Abstract

We study a dynamic model of fraud and trust-building featuring two parties: a principal, who has limited power of commitment and who wishes to accept real projects and reject fake ones, and an agent, who is either an ethical type that produces only a real project, or a strategic type that produces a real project but can also choose to generate a fake. Producing a real project takes a positive and uncertain amount of time, while a fake project can be manufactured instantaneously at some cost. In the unique equilibrium, the strategic agent randomizes the timing of fraud and the principal randomizes between acceptance and rejection. As time passes without a submission, the principal's belief that the agent is ethical grows until the point when she fully trusts the agent and accepts any subsequent submission with certainty. We explore several institutional remedies designed to improve the principal’s welfare. These include: an imperfect auditing technology, opaque standards, commitment to suppress fraud, and requiring submissions to occur on a specified date.

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

Keywords: audit, deception, fraud, innovation, opaque standards, research, trust.
1 Introduction

Fraud and deception are perpetrated in diverse and numerous economic settings. A survey conducted by the Federal Trade Commission in 2012 indicated that over 10% of adult US citizens, an estimated 25.6 million people, are victims of consumer fraud annually. A highly publicized incident of fraud in the arena of high finance is Theranos, the Silicon Valley healthcare technology startup, whose founder and former CEO, Elizabeth Holmes, and former president, Ramesh Balwani, duped private investors out of over $700 million by convincing them that the company’s key product—a portable blood analyzer—could conduct comprehensive blood tests from minute amounts of blood. Another notorious case is that of Rudy Kurniawan, (so-called “Dr. Conti”), who in 2013 was convicted to ten years in federal prison for producing counterfeit bottles of high-end wines to be sold at auction. The 2016 documentary Sour Grapes reports that as many as 10,000 bottles created by Kurniawan may still be in private collections. Deceit is also a significant concern in the production and dissemination of a wide spectrum of academic and scientific research findings, as illustrated by the high-profile cases of Diederik Stapel, formerly a professor of social psychology at Tilburg, who reportedly falsified data in at least 55 articles; Anil Potti, a physician and former cancer researcher at Duke, who allegedly altered data in 19 publications and grant applications; Brian Wansink, the founder and former director of the Cornell Food and Brand Lab, who may have manipulated data in at least 27 articles; and Piero Anversa, former director of Harvard’s Center for Regenerative Medicine, who apparently falsified data in at least 31 publications and grant applications.\(^1\) Facing time pressure, a procurer may be tempted to cut corners, as in the case of AEY Inc., which illegally purchased Chinese munitions in order to fulfill a $300 million contract with the United States government.

In this paper, we present and analyze a dynamic principal-agent model of fraud and trust. The principal has limited power of commitment and wishes to approve a real project and reject a fake. The agent is either an ethical type, who only produces a real project, or a strategic type, who produces a real project but can also generate a fake. Real projects and fakes have different arrival processes: real project development takes a positive and uncertain amount of time, while a fake project can be manufactured instantaneously at some cost. Furthermore, in the baseline model, real projects and fakes are indistinguishable

for the principal: she only observes the time at which a project is submitted when deciding whether to accept or reject it. Thus, the time of project arrival plays a critical role in the principal’s approval decision, allowing the principal to learn both about the agent’s ethics and the project’s authenticity.

In the unique equilibrium of the game, both the principal and strategic type of agent play mixed strategies on an initial, finite span of time, which we refer to as the *phase of doubt.* Specifically, during the phase of doubt, the strategic agent searches for a real discovery, but also randomly creates a fake. Thus, when the principal receives a submission, she is uncertain whether it constitutes a real breakthrough by an ethical agent, a real breakthrough by a strategic agent, or a fake. In equilibrium, the principal employs a random approval policy over this phase. To offset the strategic agent’s desire to commit fraud as early as possible, the principal’s equilibrium approval probability increases over time. In other words, she accepts early submissions with low probability and late ones with high probability. Furthermore, as time passes without receiving a submission, the principal’s confidence that the agent is ethical grows. A higher level of trust allows the strategic type to fake at a higher rate without triggering rejection. Thus, the rate of faking increases during the phase of doubt. Indeed, the strategic type submits a fake with probability one by some finite date, at which the phase of doubt ends. From that point forward, the principal is fully confident that the agent is ethical and she approves any submission with probability one.

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

The first two remedies we consider mitigate the agent’s informational advantage. Specifically, in Section 5 we allow the principal to audit the submission before deciding whether to approve it. The audit is costly and noisy: though it reduces the agent’s informational advantage, it does not eliminate it. The dynamic structure of the auditing equilibrium is especially intriguing when the auditing technology is relatively unlikely to discover fraud and the principal’s prior that the agent is ethical is low. In this case, the phase of doubt is partitioned into two stages: an initial stage in which the principal randomizes between auditing a submission and rejecting it outright and a second stage in which the principal
randomizes between auditing and outright approval. In this equilibrium, the audit plays two different roles. In the initial stage, where the principal randomizes between auditing and rejection, the audit is used to identify the rare real project which merits approval. Thus, as the principal’s trust in the agent grows, she becomes more-inclined to audit his project. When trust grows sufficiently, the principal begins to randomize between outright approval and auditing. Here, the audit serves to screen out fraud, and as the principal’s trust in the agent grows, she has less incentive to check his work. Thus, the equilibrium audit probability is non-monotonic over the phase of doubt, beginning at a low level, increasing until reaching one at the transition point between the two stages, and then decreasing until it reaches zero at the end of the phase of doubt, past which the principal is confident that she faces an ethical agent and accepts all submissions without auditing. 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.

Instead of increasing the principal’s information, the remedy that we analyze in Section 6 reduces the agent’s information. In particular, we consider a setting in which the agent interacts with one of two types of principal with different tolerances for accepting fakes. It is shown that both the high-standard and low-standard types of principal are better off (weakly or strictly) when the agent is uncertain about which type of principal he faces. In particular, when the principal is sufficiently likely to have high standards for approval, in equilibrium the agent behaves as if he only faces this type of principal. Thus, the high standards principal obtains the same equilibrium payoff under opacity as under transparency. However, if the principal’s true standard is low, then she strictly prefers to accept all submissions, even those in the phase of doubt. At the same time, the low standards principal faces a longer phase of doubt than under transparency, since the agent acts as if the principal’s standards for approval are high, making him reluctant to cheat. We show that despite this tradeoff, the low standards principal strictly benefits under opacity. In the remaining case, when the probability of the high standards principal is low, we show that the phase of doubt is divided into two stages: an initial stage of aggressive cheating, followed by the second stage of mild cheating. In this case, opacity strictly benefits both types of principal.

The remaining two remedies that we analyze endow the principal with limited commitment power. In Section 7 we consider a principal with commitment power who wishes to maximize her payoff while completely eliminating fraud. We show that this leads to a stationary optimal policy, which improves on the baseline equilibrium when the cost of producing a fake is relatively high. The fourth and final institutional remedy that we consider imposes constraints both on the principal’s ability to act and on her information. Specifically, in Section 8 we investigate an environment in which the principal makes a single acceptance decision at one specific date, and furthermore, she makes this decision without access to information.
about the time of project submission. Though it may seem that such an environment puts
the principal at a significant disadvantage—she must delay evaluation of projects that arrive
before the critical date and loses the ability to consider any projects that arrive after it—we
show that even such extreme constraints may benefit the principal when the probability of
a strategic agent is sufficiently high.

Given the dynamic nature of our model, our analysis is most applicable to settings where
real projects are associated with random breakthroughs or discoveries. For instance, the
agent could represent an entrepreneur touting a commercial innovation and the principal a
venture capitalist as in the case of Theranos mentioned above. Similarly, the agent could be
a research scientist seeking grant approval, patent, or publication of a specific discovery, as
in the cases of Stapel, Potti, Wansink, and Anversa. Another setting to which our model is
well-suited is a brokerage relationship, for example between a realtor and home seller. The
realtor may have ready access to a pool of low-value (fake) buyers, while locating a high-value
(real) buyer may take a substantial and uncertain length of time. Yet another relationship to
which our model speaks is that between a borrower (agent) and a lender (principal). In this
context, the borrower may report a negative shock to his income to the lender and request
renegotiation of the loan. The lender prefers to renegotiate if and only if the borrower has
experienced a real (not fake) shock to his income. Here, the auditing remedy is particularly
salient.

In the next section, we provide a review of related literature. We formally present the model
in Section 3. Section 4 contains a characterization of the equilibrium of the baseline model
and some comparative statics. Sections 5—8 address the four remedies mentioned above.
Concluding remarks appear in Section 9. All proofs have been relegated to the appendix.

2 Literature

At a broad level, our paper is related to a recent stream of work concerned with cheating,
consider the design of an optimal test, when the agent has the ability to manipulate the
process by which the test determines his type, showing that the possibility of manipulation
induces the principal to select a more-informative test. 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. Barron, Georgiadis, and Swinkels (2019) consider the design of compensa-
tion 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.

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 production delays 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.

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 censor less than he does in equilibrium. Similarly, Sun (2018) considers a dynamic model of censorship, demonstrating that when the arrival
of bad news is inconclusive, it is censored aggressively by the good type of ruler, which can improve information quality and lead to a Pareto improvement. In a different vein, Li, Matouschek, and Powell (2017) study power dynamics in a relational contract. In each period, the principal approves or vetoes an agent’s recommended project, without observing whether her own preferred project is available. Thus, the agent can suppress the arrival of the principal’s preferred project, hoping to implement his own.

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

In Klein (2016), the principal hires an agent to experiment by generating public information in the form of a state-contingent Poisson process. In addition to the experimentation technology, the agent has access to a specious technology which produces Poisson successes (that appear identical to the ones generated by the experimentation technology) at a rate that is independent of the state. Thus, a specious success is uninformative and worthless to the principal. The author shows that the optimal compensation contract backloads payments. Although the principal does not benefit from any genuine success after the first one, she pays the agent only when he has achieved some specified number of successes. 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. The first phase uses diminishing payments to incentivize effort as in a standard setting, but the agent’s ability to gamble limits the extent to which payments can decrease. In the second phase, the contract becomes stationary and the agent is no

---

2In Bergemann and Hege (1998, 2005), the type of the project is also uncertain, and only a good project can deliver a success. Thus when the agent shirks, he becomes more optimistic than the principal about the project’s type. In other words, shirking not only allows the agent to save on cost, but also generates private information.
longer punished for production delays. The principal in Varas (2018) learns about project quality post-submission, while the principal in our setting learns about the integrity of the agent pre-submission. More generally, Varas (2018) underscores the limits of high-powered incentive contracts, whereas our findings point to the crucial role played by reputation and trust in a setting marked by limited commitment.

Finally, the institutional remedies we consider are related in some degree to a body of prior work. Our analysis of auditing connects to the literature on costly state verification with seminal contributions by Townsend (1979) and Border and Sobel (1987). In contrast to this literature, we do not allow the principal to commit, either to her audit policy or to her acceptance rule. An intermediate case, where the principal can commit to some contract terms but cannot commit to her auditing strategy is studied by Khalil (1997) and Khalil and Lawarrée (2006). The final remedy we consider permits submissions only at a specified date. This replaces all honest arrivals with a single mass at the critical time, shortening the phase of doubt and increasing the principal’s welfare. This effect is reminiscent of Fuchs and Skrzypacz (2015), who show that committing to shut down a market at certain times stimulates trade when the market is open, enhancing efficiency.

3 The Model

A principal (she) interacts with an agent (he) over an indefinite horizon. Time is continuous and both parties discount future payoffs at rate \( \rho > 0 \). 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 \( \lambda \). The fraudulent technology allows the agent to instantly develop a fake project, at fixed cost \( \phi \in [0, 1) \). Thus, from the agent’s perspective, the authentic technology is free but slow, while the fraudulent technology is costly (if \( \phi > 0 \)) but fast.

