lambda+\rho)\tau\} a(\tau) - (1-\theta)(1-\sigma)\lambda \exp\{-(\lambda+\rho)\tau\} a(\tau) \\
+\theta\sigma(\lambda+\rho) \exp\{-(\lambda+\rho)\tau\} a(\tau) - \theta\sigma \exp\{-(\lambda+\rho)\tau\} a'(\tau) = \\
\exp\{-(\lambda+\rho)\tau\}\Big(\lambda a(\tau) - (1-\theta)(1-\sigma)\lambda a(\tau) + \theta\sigma(\lambda+\rho) a(\tau) - \theta\sigma a'(\tau)\Big) = \\
\exp\{-(\lambda+\rho)\tau\}\Big((\theta+\sigma)\lambda a(\tau) - \theta\sigma\lambda a(\tau) + \theta\sigma(\lambda+\rho) a(\tau) - \theta\sigma a'(\tau)\Big) = \\
\exp\{-(\lambda+\rho)\tau\}\Big((\theta+\sigma)\lambda a(\tau) + \theta\sigma(\rho a(\tau) - a'(\tau))\Big).
\end{aligned}$$

Because $\tau < T$, we have $u'(\tau) = 0$, and hence, $a'(\tau) - \rho a(\tau) = \phi(\rho + \lambda)$. Substituting, we have

$$\exp\{-(\lambda+\rho)\tau\}(\lambda(\theta+\sigma) a(\tau) + \theta\sigma\phi(\lambda+\rho)) > 0.$$

Thus, no cheating time below T could be optimal, because the principal's payoff is increasing on this interval. Furthermore, because $a(\cdot)$ is continuous, the principal's payoff is also continuous at $\tau = T$. Therefore, selecting $\tau = T$ dominates any other feasible cheating time. $\square$

What remains is to characterize T. To proceed, consider the principal's payoff function, evaluated at an acceptance strategy in the optimal class, with $\tau = T$. We have

$$(1-\theta)\int_0^T \lambda \exp(-(\rho+\lambda)t)a(t)dt + (1-\theta)(1-\sigma)\int_T^\infty \lambda \exp(-(\rho+\lambda)t)dt - \theta\sigma \exp(-(\rho+\lambda)T).$$

Differentiating with respect to T (note that $a(T) = 1$), we have

$$(1-\theta)\int_0^T \lambda \exp(-(\rho+\lambda)t)\frac{da(t)}{dT}dt + (1-\theta)\lambda \exp(-(\rho+\lambda)T)$$
$$-(1-\theta)(1-\sigma)\lambda \exp(-(\rho+\lambda)T) + \theta\sigma(\rho+\lambda)\exp(-(\rho+\lambda)T).$$

Note that

$$\frac{da(t)}{dT} = -\rho(1-\frac{\phi}{\widehat{\phi}})\exp(-\rho(T-t)).$$

Substituting, we have

$$-\rho(1-\frac{\phi}{\widehat{\phi}})\exp(-\rho T)(1-\theta)\int_0^T \lambda \exp(-\lambda t)dt + (1-\theta)\lambda \exp(-(\rho+\lambda)T)$$
$$-(1-\theta)(1-\sigma)\lambda \exp(-(\rho+\lambda)T) + \theta\sigma(\rho+\lambda)\exp(-(\rho+\lambda)T)$$

Integrating, we have,

$$(1-\theta)(1-\exp(-\lambda T))(-\rho+\phi(\rho+\lambda))\exp(-\rho T) + (1-\theta)\lambda \exp(-(\rho+\lambda)T)$$
$$-(1-\theta)(1-\sigma)\lambda \exp(-(\rho+\lambda)T) + \theta\sigma(\rho+\lambda)\exp(-(\rho+\lambda)T).$$

Multiplying through by $\exp(\rho T)$ results in an equation that is linear in $\exp(-\lambda T)$. Solving, we have

$$\exp(-\lambda T) = \frac{(1-\theta)(\rho-\phi(\lambda+\rho))}{(1-\theta)(\rho-\phi(\lambda+\rho)) + \sigma(\lambda+\theta\rho)},$$

$$\exp(-\lambda T) = \frac{(1-\theta)(\widehat{\phi}-\phi)}{(1-\theta)(\widehat{\phi}-\phi) + \sigma(\frac{\lambda+\theta\rho}{\lambda+\rho})}$$

$$\exp(-\lambda T) = \frac{(1-\theta)(\widehat{\phi}-\phi)}{(1-\theta)(\widehat{\phi}-\phi) + \sigma(1-\widehat{\phi}(1-\theta))}$$

Note that for $\phi < \widehat{\phi}$ the right hand side is positive, and it is obviously between 0 and 1. Furthermore, as $\phi \to \widehat{\phi}$, the right hand side goes to 0, which implies that $T \to \infty$. As $\phi \to 0$, the optimal T with commitment solves

$$\exp(-\lambda T) = \frac{(1-\theta)\rho}{(1-\theta)\rho + \sigma(\lambda+\theta\rho)}.$$

When ρ is small T is large, and vice versa. Thus, the optimal duration may be longer or shorter than in equilibrium, and by implication, the agent's level of utility may also be higher or lower.

Once we have the structure of the optimal acceptance rule, coupled with the result that the optimal recommended cheating time is $\tau = T$, the principal's tradeoff becomes relatively straightforward. If the principal selects a short duration T, then he gets to accept honest submissions before T with high probability. But if the agent is unethical and fails to have an arrival, he will have to accept his fraudulent project at an earlier time, paying a larger cost. Thus, short durations are desirable when the agent is likely to be the ethical type. Note further that even when the agent is known to be strategic, the principal benefits from a finite duration (unlike what happens in the equilibrium without commitment). Indeed, without commitment, the phase of doubt lasts forever, and the principal gets no surplus. By committing to a finite duration, the principal raises the acceptance probability for all real arrivals before date T, at the cost of accepting a fraudulent project with certainty if no real arrival comes before T.