---

3 See also Halac and Yared (2016) who embed costly verification into a model of delegation.

4 All results continue to hold qualitatively if the fraudulent technology yields a real project with a positive but relatively small probability.

5 If the use of the authentic technology imposed a positive flow cost on the agent, then he would have an additional reason to use the fraudulent technology: by doing so he could avoid paying the cost of authentic project development, i.e., shirk. Our goal is to study the dynamic incentive to commit fraud, which has received relatively little formal analysis, as opposed to the incentive to shirk, which has been researched extensively. Including a small positive flow cost for using the authentic technology would, however, have no impact on our results. Normalizing the expected flow cost under the real technology back to zero would manifest technically as a reduction in \( \phi \) (see calculations in Appendix). Furthermore, much of our analysis
The agent is one of two types: \textit{strategic} with probability $\sigma \in (0, 1)$ or \textit{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 $\sigma$ is common knowledge. Below whenever we write of the agent choosing whether to generate a fake project, it should be understood that we are referring to the strategic type.

Once a project has been developed, it is instantly submitted to the principal for approval.\footnote{This assumption rules out equilibria in which both types of agent are compelled to delay submitting projects by the principal’s off–path belief that submissions arriving at certain times must be fake. A refinement in the spirit of D1 would also eliminate such off–path beliefs and rule out equilibria involving delays. In addition, in all model variants, the equilibrium that we derive is robust to allowing the agent to delay the submission of the project.} The state of the project — real or fake — is not directly observable upon submission. In our baseline model, the principal only observes the time $t$ of submission when deciding whether to approve the project. We consider the possibility that the principal can audit a submission in Section 5.

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: false rejections are especially costly when $\theta$ is low and false approvals are especially costly when it is high. The strategic agent would like his project to be accepted regardless of its state, obtaining a gross benefit of 1 from acceptance, and 0 from rejection.\footnote{The analysis and results are similar if the agent’s payoff from having a fake project approved is smaller than his payoff from having a real project approved.}

\section{Equilibrium Characterization}

In this section, we characterize the weak Perfect Bayesian equilibrium (henceforth equilibrium) of the game and show that it is generically unique.\footnote{Multiple equilibria do exist iff $\phi = \frac{\lambda}{\rho + \gamma}$. To streamline presentation, we abstract from this and other non-generic knife-edge cases below.} 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

\footnote{6This assumption rules out equilibria in which both types of agent are compelled to delay submitting projects by the principal’s off–path belief that submissions arriving at certain times must be fake. A refinement in the spirit of D1 would also eliminate such off–path beliefs and rule out equilibria involving delays. In addition, in all model variants, the equilibrium that we derive is robust to allowing the agent to delay the submission of the project.}

\footnote{7The analysis and results are similar if the agent’s payoff from having a fake project approved is smaller than his payoff from having a real project approved.}

\footnote{8Multiple equilibria do exist iff $\phi = \frac{\rho}{\rho + \gamma}$. To streamline presentation, we abstract from this and other non-generic knife-edge cases below.}
given her beliefs, (iii) the principal’s belief about a submitted project is derived from Bayes’ rule.

**Strategies.** A pure strategy for the strategic agent is a choice of a “cheating time” \( t \in \{\mathbb{R}_+ \cup \infty\} \) at which he will submit a fake project if a real one has not yet arrived. A mixed strategy for a strategic agent is a probability measure over finite cheating times represented by cumulative distribution function \( F(\cdot) \).\(^9\) A strategy for the principal is an acceptance function \( a(\cdot) \) on the domain \( \mathbb{R}_+ \), which specifies the probability with which a submission at time \( t \) is approved.

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

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

\[
g(t) = \frac{\lambda}{\lambda + \mu(t)},
\]

where

\[
\mu(t) = \frac{\sigma f(t)}{1 - \sigma F(t)}.
\]

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

---

\(^9\)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.

\(^{10}\)Strictly speaking, the strategic agent can choose not to cheat with positive probability, thereby allocating some probability mass to \( t = \infty \). However, in all model variants that we consider except one, the strategic agent cheats with probability one in finite time in equilibrium. The only exception is considered in Section 7, where the principal makes a commitment that eliminates cheating.
Sequential Rationality. If the principal believes that a project that arrives at time $t$ is real (i.e., good) 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

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

The principal’s belief $g(t)$ given in (1) is decreasing in the hazard rate of the false arrival $\mu(t)$. Thus, a greater likelihood of fraudulent arrival reduces the principal’s confidence in the project, and (weakly) reduces the probability with which it is accepted.

Agent’s Optimal 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 his expected payoff is

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

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 will submit a fake project which costs $\phi$ 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 $\phi > \hat{\phi} \equiv \frac{\rho}{\rho + \lambda}$, then there is a unique equilibrium of the game and it involves the agent never submitting a fake project and the principal approving any submission she receives with probability 1.

The intuition is straightforward. Suppose for the moment that the principal approves any submission she receives with probability 1. Given the stationarity of the environment, the strategic agent effectively faces two alternatives: he can submit a fake immediately and earn payoff $1 - \phi$, or he can wait for a real arrival, earning payoff $\lambda/(\rho + \lambda) = 1 - \hat{\phi}$. Obviously, when $\phi > \hat{\phi}$, he prefers the second alternative. In this case, the agent never submits a fake, and it is sequentially rational for the principal to approve any submission she receives with probability 1. Indeed, this is the unique equilibrium of the game for $\phi > \hat{\phi}$. Motivated by this observation, we maintain the following assumption (without restatement) for the rest of the paper.
**Assumption 1.** The cost of faking is sufficiently low that an equilibrium in which all submissions are real does not exist: \( \phi < \hat{\phi} \).

When \( \phi < \hat{\phi} \), 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 principal’s best response would be to reject such a submission outright. 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 in equilibrium. Thus, given impatience, it must be that the higher payoff generated by the approval of an early fake project is offset by a lower probability of approval. 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 \( [0, \tilde{t}] \), where \( \tilde{t} \in (0, \infty) \).

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

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

In equilibrium, the interval of arrival times is divided into two phases: an early *phase of doubt* \( [0, \tilde{t}] \), in which the agent’s submission is treated with skepticism, inducing the principal to reject with positive probability, and a late *phase of credibility* \( [\tilde{t}, \infty) \) 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. Indeed, because the arrival time of real projects is distributed continuously, the strategic type must also mix continuously to keep a fake submission from being instantly identified and rejected. Furthermore, with \( \phi < \hat{\phi} \), even if the principal always accepts all arrivals, the agent prefers to submit a costly fake immediately rather than wait for a (free) real arrival. Thus, if the principal accepts with probability one at some time, then the strategic agent would not wait past this time to fake. Consequently, the equilibrium mixed strategy has no gaps.

Building on these observations, it is also possible to show that the phase of doubt must be finite (for \( \sigma < 1 \)). Mathematically, it is simple to show that the cheating rate, \( \mu(t) \), approaches zero as time passes, regardless of the agent’s strategy. It follows that at some finite time, the cheating rate becomes small enough that the principal strictly prefers to accept an arrival, and the agent never waits past this time to submit a fake, resulting in a finite phase of doubt.
Because the strategic agent mixes over the phase of doubt in equilibrium, his payoff \( u(t) \) must be constant. Using this observation, we show that the approval probability \( a(\cdot) \) is continuous and differentiable. Furthermore, because the agent is impatient, the approval probability must rise over time so as to maintain indifference between early fake submissions and later ones. Moreover, the acceptance probability approaches one at the end of the phase of doubt. Indeed, once the phase of doubt ends, the principal knows that the agent is ethical and she always approves his submissions. By implication, for times near the end of the phase of doubt, the principal must accept with probability approaching one; otherwise, the agent could benefit by delaying his fake until the phase of doubt is over.

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

\[
g(t) = \theta \Rightarrow \frac{\sigma f(t)}{1 - \sigma F(t)} = \frac{\lambda (1 - \theta)}{\theta}, \quad (4)
\]

\[
u'(t) = 0 \Rightarrow a'(t) - \rho a(t) + \phi (\rho + \lambda) = 0. \quad (5)
\]

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 faking 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 \( \overline{t} \) by solving \( F(\overline{t}) = 1 \). Finally, solving the second differential equation with boundary condition \( a(\overline{t}) = 1 \) yields the principal’s acceptance strategy. To characterize the equilibrium most succinctly, define

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

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

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

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

\[
F(t) = \frac{1}{\sigma} (1 - \exp(-\mu t)) \quad (6)
\]

supported on interval \([0, \overline{t})\), where

\[
\overline{t} = -\frac{\ln(1 - \sigma)}{\mu}. \quad (7)
\]
If $t \in [0,\bar{t})$, then the principal accepts with probability
\[ a(t) = \frac{\phi}{\phi^*} + \left(1 - \frac{\phi}{\phi^*}\right) \exp\{-\rho(\bar{t} - t)\}, \tag{8} \]
and with probability 1 otherwise.

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

**Payoffs.** The strategic agent’s equilibrium payoff is $U_S = a(0) - \phi$, and the ethical agent’s payoff is $U_E = U_S - (\phi^* - \phi) \exp(-\rho + \lambda)\bar{t})$. The principal’s payoff is
\[ V = (1 - \sigma)(1 - \theta) \int_{\bar{t}}^{\infty} \lambda \exp\{-\rho + \lambda\} s \, ds = (1 - \sigma)(1 - \theta) \frac{\lambda}{\rho + \lambda} \exp(-\rho + \lambda)\bar{t}) \tag{9} \]

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 function of a particular functional form: a constant, plus an exponential function with growth rate $\rho$, the discount rate.

To understand the shape of the acceptance strategy it is helpful to consider the differential equation for $a(\cdot)$, given in (4). This equation admits a particular solution with a constant acceptance function $a(t) = \frac{\phi}{\phi^*}$ and a complementary solution $a(t) = \exp(\rho t)$.\(^\text{11}\) 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 $\phi$, either if the game ends due to an exogenous shock (at rate $\rho$) or a real project arrives (at rate $\lambda$), resulting in a marginal gain of $\phi(\rho + \lambda)$. Thus, the constant term equates the marginal loss and gain from delay, leaving the agent indifferent. To see where the exponential term appears, note that (4) also accounts for the increases in the acceptance probability over time, reflected in the term $a'(t)$. Thus, by marginally delaying, the agent also increases the probability that his fake will be approved. To maintain the agent’s indifference, the increase in the acceptance probability must be exactly offset by discounting, resulting in an acceptance strategy that grows at rate $\rho$.

What about the agent’s strategy? If the principal mixes over the phase of doubt, then her belief that a submission is real must be constant, $g(t) = \theta$. Thus, from the principal’s

\(^\text{11}\)A constant particular solution satisfies $-\rho a + \phi(\rho + \lambda) = 0$; dividing by $\rho$ and substituting yields the claimed solution. The complementary solution satisfies $a'(t) - \rho a(t) = 0$ and is evidently exponential with growth rate $\rho$. \[ \text{13} \]
perspective, the arrival rate of a fake must also be constant, as in (5). In other words, accounting for her uncertainty about the agent’s type, the principal believes that a fake project arrives at constant rate $\mu$. Because the authentic technology has a constant arrival rate, the fraudulent technology must appear to the principal to be deployed with a constant rate so that she does not learn about the state of a project from observing the time at which it was submitted. Solving (5), we find that the only distributions with constant arrival rate $\mu$ are truncated exponentials of the form $\frac{1}{\sigma}(1 - \kappa \exp(-\mu t))$, where $\kappa$ is an integration constant. Finally, ruling out a mass point at zero yields, $\kappa = 1$, delivering the stated distribution.

Although the rate of faking is constant from the principal’s perspective, $\mu(t) = \mu$, the rate of faking is increasing from the agent’s perspective. Indeed, conditional on the agent being strategic, we can solve for the likelihood that the strategic agent fakes at time $t$, conditional on reaching time $t$ with no arrival,$^{12}$

$$\mu_S(t) = \frac{f(t)}{1-F(t)}.$$

Substituting the equilibrium mixing distribution, it is easy to see that $\mu_S(t)$ is increasing and approaches infinity as $t \to \bar{t}$. Thus, as time passes without an arrival and the principal becomes more convinced that the agent is the ethical type, the rate of faking by the strategic type accelerates, approaching infinity as the principal becomes certain that the agent is ethical.

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 equilibrium payoff must equal the expected return to faking at $t = 0$, where it is accepted with the lowest probability on the equilibrium path, $a(0)$. Unlike the strategic agent, the ethical agent has no opportunity to fake. In other words, he must “wait forever” to fake, which yields payoff $u(\infty)$. By implication, the ethical agent’s equilibrium payoff is strictly lower than the strategic agent’s, who could always mimic the ethical type’s strategic but strictly prefers not to (at least when $\phi < \hat{\phi}$). 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 faking, note that for a submission in the phase of doubt, the principal is indifferent between approval and rejection. She, therefore, expects zero from any submission during this phase. At $\bar{t}$, however, only the ethical agent is still active, because the strategic agent would have submitted a fake.

$^{12}$Note that if the agent is strategic, then the likelihood of a fake being submitted at time $t$ is $\exp(-\lambda t)f(t)$, while the likelihood of no arrival before time $t$ is $\exp(-\lambda t)(1 - F(t))$. Taking the ratio delivers the expression in the text.
with probability 1 by this point. Hence for \( t \geq \bar{t} \), the principal approves all submission for sure and earns \( 1 - \theta \). This is reflected in the formula for her expected payoff \( V \) given in Proposition 4.1.

When the principal must contend with faking, she obtains positive surplus 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 one, and his arrival, be it real or fake, yields the principal no surplus. Second, even if he is ethical, the agent must also be unlucky: if the ethical agent happens to be lucky and discover a real solution relatively quickly (i.e., before \( \bar{t} \)), the principal’s doubt about the agent’s credibility undermines the surplus that could be generated by this arrival. This normative implication is particularly pernicious for the principal, since the early arrivals, which under ordinary circumstances are least discounted and most valuable, are rendered worthless by the possibility of fraud.

We now turn to the comparative statics.

**Corollary 4.1** (Comparative Statics). The equilibrium exhibits the following comparative statics:

(i) A rise in \( \lambda \) results in a first order dominance shift of the agent’s cheating time distribution toward earlier cheating times, a decrease in \( \bar{t} \), and a strict increase in \( a(t) \) for all \( t < \bar{t} \). Payoffs increase for the principal and both types of agent.

(ii) A rise in \( \phi \) increases the approval probability for all \( t < \bar{t} \). It raises the payoff of the ethical agent and increases (decreases) the strategic agent’s payoff if \( \sigma \) is high (low).

(iii) A rise in \( \theta \) results in a first order dominance shift of the agent’s cheating time distribution toward later cheating times, an increase in \( \bar{t} \), and strict decrease in \( a(t) \) for all \( t < \bar{t} \). The strategic agent, the ethical agent, and the principal are all worse off.

(iv) A rise in \( \rho \) reduces the approval probability for all \( t < \bar{t} \). The strategic agent, ethical agent, and principal are all worse off.

First, consider an increase in \( \lambda \). When the real project’s arrival rate increases, the strategic type has more leeway to submit a fake and have it approved. Thus, the equilibrium arrival rate of a fake project, \( \mu \), scales up proportionately to the increase in \( \lambda \), and the principal’s belief stays fixed at \( \theta \) during the phase of doubt. Because the faking rate is higher, the waiting time for a fake to arrive is smaller, resulting both in a first order dominance shift toward earlier faking times and a collapse of the support of the faking time distribution toward zero, i.e., the phase of doubt becomes shorter (\( \bar{t} \) falls). The reduction in the phase of doubt suggests that the principal gains confidence in the agent more quickly when no
arrival is observed, shifting up her approval strategy. Indeed, a sufficient increase in $\lambda$ may even switch $\hat{\phi} > \phi$ to $\hat{\phi} < \phi$, in which case the strategic agent never fakes and the principal always approves in equilibrium. The fact that the agent’s equilibrium payoffs increase if the authentic technology improves is a direct consequence of these observations.

For the principal’s welfare, an increase in $\lambda$ has conflicting effects. On one hand, if the phase of credibility is reached, then the principal is better off because her valuable real arrival is less discounted. On the other hand, with a higher real arrival rate, it is more likely that an ethical agent will generate an arrival during the phase of doubt, which is worthless to the principal. Crucially, an increase in the real arrival rate also collapses the phase of doubt, so the phase of credibility begins earlier. Taken together, it is possible to show that the principal’s payoff improves as the real arrival rate increases, but for this result, the third effect is crucial.

Next, consider an increase in the cost of generating a fake project, $\phi$. This change makes faking less attractive to the strategic agent, so the principal must approve submissions in the phase of doubt with higher probability in order to maintain his indifference. This indirect effect has natural intuition: the higher faking cost acts as a deterrent against cheating, which induces the principal to accept more often. By implication, the payoff of the ethical agent is higher. Furthermore, the indirect effect also has a positive effect on the payoff of the strategic agent, but this must be compared with the higher cost of faking, which directly harms the strategic agent. For low values of $\sigma$, the phase of doubt is short and approval rates are high – so the indirect effect is muted and dominated by the direct effect. The reverse holds for large values of $\sigma$. The principal’s payoff does not depend on $\phi$ directly or indirectly when $\phi < \hat{\phi}$. It is interesting to note, however, that the principal’s payoff jumps down discontinuously, from $\lambda/ (\lambda + \rho)$ to $V$, as $\phi$ crosses through $\hat{\phi}$ from above.

In the ensuing four sections, we explore distinct institutional remedies designed to benefit the principal relative to her equilibrium payoff in the baseline model. These include: an imperfect auditing technology, opaque standards, commitment to suppress fraud, and requiring submissions to occur on a specified date.

5 Auditing

In this section, we improve the principal’s access to information, endowing her with the ability to uncover fraud before deciding whether to approve a submission. In particular, suppose that the principal can pay $k \in (0, 1)$ to perform a test on a submission that reveals fraud with probability $\alpha \in (0, 1]$. That is, if the project is fake, it fails the test with probability $\alpha$ and (falsely) passes the test with probability $1 - \alpha$. Furthermore, a real project always
passes the test.\footnote{Allowing for real submissions to falsely fail the test with positive probability complicates calculations significantly without adding additional insight.}

\begin{center}
\begin{tabular}{|l|c|c|}
\hline
 & Pr(Pass) & Pr(Fail) \\
\hline
Real Project & 1 & 0 \\
Fake Project & $1 - \alpha$ & $\alpha$ \\
\hline
\end{tabular}
\end{center}

We classify an audit into one of two groups according to its ability to uncover fraud.

**Definition 1 (Strength of Audit).** The test is called strong if $\alpha > \hat{\phi} - \phi$, and weak if $\alpha < \hat{\phi} - \phi$.

Throughout we focus on audits that are either strong or weak, abstracting from the knife-edge case $\alpha = \hat{\phi} - \phi$.

As in the main model, let $g(t)$ represent the principal’s belief that the project is good, conditional solely on the project’s arrival time. We begin with the following observation that comes from basic reasoning about the value of information.

**Observation 5.1.** Suppose the principal believes a project is good with probability $g(t)$. If it is sequentially rational for the principal to test the project, then it is also optimal to approve the submission if and only if it passes the test.

To see why this must hold, 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.

Following an arrival at time $t$, the principal chooses between testing, accepting with no audit, and rejecting with no audit. She, therefore, compares the following three payoffs,

\[
\begin{align*}
\text{Audit:} & \quad g(t)(1 - \theta) - \theta(1 - g(t))(1 - \alpha) - k \\
\text{Accept:} & \quad g(t) - \theta \\
\text{Reject:} & \quad 0
\end{align*}
\]

Note that, consistent with Observation 5.1, the expected payoff of auditing is calculated assuming that the outcome of the test determines the principal’s acceptance decision: with probability $g(t)$ the project is authentic and passes the test, with probability $(1 - g(t))(1 - \alpha)$ the project is fraudulent but nevertheless passes and is approved. Let $p(t)$ be the probability that the principal tests a project that arrives at time $t$, $a(t)$ be the probability that the
principal approves the arrival with no test, and \( r(t) \) be the probability that the principal rejects with no test.

Comparing the principal’s payoffs from each of the three alternatives, we find the following sequentially rational strategy. If \( k > k^* \equiv \alpha \theta (1 - \theta) \), then the principal never audits, and follows an identical strategy to the one in equation (2). If \( k \leq k^* \), then the principal follows the following strategy,

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

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 if \( k < k^* \), then \( \theta_0 < \theta < \theta_1 \). Thus, if the auditing cost is not too high, then the principal accepts when her belief is high, rejects when her belief is low, and audits for an interval of moderate beliefs surrounding \( \theta \). The following result is immediate.

**Observation 5.2** (Equilibrium with no testing.). *An equilibrium in which the principal never tests any arrival exists if and only if \( k \geq k^* \).*

Motivated by this observation, assume \( k \in (0, k^*) \) for the rest of the section. By implication, the principal must audit some arrivals in equilibrium.

We argue next that the principal’s audit must be conducted at random. To see why, suppose to the contrary there is an equilibrium in which the principal audits every submission in some interval \((t_1, t_2)\). Intuitively, if the test is strong, then the agent’s expected payoff from submitting a fake is increasing throughout the interval (i.e., he would like to submit one as late as possible), and if the test is weak, then his expected payoff from submitting a fake is decreasing throughout the interval (i.e., he would like to submit one as soon as possible). In either case, fakes are not submitted in the interior of the interval, and the principal should, therefore, approve any submission at \( t \in (t_1, t_2) \) without performing a costly audit.

**Observation 5.3.** In equilibrium, no interval \( T \) exists such that \( p(t) = 1 \) for all \( t \in T \).

Given that the principal must audit randomly in equilibrium, there are two types of strategies that she can follow. In the “default approval” strategy, the principal mixes between auditing

\[14\text{Furthermore, } k = k^* \iff \theta_0 = \theta = \theta_1.\]
and outright approval, \( a(t) + p(t) = 1 \), and in the “default rejection” strategy, she mixes between auditing and rejection, \( p(t) + r(t) = 1 \). Intuitively, we might expect that the principal uses the default approval strategy when she has a sufficient degree of trust in the agent. Thus, if she adopts this strategy at some time, then she will maintain this strategy at all subsequent times, where her trust in the agent is higher. Conversely, if the principal does not have sufficient trust in the agent to maintain default approval at some time, then the principal will reject by default. At all earlier times, when the principal has even less trust in the agent, she will also reject by default. This intuition is formalized in the following lemma, which characterizes the equilibrium structure when \( k < k^* \).

**Lemma 5.1** (Auditing equilibrium structure). In an equilibrium with auditing, there exists \( \tilde{t}_A \in (0, \infty) \) and \( \tilde{t}_A \in [0, \tilde{t}_A) \) such that

(i) the agent’s falsification time is drawn from a continuous mixed strategy with no mass points or gaps supported on an interval \([0, \tilde{t}_A] \).

(ii) for \( t \in [\tilde{t}_A, \infty) \), the principal accepts the project with no audit, \( a(t) = 1 \).

(iii) for \( t \in (\tilde{t}_A, \tilde{t}_A) \) the principal mixes between auditing and outright acceptance, \( p(t) \in (0, 1) \) and \( a(t) + p(t) = 1 \). The probability of auditing is strictly decreasing, continuous, and differentiable almost everywhere, with \( \lim_{t \to \tilde{t}_A} p(t) = 0 \).

(iv) if \( \tilde{t}_A > 0 \), then for \( t \in [0, \tilde{t}_A) \) the principal mixes between auditing and outright rejection, \( p(t) \in (0, 1) \) and \( a(t) = 0 \). The probability of auditing is strictly increasing, continuous, and differentiable, with \( \lim_{t \to \tilde{t}_A} p(t) = 1 \). Furthermore, for \( t \in (\tilde{t}_A, \tilde{t}_A) \), we have \( \lim_{t \to \tilde{t}_A} p(t) = p(\tilde{t}_A) = 1 \).

The structure of the auditing equilibrium has two particularly intriguing features. First, the principal’s default action may switch from rejection to acceptance at some interior time \( \tilde{t}_A \). That is, the phase of doubt may consist of two stages, one involving default rejection followed by one involving default acceptance. Second, in an equilibrium with this feature, the audit probability is non-monotone, increasing until the transition point, \( \tilde{t}_A \), and decreasing thereafter.

To understand these features formally, note that (as in the main model) the agent gains credibility over time, and, it is fully restored by some finite date. Because the principal accepts outright once the agent’s credibility is fully restored, if the agent were to submit a fake at such a time, then the fake would be accepted with probability one. Thus, to avoid a discontinuity in the agent’s payoff near the end of the phase of doubt (which would violate his indifference condition), the probability of accepting a fake submission must approach one.
as the phase of doubt ends. This is only possible if the principal plays a default acceptance strategy at the end of the phase of doubt. Indeed, if the principal plays a default rejection strategy, then a fraudulent project is only accepted if it passes the audit by mistake (i.e., with probability $p(t)(1 - \alpha)$). Thus, under default rejection, the probability that a fake is accepted is bounded above by $1 - \alpha$, and it obviously cannot approach one as the phase of doubt ends. In contrast, if the principal plays a default acceptance strategy, then a fake is accepted unless it is audited and fails (i.e., with probability $1 - \alpha p(t)$). Thus, if the audit probability approaches zero at the end of the phase of doubt, then the probability that a fake is accepted approaches one, as required. Building on this observation, it is possible to show that the probability of an audit must be decreasing during the default acceptance stage, and approach zero at the end of the phase of doubt.

Depending on parameter values, the equilibrium may also involve a default rejection stage. A similar continuity argument establishes that at the threshold time $\tilde{t}_A$, the probability of audit must be one. Indeed, during the default rejection stage the probability that a fraudulent arrival is accepted is $p(t)(1 - \alpha)$, while during the default acceptance stage it is $1 - \alpha p(t)$. Continuity at $\tilde{t}_A$ implies $p(\tilde{t}_A) = 1$, and it is possible to show that the audit probability must therefore be strictly increasing over the default rejection stage.\textsuperscript{15}

Lemma 5.1 suggests that the audit may play two distinctly different roles in equilibrium. The agent’s reputation for being ethical rises steadily over the phase of doubt. Audits are rarely performed for early submissions because they are likely to be fakes produced by an unethical agent and are therefore typically rejected out of hand. Audits are also rarely performed for submissions late in the phase of doubt, but for the opposite reason. These submissions are most likely real ones produced by an ethical agent and are, therefore, usually approved without an audit. Hence, the probability of an audit rises over the initial stage of default rejection; peaks at a value of 1 at the transition point; and declines over the stage of default approval.

To study further the interplay between auditing, fraud, and the agent’s credibility, we provide a complete characterization of the two possible equilibrium structures. We begin by analyzing the equilibrium in which the entire phase of doubt involves default acceptance, corresponding to $\tilde{t}_A = 0$ in Lemma 5.1, which we refer to as the “one stage auditing equilibrium.” We then analyze the “two stage auditing equilibrium,” in which a phase of default rejection is followed by a phase of default acceptance, corresponding to $\tilde{t}_A > 0$ in Lemma 5.1.

\textsuperscript{15}Because the audit probability must be increasing during the penultimate default rejection stage, decreasing during the final default acceptance stage, and a transition between default rejection and acceptance is possible only if the audit probability is one, in equilibrium the principal can switch only once between a default rejection and acceptance strategy.
To facilitate exposition, for \( \theta_i \in \{ \theta_0, \theta_1 \} \) (as defined in (10)) let

\[
\mu_i \equiv \lambda \frac{1 - \theta_i}{\theta_i} \quad \tilde{t}_i \equiv -\frac{\ln(1 - \sigma)}{\mu_i}.
\]

Note that \( \mu_i \) is the equilibrium cheating rate in the equilibrium of the baseline model when the principal’s preference parameter is \( \theta_i \), and \( \tilde{t}_i \) is the duration of the corresponding phase of doubt. Furthermore, for a weak test let

\[
\delta_A \equiv -\frac{-\ln(1 - \frac{\alpha}{\phi - \rho})}{\lambda + \rho} \quad \tilde{t} \equiv \max \left\{ 0, \left( \frac{\mu_1}{\mu_0} (\tilde{t}_1 - \delta_A) \right) \right\}.
\]

As we will see, parameter \( \delta_A \) gives the duration of the second “default acceptance” phase in the two stage auditing equilibrium.

**Proposition 5.1** (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 \( \tilde{t} = 0 \), 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 \([0, \tilde{t}_1] \). If \( t \in [0, \tilde{t}_1] \), then the principal mixes between auditing and outright approval, and the probability of audit is

\[
p(t) = \frac{\phi - \phi}{\alpha}(1 - \exp\{-\rho + \lambda (\tilde{t}_1 - t)\}).
\]

If \( t \geq \tilde{t}_1 \), then the principal does not audit and accepts with probability 1.

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

**Payoffs.** The strategic agent’s equilibrium payoff is \( U_{SA}^S = 1 - \alpha p(0) - \phi \). The principal’s payoff is

\[
V_A = \lambda \left( 1 - \frac{\theta}{\theta_1} \right) \int_0^{\tilde{t}_1} \exp\{-(\rho + \frac{\lambda}{\theta_1})t\} dt + (1 - \theta)(1 - \sigma) \int_{\tilde{t}_1}^{\infty} \lambda \exp\{-(\rho + \lambda)t\} dt.
\]

Thus a one stage auditing equilibrium—in which the principal mixes between approving and auditing during the entire phase of doubt—exists in two scenarios: when the test is strong or when it is weak but the agent has sufficiently high initial credibility (\( \tilde{t}_1 < \delta_A \)). If the test is strong, then the strategic type is reluctant to submit a fake project. If the agent has high credibility, then his submission is again, likely to be real. In both cases, the principal can afford to rubber-stamp some submissions without auditing, and as the agent gains credibility, the principal becomes even more inclined to do so. On the other hand, when the test is weak and the agent’s initial credibility is low, a two stage auditing equilibrium must obtain.
Proposition 5.2 (Two stage auditing equilibrium.). A two stage auditing equilibrium exists if and only if the test is weak and \( \tilde{t} > 0 \), 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

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

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

\[
p(t) = \frac{\phi}{\phi - \alpha} + \left(1 - \frac{\phi}{\phi - \alpha}\right) \exp \left(-\left(\frac{\phi}{1 - \alpha}\right)(\rho + \lambda)(\tilde{t}_A - t)\right).
\]

For \( t \in [\tilde{t}_A, t_A] \) the principal randomizes between approval and auditing, and the audit probability is

\[
p(t) = \frac{\phi}{\phi - \alpha} (1 - \exp(-(\rho + \lambda)(\tilde{t}_A - t)));
\]

and for \( t \geq \tilde{t}_A \) the principal approves all submissions without performing an audit.

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

**Payoffs.** The agent’s equilibrium payoff is \( U^S_A = (1 - \alpha)p(0) - \phi \). The principal’s payoff is

\[
V_A = \exp\left(-(\mu_0 - \mu_1)\tilde{t}_A\right) \left(1 - \frac{\theta}{\theta_1}\right) \int_{\tilde{t}_A}^{t_A} \lambda \exp\left(-(\rho + \frac{\lambda}{\theta_1})t\right) dt + (1 - \theta)(1 - \sigma) \int_{t_A}^{\infty} \lambda \exp(-(\rho + \lambda)t) dt.
\]

When the test is weak and the agent’s initial credibility is low, the relationship between the principal and agent passes through two stages. In the first stage, the strategic agent cheats aggressively, and the principal views submissions skeptically: in order to be approved, the submission must pass the test. At early times, the principal uses an occasional test to try to fish out an unlikely real arrival. Once enough time passes, the agent gains enough credibility that the principal’s strategy changes: she accepts the project unless it fails the test. At later times, the principal uses the test as a monitoring device, which limits the agent’s incentive to cheat. As the agent gains additional credibility, the principal checks his work less often, until she stops checking entirely, at the point that she becomes convinced that the agent must be ethical.

It is worth pointing out that the conditions for existence of the one stage and two stage auditing equilibrium are mutually exclusive, and thus, the equilibrium in the model with auditing is (generically) unique.
To understand auditing’s normative implications, let us first consider the one stage equilibrium. Here the agent’s mixed strategy is identical to that of the main model, where the principal does not audit but has preference parameter $\theta_1 > \theta$. Thus, during the phase of doubt, the agent cheats more slowly, at 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 accepting outright and auditing, and thus her expected payoff is the same as if she outright accepts. Therefore, from a normative perspective, the principal faces a tradeoff between short-term gain and long-term loss. On one hand, an arrival during the phase of doubt generates positive surplus for the principal, which improves her payoff. On the other hand, the strategic agent cheats more slowly, and it takes more time for credibility of the ethical agent to be fully restored, which hurts her payoff. In the proposition below, we show that the short-term gain dominates—auditing benefits the principal in the one stage equilibrium.

In the two stage equilibrium, an additional complication arises. In the second stage (with default acceptance), the same basic tradeoff applies as in the one stage equilibrium—the slowdown in cheating increases the principal’s surplus, but delays the onset of the phase of credibility. However, during the first stage the principal is indifferent between testing and rejection, and therefore obtains no surplus. At the same time, the agent cheats rapidly during the default rejection stage (i.e., $\mu_0 > \mu > \mu_1$), and thus, full credibility is restored more quickly than it would have been if the equilibrium consisted of a single default acceptance stage (i.e. $\bar{t}_A < \bar{t}_1$). We show that in this case the principal also benefits from the ability to audit.

It is interesting to note that in both types of equilibrium, the principal is indifferent between auditing and her default action, and thus, the value of information coming from the audit is exactly equal to its cost. Therefore, the normative gains for the principal do not come directly from her ability to acquire information about the project’s type once it has been submitted. Rather, these gains come indirectly, via the effect on the strategic agent’s incentive to cheat.\footnote{In other words, if the agent continues to play his equilibrium strategy, then removing the principal’s ability to audit has no effect on her payoff.}

**Proposition 5.3** (Normative Analysis of Auditing.). In either the one stage auditing equilibrium or the two stage auditing equilibrium, 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 $\sigma$ is sufficiently large.

Remarkably, both types of agent can also benefit when the principal has the ability to audit a submission. Because the ethical agent only submits a real project, under default
acceptance the ethical agent’s submissions are always approved. Thus, in the one stage auditing equilibrium, the ethical agent gets his first best payoff. Similarly, in the two stage equilibrium, the ethical agent’s projects are always approved past time $t_A$; before the default acceptance phase, $t < t_A$ the ethical agent’s project is approved whenever it is tested. It is relatively straightforward to show that in the two stage equilibrium, the probability of testing for $t < t_A$ 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. Consequently, the ethical agent’s equilibrium payoff is also higher.

Notably, 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 $\sigma$ is large, the strategic type also has a higher payoff under auditing than in the baseline model, and in these cases, auditing generates a Pareto improvement.

6 Opaque Standards

In this section, we consider the possibility that the principal withholds information from the agent. In particular, we analyze an environment in which the principal can be one of two types. With probability $\nu$ the principal has a high standard, $\theta_H$, and with probability $1 - \nu$, the principal has a low standard, $\theta_L$ where $\theta_H > \theta_L$. With “transparent standards” the principal’s type is observed by the agent at the beginning of the interaction. For each realization of the principal’s type, the equilibrium is identical to the main model. With “opaque standards” the agent cannot observe the principal’s type. We characterize the equilibrium with opaque standards, and analyze its normative properties.

Denote the acceptance strategy of each principal $a_i(\cdot)$, where $i \in \{H, L\}$, and let $a_U(t) \equiv \nu a_H(t) + (1 - \nu) a_L(t)$ denote the expected probability of acceptance at time $t$, accounting for the agent’s uncertainty about the principal’s type. The agent’s expected payoff of se-
lecting cheating time \( t \) is identical to his payoff in the main model, substituting the expected acceptance probability \( a_U(\cdot) \) for the acceptance probability \( a(\cdot) \) of the main model. Similar arguments to those in Lemma 4.3 establish that the agent mixes continuously on an interval from time zero to some finite threshold time \( \tilde{t}_U \), defining a finite “phase of doubt.” Furthermore, over this interval of times, the expected acceptance probability inherits the features of the acceptance probability in the main model: \( a_U(\cdot) \) is strictly greater than \( \phi \), increasing, continuous, differentiable, and approaches one at the end of the phase of doubt. After the phase of doubt, the expected acceptance probability is one, \( a_U(t) = 1 \) for \( t > \tilde{t}_U \).

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

**Lemma 6.1** (Opaque standards equilibrium structure.). In equilibrium with uncertain standards, there exists \( \tilde{t}_U \in (0, \infty) \) and \( \tilde{t}_U \in [0, \tilde{t}_U) \) such that

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

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

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

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

her preferences, the agent does not update beliefs about the principal over time.
To complete the characterization, we separately consider equilibria with a “one stage” structure, corresponding to the case $\bar{t}_U = 0$, and a “two stage” structure, corresponding to $\bar{t}_U > 0$. In anticipation of the characterization, we introduce some additional notation that simplifies the exposition. For $i \in \{H, L\}$, let

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

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

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

**Proposition 6.1** (Opaque standards, one stage.). *With opaque standards, a one phase equilibrium exists if and only if $\nu > \nu^*$, and it is characterized below. Furthermore, with opaque standards, no other one phase equilibrium exists.*

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

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

26
supported on interval $[0, \bar{t}_H]$. If $t \in [0, \bar{t}_H]$, then the high type principal accepts with probability
\[ a_H(t) = \frac{1}{\nu} \left( \frac{\phi}{\phi} + (1 - \frac{\phi}{\phi}) \exp\{ -\rho(\bar{t}_H - t) \} - (1 - \nu) \right), \]
and with probability 1 otherwise. The low type principal always accepts, $a_L(t) = 1$. The expected acceptance probability $a_U(\cdot)$ is identical to the acceptance probability in the main model, with principal’s standard known to be $\theta_H$.

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

**Payoffs.** The strategic agent’s equilibrium payoff is $a_U(0) - \phi$, identical to the main model with principal’s standard known to be $\theta_H$. The high standards principal’s payoff is
\[ V_H = (1 - \theta_H)(1 - \sigma) \int_{\bar{t}_H}^{\infty} \lambda \exp\{ -(\rho + \lambda)t \} dt, \]
and the low type principal’s payoff is
\[ V_L = \lambda (1 - \frac{\theta_L}{\theta_H}) \int_{0}^{\bar{t}_H} \exp\{ -(\rho + \frac{\lambda}{\theta_H})t \} dt + (1 - \theta)(1 - \sigma) \int_{\bar{t}_H}^{\infty} \lambda \exp\{ -(\rho + \lambda)t \} dt. \]

**Normative Ranking.** In the one phase equilibrium with opaque standards, (i) the high type principal’s payoff is the same as in the unique equilibrium with transparent standards. (ii) The low type principal’s payoff is strictly higher than in the unique equilibrium with transparent standards. (iii) the strategic and ethical agents’ expected payoffs are strictly lower than in the unique equilibrium with transparent standards.

When the principal is likely to have high standards, opacity (weakly) increases the payoffs of both types of principal and strictly reduces the payoffs of both types of agent. Because the agent is likely to interact with the high standards principal, it is the high type’s incentive to accept that restrains the strategic agent’s incentive to cheat. In equilibrium, the strategic type effectively targets only the high type principal, completely ignoring the low type. In particular, the agent’s strategy is identical to the baseline model, assuming that the principal’s standard is known to be $\theta_H$. Thus, the high type principal obtains the same equilibrium payoff with opacity as with transparency. In contrast, when the principal’s realized standard is low, she accepts every submission, obtaining a positive payoff in both the doubt and credibility phases. However, the agent cheats more slowly with opacity than transparency if he faces the low type principal; thus, opacity also delays the onset of the credibility phase for the low type. Indeed, the low type principal faces the exact same tradeoff with opacity as she does in the “default acceptance” equilibrium with auditing: a higher belief during the phase of doubt (and acceptance of all arrivals), but a longer phase of doubt. In our analysis of
auditing, we showed that the benefit generated by a higher belief during the phase of doubt outweighed the delay in restoring credibility. The analogous finding holds under uncertain standards; i.e., opacity strictly benefits the low type principal.

From the strategic agent’s perspective, the expected acceptance probability $a_U(\cdot)$ is identical to the acceptance probability in the main model, assuming he faces a principal with high standards. Thus, the strategic agent’s payoff is as if he always faces a principal with high standards (regardless of the principal’s actual type). Under transparency, whenever the principal actually has low standards, the strategic agent cheats more aggressively in equilibrium, and his submission is accepted more often, both of which benefit the agent. It follows that the strategic agent is harmed by opacity in the one phase equilibrium: when the principal is a high type the agent is indifferent, but when the principal is a low type he prefers transparency. Similar logic applies to the ethical agent, who (from Corollary 4.1), also benefits when facing a principal with lower standards.

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

**Proposition 6.2** (Opaque standards, two Stages.). With opaque standards, a two stage equilibrium exists if and only if $\nu < \nu^*$, and it is characterized below. Furthermore, with opaque standards, no other two stage equilibrium exists.

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

$$\tilde{t}_U = \bar{t}_L - \frac{\mu_H}{\mu_L} \delta_U \quad t_U = \tilde{t}_L + \delta_U.$$  

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

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

$$F(t) = \begin{cases} \frac{1}{\sigma}(1 - \exp(-\mu_L t)) & \text{for } t \in [0, \tilde{t}_U) \\ \frac{1}{\sigma}(1 - \exp(-\mu_H t - (\mu_L - \mu_H)\tilde{t}_U)) & \text{for } t \in [\tilde{t}_U, t_U] \end{cases}$$
supported on $[0, \tilde{t}_U]$. If $t \in [0, \tilde{t}_U]$, then the high type principal always rejects, $a_H(t) = 0$, and the low type principal accepts with probability
\[
a_L(t) = \left( \frac{1}{1 - \nu} \right) \left( \frac{\nu}{\phi} + \frac{\phi - \nu}{\phi} \right) \exp \{ -\rho(\tilde{t}_U - t) \}.
\]
If $t \in [\tilde{t}_U, t_U]$, then the high type principal accepts with probability
\[
a_H(t) = \frac{1}{\nu} \left( \frac{\nu}{\phi} + \frac{\phi - \nu}{\phi} \right) \exp \{ -\rho(\tilde{t}_U - t) \} - (1 - \nu),
\]
and the low type principal always accepts, $a_L(t) = 1$. If $t \geq t_U$, then both types of principal always accept, $a_L(t) = a_H(t) = 1$.

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

**Payoffs.** The agent’s equilibrium payoff is $a_U(0) - \phi$. The high standards principal’s payoff is
\[
V_H = (1 - \theta_H)(1 - \sigma) \int_{\tilde{t}_U}^{\infty} \lambda \exp \{ -(\rho + \lambda) s \} ds.
\]
The low type principal’s payoff is
\[
V_L = \exp \{ -(\mu_L - \mu_H) \tilde{t}_U \} (1 - \frac{\theta_L}{\theta_H}) \int_{\tilde{t}_U}^{t_U} \lambda \exp \{ -(\rho + \frac{\lambda}{\theta_H}) t \} dt + (1 - \theta_L)(1 - \sigma) \int_{\tilde{t}_U}^{\infty} \lambda \exp \{ -(\rho + \lambda) t \} dt.
\]

**Normative Ranking.** In the two stage equilibrium with opaque standards, (i) the high type principal’s payoff is strictly higher than in the unique equilibrium with transparent standards. (ii) The low type principal’s payoff is strictly higher than in the unique equilibrium with transparent standards. (iii) the strategic agent’s payoff is strictly lower with opacity for $\nu$ close to $\nu^*$ than in the unique equilibrium with transparency. If $\nu$ is sufficiently small, then the strategic agent’s payoff may be either larger or smaller under opacity than transparency, depending on parameters.

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

It is worth pointing out that endowing the principal with a disclosure technology by which she can prove her type to the agent at the outset does not undermine the equilibrium that we construct. That is, even if the principal has such a technology, an equilibrium exists in which neither type of principal uses it.\textsuperscript{19}

7 Deterring Fraud

In this section, we explore the possibility that the principal can use her acceptance strategy to deter fraud. Specifically, we ask, if the principal could commit to any approval strategy, what is the optimal way to ensure that no fraud takes place in equilibrium? Furthermore, under what conditions is deterring fraud unambiguously better for the principal than the equilibrium of the baseline model?

Formally, we consider a benchmark model in which the principal has the ability to commit to any measurable strategy \( a(\cdot) \). We are interested in optimal “fraud deterrence,” and thus, we focus on acceptance strategies that induce the strategic agent to never submit a fake. If the agent never fakes, then he expects payoff \( u(\infty) \). To induce the agent never to fake, the principal must ensure that his payoff of faking at any time \( t \) is smaller than his payoff of waiting forever. Thus, the principal faces an incentive constraint, \( u(t) \leq u(\infty) \) for all \( t \geq 0 \).

Given that the incentive constraint deters faking, the principal’s expected payoff is simply the discounted payoff of accepting a real arrival under her strategy,

\[
(1 - \theta) \int_0^\infty \lambda \exp(-\rho t) a(t) dt = (1 - \theta)u(\infty).
\]

Thus, the principal’s problem is to maximize \( u(\infty) \), subject to the incentive constraint \( u(t) \leq u(\infty) \) for all \( t \geq 0 \).\textsuperscript{20}

We make the crucial observation that the incentive constraint must bind for all \( t \geq 0 \).

\textsuperscript{19}Because both types of principal prefer opacity to transparency (at least weakly), if the type \( i \) principal is expected not to disclose her type, then it is a best response for the type \( j \) principal not to disclose. In some cases, equilibria with partial disclosure also exist.

\textsuperscript{20}Note that neither the equilibrium strategy, nor \( a(\cdot) = 1 \) are incentive compatible when \( \phi < \hat{\phi} \). In particular, when \( a(\cdot) = 1 \) the payoff function \( u(\cdot) \) is strictly decreasing, and in equilibrium with \( \phi < \hat{\phi} \), the agent’s payoff function \( u(\cdot) \) is constant on the phase of doubt, and then strictly decreasing.
Lemma 7.1. In the solution to the principal’s problem, \( u(t) = u(\infty) \) for all \( t \geq 0 \).

Because the incentive constraint binds at all times in the solution to the principal’s problem, i.e. \( u(t) = u(\infty) \) for all \( t \), we can use an argument analogous to Lemma 4.3 to show that \( a(\cdot) \) must be continuous and differentiable. Furthermore, \( u(t) = u(\infty) \) also implies \( u'(t) = 0 \) for all \( t \geq 0 \), and hence, (5) must hold at all times. Thus, we find

\[
a(t) = \frac{\phi}{\hat{\phi}} + \kappa \exp(\rho t)
\]

for all \( t \geq 0 \). It immediately follows that \( \kappa = 0 \); otherwise, \( a(\cdot) \) becomes negative or exceeds one for large \( t \).

Proposition 7.1 (Deterring Fraud.). The optimal acceptance strategy that deters fraud is \( a(t) = \frac{\phi}{\hat{\phi}} \). If the principal commits to this acceptance strategy, then her payoff is

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

There exists \( \phi^* < \hat{\phi} \) such that the principal’s payoff under the optimal fraud deterrence strategy exceeds her payoff in the baseline model if and only if \( \phi > \phi^* \).

The optimal acceptance strategy in the principal’s problem is constant over time—every arrival is simply accepted with probability \( \frac{\phi}{\hat{\phi}} \) regardless of when it arrives. Thus, when faking is difficult, i.e. \( \phi \) is close to \( \hat{\phi} \), the principal does not need to distort her acceptance strategy much from the first best \( (a(\cdot) = 1) \) in order to deter faking, and when faking is easy she must reduce the acceptance probability almost to zero. Because the optimal policy does not depend on time, it can be implemented in a fairly straightforward manner. However, doing so requires significant commitment power: the principal must commit to reject arrivals with a positive probability that she knows are real.

To understand the normative implications of the fraud deterrence strategy, recall that in the absence of commitment, the principal obtains no surplus from arrivals during the phase of doubt, but obtains surplus \( (1 - \theta) \) from arrivals during the phase of credibility, and this phase is reached only if the agent is ethical, probability \( 1 - \sigma \). Thus, her payoff in the baseline equilibrium can be computed as if all arrivals are real, but her acceptance strategy is a step function: \( a(t) = 0 \) during the phase of doubt \( (t < \bar{t}) \), and \( a(t) = 1 - \sigma \) during the phase of credibility \( (t \geq \bar{t}) \). In contrast, with the fraud deterrence strategy, all arrivals are real, but the principal’s acceptance probability is constant over time, equal to \( \phi/\hat{\phi} \). Thus, from a normative perspective, fraud deterrence increases the acceptance probability for early arrivals

\[ \text{Recall that } V = (1 - \theta)(1 - \sigma) \int_T^\infty \lambda \exp(- (\rho + \lambda) t) dt. \]
and (potentially) reduces it for late arrivals. Furthermore, this shift may increase the overall acceptance probability: indeed, when the cost of faking is close to $\hat{\phi}$, under fraud deterrence, all real arrivals are accepted with probability close to one, and thus, the principal’s payoff is higher than in the baseline model. Conversely, when the cost of producing a fake is low, the principal must commit to accept with low probability in order to deter fraud. In this case, not making any commitment to deter fraud delivers a higher payoff to the principal.

8 Fixed Duration

In the baseline model, the principal evaluates all submissions, no matter how early. The possibility that the real technology could generate an early arrival gives the strategic agent sufficient cover to submit fakes early, undermining the credibility of the most valuable (least discounted) real arrivals, thereby rendering them worthless to the principal. In this section, we endow the principal with a simple tool to blunt this incentive: we allow her to select a fixed project duration, $T$, committing not to allow submissions at any other time. Thus, the strategic agent does not have any incentive to submit a fake before $T$, since he can always “wait and see” whether he will get lucky with a real arrival in the next instant. Because a longer duration increases the probability of a real arrival happening before $T$ and thereby decreases the probability of a fake, the principal can affect the probability that the final submission is real by changing the project duration. Thus, the principal faces a tradeoff: extending the duration increases the probability of a real submission and the principal’s payoff from accepting, but any surplus she obtains from an acceptance is discounted more heavily.

If the model with fixed duration does better than the baseline equilibrium, then it must be strictly optimal for the principal to accept any submission at $T$ with probability 1. Otherwise, the principal’s payoff is the same as always rejecting, which is worse than her payoff in the baseline model. Consequently, we focus on deriving equilibria in which the principal accepts for sure if she observes an arrival at the project’s chosen duration, i.e. $a(T) = 1$. In such an equilibrium, it is optimal for the ethical agent to submit his project at time $T$ if he has had an arrival and otherwise does nothing. Furthermore, it is optimal for the strategic agent to generate a fake, if he has not had a real arrival by time $T$. Thus, in such an equilibrium, the principal’s payoff of setting duration $T$ (and accepting an arrival) is

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

By implication, such an equilibrium exists if and only if $v(T) > 0$. A simple calculation shows that for such an equilibrium to exist, we must have a sufficiently long duration.
Lemma 8.1 (Long Duration.). In the fixed duration model, an equilibrium in which the principal approves any arrival at time $T$ exists if and only if $T > \overline{T}$, where

$$\exp\{-\lambda \overline{T}\} = \frac{1 - \theta}{1 - \theta + \theta \sigma}.$$ 

As the probability that the agent is strategic increases, the phase of doubt becomes unbounded, and the principal’s payoff in the baseline model approaches zero. Meanwhile, even when the probability of a strategic agent is high, with a long project duration, the principal can always ensure that an arrival is sufficiently likely to be real that she has a strict incentive to accept it. Furthermore, if she sets the optimal duration, $T^* > \overline{T}$, such that

$$\exp\{-\lambda T^*\} = \frac{\rho}{\rho + \lambda} \left(\frac{1 - \theta}{1 - \theta + \theta \sigma}\right),$$

then the principal always achieves a payoff that is bounded away from zero. Consequently, when she is likely to be interacting with a strategic agent, the principal can always increase her payoff over the baseline model by setting a project duration that is long enough to allow for real projects to arrive with high probability, but is not so long that all of the gains are dissipated through discounting.

Proposition 8.1 (Benefit from fixed duration.). When $\sigma$ is sufficiently high, there exists an interval of durations $[T_L, T_H]$ with $\overline{T} < T_L < T_H$, such that setting a project duration $T \in [T_L, T_H]$ delivers the principal a higher payoff than in the baseline model.

While doing so may be valuable for the principal, imposing a fixed project duration requires commitment power. With a fixed duration, a strategic agent has no incentive to generate a fake at $t < T$. Thus, from an attempt to submit a project before $T$, the principal can safely infer that the project is real, and she would like to deviate by immediately accepting it. Similarly, if no project is submitted by $T$, the principal must forego all future arrivals. But the absence of a submission at $T$ implies that the agent must be ethical, and therefore, the principal would like to accept any future submission. While it does require commitment, a fixed duration does not require that the principal has access to a public randomization device, which she would need in order to play her strategy in its most general form (i.e. a probability of acceptance as a function of time).

9 Conclusion

In this paper, we have investigated a dynamic model of fraud and trust building in which a principal with limited power of commitment faces an agent whose type — ethical or strategic
— is private information. An ethical agent only produces a real project, while a strategic agent chooses between producing a real project and a fake one. Producing a real project takes a positive and uncertain amount of time while a fake can be produced instantly.

In the unique equilibrium of the baseline model, the strategic agent randomizes about when to commit fraud and the principal randomizes about whether to accept a project as a function of its time of submission. As time passes without a project arrival, the principal’s belief that the agent is ethical rises until the point when she obtains full trust and accepts any subsequent submission with certainty. Indeed, the principal receives a positive expected payoff from a submission if and only if it arrives after she obtains full trust in the agent. This generates a relatively low present expected payoff for the principal in the equilibrium of the baseline model. We, therefore, explore several institutional remedies designed to improve the principal’s welfare, including: an auditing technology, opaque standards, commitment to suppress fraud, and requiring submissions to occur on a specified date.

We believe that the analysis in this paper can be extended fruitfully to study fraud and trust building in a number of related settings. For instance, it would be edifying to know how the potential for repeated interactions influences the agent’s equilibrium incentives to commit fraud and the principal’s incentives to approve his project submissions. Another intriguing possibility would be to investigate the impact on incentives of competition between several agents engaged in an innovation race. We plan to extend the analysis presented above to address such related questions in future work.
References


35
A Proofs

Calculations for Footnote 5. Suppose that the real technology imposes flow cost $c$ on

the agent and the agent pays this flow cost until either (i) he has a real arrival, at rate $\lambda$, (ii) the game exogenously ends, at rate $\rho$, (iii) he submits a fake. Equation (3) becomes,

$$u_s(t) = \int_0^t \lambda \exp(-\rho + \lambda) s (a(s) - cs) - \rho \exp(-\rho + \lambda) cs ds + \exp(-\rho + \lambda) t (a(t) - ct - \phi)$$

$$= u(t) - c \int_0^t (\lambda + \rho) \exp(-\rho + \lambda) s ds + \exp(-\rho + \lambda) t,$$

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

$$\int_0^t (\lambda + \rho) \exp(-\rho + \lambda) s ds + \exp(-\rho + \lambda) t = K - \frac{\exp(-\rho + \lambda) t}{\lambda + \rho},$$
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 $\phi$.

A.1 Proofs for Baseline Model

Proof of Lemma 4.1. To prove the claim, we need to derive the probability distributions for the overall time of submission and the time of submission for a real project. First, if the agent only uses the authentic technology, then the submission time for his real project is an exponentially distributed random variable $T_R \sim H(t) \equiv 1 - \exp(-\lambda t)$. Denote by $T_F \sim F(t)$ the waiting time for a fake project under the strategic agent’s strategy. Then the overall waiting time for a submission from the perspective of the principal is the random variable

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

That is, the arrival time for an ethical agent is always drawn from the authentic technology, while the arrival time of the strategic agent is the smaller of his cheating time and the authentic waiting time. Therefore, the cumulative distribution function of the project’s overall arrival time when the strategic agent uses strategy $F(\cdot)$ is

$$W(t) = (1 - \sigma)H(t) + \sigma(\lambda t + F(t) - F(t)H(t)) = H(t) + \sigma F(t)(1 - H(t))$$

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

with associated density

$$w(t) = \exp(-\lambda t)[\lambda(1 - \sigma F(t)) + \sigma f(t)].$$ 

By similar reasoning, when the strategic agent uses strategy $F(\cdot)$, the waiting time for a real arrival is distributed according to

$$W_A(t) = (1 - \sigma)H(t) + \sigma \int_0^t h(s)(1 - F(s))ds$$

$$= (1 - \sigma)(1 - \exp(-\lambda t)) + \sigma \int_0^t \lambda \exp(-\lambda s)(1 - F(s))ds.$$ 

If the agent is ethical, then a real project arrives before time $t$ with probability $H(t)$. If the agent is strategic, then a real project arrives before time $t$ if the waiting time under the authentic technology is some $s \leq t$, and simultaneously, if the cheating time exceeds this $s$. 

37
Integrating over \( s \leq t \) yields the preceding expression. The density of the waiting time for a real arrival is therefore,
\[
 w_A(t) = \lambda \exp(-\lambda t)(1 - \sigma F(t)).
\]
Then 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. \( \square \)

**Proof of Lemma 4.2.** First we verify that the prescribed behavior is an equilibrium when \( \phi > \bar{\phi} \). Suppose that \( a(t) = 1 \) for all \( t \geq 0 \). Then we have
\[
 u(t) = \frac{\lambda(1 - \exp(-(\rho + \lambda)t))}{\rho + \lambda} + \exp((\rho + \lambda)t)(1 - \phi),
\]
and
\[
 u'(t) = \exp(-(\rho + \lambda)t)(\phi(\rho + \lambda) - \rho) > 0 \iff \phi > \bar{\phi}.
\]
Thus, when \( \phi > \bar{\phi} \), the agent’s best response to \( a(t) = 1 \) for all \( t \geq 0 \) is never to submit a fake project. Moreover, if the agent never submits a fake, then \( g(t) = 1 \) for all \( t \geq 0 \), and the principal’s sequentially rational strategy is to approve any submission with probability 1.

Next we verify uniqueness. By way of contradiction, suppose that there is an equilibrium in which the agent submits a fake project with positive probability. Then he must do so using a density \( f(\cdot) \). To see this, note that the agent cannot submit a fake with positive probability at any distinct time \( t \) because this would yield \( g(t) = 0 \) and certain rejection by the principal. Let \( \bar{t} \leq \infty \) denote the upper bound of the support of \( F(\cdot) \). We consider two cases: \( \bar{t} < \infty \) and \( \bar{t} = \infty \).

Suppose \( \bar{t} < \infty \). Then for all \( t > \bar{t} \), \( \mu(t) = 0 \) which implies \( g(t) = 1 \) and by (2) \( a(t) = 1 \). Moreover \( a(\bar{t}) = 1 \) or else the agent could get a higher payoff than \( u(\bar{t}) \) by waiting to submit a fake until \( \bar{t} + \epsilon \). Therefore, the equilibrium payoff to the agent must satisfy
\[
 u(\bar{t}) = \int_0^{\bar{t}} \lambda \exp(-(\rho + \lambda)s)a(s) \, ds + \exp(-(\rho + \lambda)\bar{t})(1 - \phi).
\]
This must be at least as high as the payoff from never submitting a fake
\[
 u(\infty) = \int_0^{\bar{t}} \lambda \exp(-(\rho + \lambda)s)a(s) \, ds + \int_{\bar{t}}^{\infty} \lambda \exp(-(\rho + \lambda)s) \, ds
\]
Thus we have
\[ u(t) \geq u(\infty) \iff \exp(-(\rho + \lambda)t)(1 - \phi) \geq \exp(-(\rho + \lambda)\bar{t}) \frac{\lambda}{\rho + \lambda} \iff \hat{\phi} \geq \phi. \]

The last line contradicts the maintained hypothesis.

Next, suppose \( \bar{t} = \infty \). For this case we need to show \( a(t) \in (0, 1) \) almost everywhere in the support of \( f(\cdot) \). To see this, suppose to the contrary that there is a non-negligible set of points \( T \) such that for every \( t \in T \) \( a(t) = 1 \). Let \( t' \) be the smallest element of \( T \). Then for every other \( t \in T \)
\[ (1 - \phi)(\exp(-(\rho + \lambda)t') - \exp(-(\rho + \lambda)t)) + \int_{t'}^{t} \lambda \exp(-(\rho + \lambda)s)a(s)ds > 0 \implies u(t') > u(t). \]

But then the agent can profitably deviate by shifting the positive probability \( f(\cdot) \) places on \( T \) to \( t' \). Next suppose there exists a non-negligible set of points \( T \) such that \( a(t) = 0 \) for all \( t \in T \). Then for every \( t \in T \)
\[ \exp(-(\rho + \lambda)t)(0 - \phi) < \int_{t}^{\infty} \lambda \exp(-(\rho + \lambda)s)a(s)ds \implies u(t) < u(\infty). \]

But then the agent can profitably deviate by shifting the positive probability \( f(\cdot) \) places on \( T \) to \( \infty \) (i.e., never submitting a fake). Thus we have \( a(t) \in (0, 1) \) for almost every \( t \) in the support of \( f(\cdot) \). By (2) this requires \( \mu(t) = \lambda(1 - \theta)/\theta \) for all \( t \). By implication, \( f(t) = (1 - \sigma F(t))\lambda(1 - \theta)/\sigma \theta \geq (1 - \sigma)\lambda(1 - \theta)/\sigma \theta > 0 \). Thus the integral of \( f(\cdot) \) is unbounded, a contradiction.

\[ \square \]

**Proof of Lemma 4.3.** We prove several steps and then combine these to prove the claims in the lemma.

**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 \( \int_{t}^{\infty} \exp\{-(\rho + \lambda)s\}a(s)ds \geq 0 \), submitting a fake at time \( t \) is worse for the strategic agent than never submitting a fake,
\[ \int_{0}^{t} \exp\{-(\rho + \lambda)s\}a(s)ds + \exp\{-(\rho + \lambda)t\}(a(t) - \phi) < \int_{0}^{\infty} \exp\{-(\rho + \lambda)s\}a(s)ds. \]

Then, \( \mu(t) = 0 \) which implies \( g(t) = 1 \), and hence \( a(t) = 1 \) by (2), contradicting \( \phi < 1 \).

**Step 2.** We show \( f(t) = 0 \implies a(t) = 1 \). This follows from \( f(t) = 0 \implies \mu(t) = 0 \implies g(t) = 1 \implies a(t) = 1 \), by (2).
Step 3. We show that if \( a(t) = 1 \), then \( f(t') = 0 \) and \( a(t') = 1 \) for all \( t' > t \). Suppose \( a(t) = 1 \). Then for all \( t' > t \) we have,

\[
\begin{align*}
\text{Step 3.} & \quad u(t') - u(t) \\
& = \int_t^{t'} \lambda \exp(-(\rho + \lambda)s) a(s) \, ds + \exp(\rho - (\rho + \lambda)t') (a(t') - \phi) - \exp(\rho - (\rho + \lambda)t)(1 - \phi) \\
& \leq \int_t^{t'} \lambda \exp(-(\rho + \lambda)s) \, ds + \exp(\rho - (\rho + \lambda)t') (1 - \phi) - \exp(\rho - (\rho + \lambda)t)(1 - \phi) \\
& = \left( \phi - \frac{\rho}{\rho + \lambda} \right) \exp(\rho - (\rho + \lambda)t') - \exp(\rho - (\rho + \lambda)t') \quad < 0,
\end{align*}
\]

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

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

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

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

Step 6. We show that \( \bar{t} \in (0, \infty) \). first suppose \( \bar{t} = 0 \). Then \( a(t) = 1 \) for all \( t \geq 0 \) by Step 3. This yields

\[
\begin{align*}
\text{Step 6.} & \quad u(t) = \frac{\lambda(1 - \exp(-\rho \lambda t))}{\rho + \lambda} + \exp(\rho - (\rho + \lambda)t)(1 - \phi), \\
\quad u'(t) = (\phi - \hat{\phi})(\rho + \lambda) \exp(\rho - (\rho + \lambda)t).
\end{align*}
\]

This is negative because \( \phi < \hat{\phi} \). Hence If the principal approves all submissions, then it is optimal for the agent to submit a fake at \( t = 0 \), but then \( a(0) = 0 \) is sequentially rational for the principal, a contradiction.

Next, suppose that \( \bar{t} = \infty \), i.e. for all \( t \) we have \( f(t) > 0 \). From Step 4, we have \( a(t) \in (0, 1) \) for all \( t \). Hence, \( \mu(t) = \lambda(1 - \theta)/\theta \) for all \( t \). By implication, \( f(t) = (1 - \sigma F(t))\lambda(1 - \theta)/\sigma \theta \geq (1 - \sigma) \lambda(1 - \theta)/(\sigma \theta) > 0 \). Thus the integral of \( f(\cdot) \) is unbounded, a contradiction.

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

Proof of (ii). This follows from Steps 3 and 6.
Step 7. We show that \( a(\cdot) \) is continuous at \( \bar{t} \). Note that for \( t > \bar{t} \) we have \( \lim_{t \to \bar{t}} a(t) = a(\bar{t}) = 1 \). We seek to show that for \( t < \bar{t} \), we have \( \lim_{t \to \bar{t}} a(t) = 1 \). Consider \( t < \bar{t} \). Because \( f(t) > 0 \), we must have \( u(t) \geq u(\bar{t}) \). Hence,

\[
\begin{align*}
u(\bar{t}) - u(t) &= \\
&= \int_t^{\bar{t}} \lambda \exp(-(\rho + \lambda)s)a(s)\,ds + \exp(-(\rho + \lambda)\bar{t})(1 - \phi) - \exp(-(\rho + \lambda)t)(a(t) - \phi) \leq 0
\end{align*}
\]

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,

\[
0 = u(t') - u(t)
= \int_t^{t'} \lambda \exp(-(\rho + \lambda)s)a(s)\,ds + \exp(-(\rho + \lambda)t')(a(t') - \phi) - \exp(-(\rho + \lambda)t)(a(t) - \phi)
= \int_t^{t'} \lambda \exp(-(\rho + \lambda)s)a(s)\,dt + [\exp(-(\rho + \lambda)t') - \exp(-(\rho + \lambda)t)](a(t') - \phi)
+ \exp(-(\rho + \lambda)t)(a(t') - a(t)).
\]

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,

\[
\frac{\int_t^{t'} \lambda \exp(-(\rho + \lambda)s)a(s)\,ds}{t' - t} + \frac{\exp(-(\rho + \lambda)t') - \exp(-(\rho + \lambda)t)}{t' - t}(a(t') - \phi)
+ \exp(-(\rho + \lambda)t)\frac{a(t') - a(t)}{t' - t} = 0.
\]

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,

\[
a'(t) = (\rho + \lambda)(a(t) - \phi) - \lambda a(t) = (\rho + \lambda)[a(t) - \frac{\rho}{\rho + \lambda} - \phi] = (\rho + \lambda)[a(t)\hat{\phi} - \phi].
\]
It follows that \( a(t) \) does not change monotonicity for any \( t \in (0, \bar{t}) \). It is either constant, strictly decreasing, or strictly decreasing. Suppose \( a(t) \) is constant or strictly decreasing for all \( t < \bar{t} \). It follows that \( a(t) \leq \phi/\hat{\phi} \) for all \( t < \bar{t} \). Because \( \phi/\hat{\phi} < 1 \), we have \( \lim_{t \to \bar{t}} a(t) < 1 \), contradicting Step 7.

**Proof of (iii).** This follows from Steps 7 and 8. \qed

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

\[
\begin{align*}
  u'(t) &= \exp(-\rho t) \cdot \{\lambda a(t) - (\rho + \lambda)(a(t) - \phi) + a'(t)\} \\
  &= \exp(-\rho t) \cdot \{a'(t) - \rho a(t) + \phi(\rho + \lambda)\} = 0,
\end{align*}
\]

and hence, for \( t \in [0, \bar{t}] \)

\[
a(t) = \phi(1 + \frac{\lambda}{\rho}) + \kappa_1 \exp(\rho t)
\]

for some integration constant \( \kappa_1 > 0 \). Because \( a(t) \in (0, 1) \) for such \( t \in (0, \bar{t}) \), we also have

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

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

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

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

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

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

**Beliefs.** Obvious.

**Payoffs.**

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

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

**Ethical Agent.** Payoff of the ethical agent is

\[
U^E = \int_0^\infty \lambda \exp(-\rho t) a(t) \, dt.
\]

42
Note that
\[ U^E = \lim_{t \to \infty} u(t). \]
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
\[ u'(t) = - (\hat{\phi} - \phi)(\rho + \lambda) \exp(-(\rho + \lambda)t). \]
It follows that
\[ U^E = \lim_{t \to \infty} u(t) \]
\[ = u(\bar{t}) + \int_{\bar{t}}^{\infty} u'(t) \, dt \]
\[ = U^S - (\hat{\phi} - \phi) \exp(-(\rho + \lambda)\bar{t}). \]

**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,
\[ V = (1 - \theta)(1 - \sigma) \int_{\bar{t}}^{\infty} \lambda \exp\{-\lambda(\rho + \lambda)t\} \, dt \]
\[ = (1 - \theta)(1 - \sigma) \frac{\lambda}{\rho + \lambda} \exp\{-\lambda(\rho + \lambda)\bar{t}\} \]
\[ = (1 - \theta) \frac{\lambda}{\rho + \lambda} \exp\{\left((1 + \rho)(\frac{\theta}{\lambda(1 - \theta)}) \ln(1 - \sigma)\right)\} \]
\[ = (1 - \theta) \frac{\lambda}{\rho + \lambda} \exp\{-\left(\frac{\lambda}{\theta} + \rho\right)(\frac{\theta}{\lambda(1 - \theta)}) \ln(1 - \sigma)\} \]
\[ = (1 - \theta) \frac{\lambda}{\rho + \lambda} \exp\{-\left(\frac{\lambda}{\theta} + \rho\right)\bar{t}\} \]
\[ = (1 - \theta) \frac{\lambda}{\rho + \lambda} (1 - \sigma)^{\frac{\rho + \lambda}{\lambda(1 - \theta)}}. \]

**Proof of Corollary 4.1.** We prove each part in turn. **Part (i).** Observe that
\[ \frac{\partial F(t)}{\partial \lambda} = \frac{t(1 - \theta)}{\theta \sigma} \exp\{-\lambda \frac{1 - \theta}{\theta} t\} > 0. \]
Hence, an increase in \( \lambda \) generates a first order dominance shift down in the cheating time distribution.
We show that an increase in $\lambda$ increases the acceptance probability at all $t < \bar{t}$. We find that

$$\frac{\partial a(t)}{\partial \lambda} = \frac{\phi}{\rho} (1 - \exp\{-\rho(\bar{t} - t)\}) + (1 - \frac{\phi}{\bar{\phi}}) \frac{d}{d\lambda} (\exp\{-\rho(\bar{t} - t)\})$$

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

Note that $\partial \bar{t}/\partial \lambda = -\bar{t}/\lambda$, and hence

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

Each term in the sum is positive for $t < \bar{t}$ and $\phi < \bar{\phi}$.

That the strategic agent’s payoff increases in $\lambda$ is an immediate implication of the preceding observation.

For the ethical agent we have

$$\frac{\partial U^E}{\partial \lambda} = \frac{\partial U^S}{\partial \lambda} + \bar{t}(\bar{\phi} - \phi) \left( \frac{\rho + 2\lambda}{\lambda} \right) \exp(-(\rho + \lambda)\bar{t}),$$

which is evidently positive.

To see that the principal’s payoff is increasing in $\lambda$, note that $\lambda/(\rho + \lambda)$ is increasing in $\lambda$ and that

$$\frac{\partial}{\partial \lambda} \exp\{-\left(\frac{\lambda}{\theta} + \rho\right)\bar{t}\} = \left(-\frac{\bar{t}}{\theta} - \left(\frac{\lambda}{\theta} + \rho\right) \frac{d\bar{t}}{d\lambda}\right) \exp\{-\left(\frac{\lambda}{\theta} + \rho\right)\bar{t}\} = \frac{\bar{t} \lambda}{\rho} \exp\{-\left(\frac{\lambda}{\theta} + \rho\right)\bar{t}\} > 0,$$

where the last equality uses the observation that $\partial \bar{t}/\partial \lambda = -\bar{t}/\lambda$.

**Part (ii).** That $a(t)$ increases with $\phi$ is obvious. For the strategic agent’s payoff, note that

$$\frac{\partial U^S}{\partial \phi} = \frac{\lambda}{\rho} - (1 + \frac{\lambda}{\rho})(1 - \sigma) \frac{\lambda^\phi}{\lambda(1-\theta)}.$$ 

Clearly, $\partial U^S/\partial \phi$ is increasing in $\sigma$, strictly negative at $\sigma = 0$ and strictly positive at $\sigma = 1$ which proves this part of the claim.

For the ethical agent we have

$$\frac{\partial U^E}{\partial \phi} = \frac{\lambda}{\rho} - (1 + \frac{\lambda}{\rho})(1 - \sigma) \frac{\lambda^\phi}{\lambda(1-\theta)} + \ln(1 - \sigma) \frac{\lambda^\phi}{\lambda(1-\theta)}.$$ 

This is 0 at $\sigma = 0$ and it is straightforward to verify that $\frac{\partial^2 U^E}{\partial \phi^2} > 0$, which establishes the claim for the ethical agent’s payoff. Finally, $v$ does not depend directly or indirectly on $\phi$.

**Part (iii).** We have

$$\frac{\partial F(t)}{\partial \theta} = -\frac{\lambda t}{\theta^2 \sigma} \exp\left(-\frac{(1 - \theta)}{\theta} t\right).$$
This is strictly negative for $t > 0$. Similarly

$$\frac{\partial \bar{t}}{\partial \theta} = -\frac{1}{\lambda \theta^2} \ln(1 - \sigma) > 0.$$ 

Also, we have

$$\frac{\partial a(t)}{\partial \theta} = -\lambda \frac{\partial \bar{t}}{\partial \theta} \left(1 - \frac{\phi}{\hat{\phi}}\right) \exp(-\lambda(\bar{t} - t)) < 0.$$ 

Moreover, it follows that

$$\frac{\partial U^S}{\partial \theta} = \frac{\partial a(0)}{\partial \theta} < 0.$$ 

For the ethical agent we have

$$\frac{\partial U^E}{\partial \theta} = \frac{\partial}{\partial \theta} \left(a(0) - \phi - (\hat{\phi} - \phi) \exp(-(\rho + \lambda)\bar{t})\right) = \frac{\partial \bar{t}}{\partial \theta} (\hat{\phi} - \phi) \left(\exp(-(\rho + \lambda)\bar{t}) - \exp(-\rho \bar{t})\right).$$ 

The first two terms are positive and the third is negative. That the principal’s payoff is decreasing is an immediate implication of $d\bar{t}/d\theta > 0$.

**Part (iv).** A change in $\rho$ has no effect on the agent’s mixed strategy or on the duration of the phase of doubt. Because arrivals during the phase of credibility are discounted more heavily, an increase in $\rho$ clearly reduces the principal’s welfare.

Furthermore, we have

$$a(t) = \phi(1 + \frac{\lambda}{\rho}) + (1 - \phi - \frac{\phi \lambda}{\rho}) \exp(-\rho(\bar{t} - t)) \Rightarrow$$

$$\frac{da(t)}{d\rho} = -\frac{\phi \lambda}{\rho^2} + \frac{\phi \lambda}{\rho^2} \exp(-\rho(\bar{t} - t)) - (\bar{t} - t)(1 - \frac{\phi \lambda}{\rho}) \exp(-\rho(\bar{t} - t)))$$

$$= -\frac{\phi \lambda}{\rho} [1 - \exp(-\rho(\bar{t} - t))] - (\bar{t} - t)(1 - \frac{\phi \lambda}{\rho}) \exp(-\rho(\bar{t} - t)) < 0.$$ 

Finally, because an increase in $\rho$ reduces $a(0)$, the strategic agent is worse off. Furthermore, the ethical agent is hurt, both because his payoff from every arrival is discounted more heavily, and because $a(t)$ is reduced.

\[\square\]

### A.2 Proofs for Auditing

**Proof of Lemma 5.2.** Consider a possible equilibrium in which the principal never audits an arrival. In this case, the equilibrium strategies and beliefs must be the ones characterized in Proposition 4.1. It is never sequentially rational for the principal to test during the phase of credibility, since $g(t) = 1$. Thus, an equilibrium with no testing exists if and only if it is
sequentially rational for the principal not to test within the phase of doubt. For any arrival within the phase of doubt, the principal’s belief is \( g(t) = \theta \), and thus, the payoff of testing is 
\[
\theta(1 - \theta) - \theta(1 - \theta)(1 - \alpha) - k = \alpha \theta(1 - \theta) - k,
\]
and the payoff of either accepting or rejecting is zero. It is sequentially rational for the principal not to test during the phase of doubt if and only if \( k \geq \alpha \theta(1 - \theta) \).

**Proof of Lemma 5.3.** 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,
\[
u'_A(t) = \exp\{- (\rho + \lambda)t\}(\rho + \lambda)(\alpha - (\hat{\phi} - \phi)).
\]
Thus, for \( \alpha \neq \hat{\phi} - \phi \), 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 5.1.** 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
\[
u_A(t) = \int_0^t \exp\{- (\rho + \lambda)s\} \lambda a_R(s) ds + \exp\{- (\rho + \lambda)t\}(a_F(t) - \phi).
\]

**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, \( \mu(t) = 0 \) which implies \( g(t) = 1 \), and hence \( a_F(t) = 1 \) by (10), contradicting \( \phi < 1 \).

**Step 2.** We show \( f(t) = 0 \implies a(t) = 1 \). This follows from \( f(t) = 0 \implies \mu(t) = 0 \implies g(t) = 1 \implies a(t) = 1 \), by (10).
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,

\[
u_A(t') - u_A(t)
= \int_t^{t'} \lambda \exp(-(\rho + \lambda)s)a_R(s) \, ds + \exp(-(\rho + \lambda)t')(a_F(t') - \phi) - \exp(-(\rho + \lambda)t)(1 - \phi)
\leq \int_t^{t'} \lambda \exp(-(\rho + \lambda)s) \, ds + \exp(-(\rho + \lambda)t')(1 - \phi) - \exp(-(\rho + \lambda)t)(1 - \phi)
= \frac{\lambda}{\rho + \lambda} \left( \exp(-(\rho + \lambda)t) - \exp(-(\rho + \lambda)t') \right) - (1 - \phi) \left( \exp(-(\rho + \lambda)t) - \exp(-(\rho + \lambda)t') \right)
= \left( \phi - \frac{\rho}{\rho + \lambda} \right) \left( \exp(-(\rho + \lambda)t) - \exp(-(\rho + \lambda)t') \right) < 0,
\]

where the last inequality follows because \( \phi < \hat{\phi} \) 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 (10) \( a(t') = 1 \).

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 (10), we get \( g(t) \leq \theta_1 \), and hence, \( \mu(t) \geq \lambda(1 - \theta_1)/\theta_1 > 0 \). 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 \( \mu(t) = \infty \), 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, \( \phi < \hat{\phi} \) 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 (10), we have \( g(t) \leq \theta_1 \) for all \( t \). Hence, \( \mu(t) \geq \lambda(1 - \theta_1)/\theta_1 \) 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,

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

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,

\[
0 = u_A(t') - u_A(t)
= \int_t^{t'} \lambda \exp(-(\rho + \lambda)s)a_R(s) \, ds + \exp(-(\rho + \lambda)t')(a_F(t') - \phi) - \exp(-(\rho + \lambda)t)(a_F(t) - \phi)
= \int_t^{t'} \lambda \exp(-(\rho + \lambda)s)a_R(s) \, ds + [\exp(-(\rho + \lambda)t') - \exp(-(\rho + \lambda)t)](a_F(t') - \phi)
+ \exp(-(\rho + \lambda)t)(a_F(t') - a_F(t)).
\]

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 (10). Suppose \( g(t) = \theta_1 \). From (10), 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 (10), 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 (10), 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

\[
0 = u_A(t') - u_A(t)
\]

\[
= \int_t^{t'} \lambda \exp(-\rho s) \, ds + \exp(-\rho \lambda t')(a_F(t') - \phi) - \exp(-\rho \lambda t)(a_F(t) - \phi)
\]

\[
= \int_t^{t'} \lambda \exp(-\rho s) \, dt + [\exp(-\rho \lambda t') - \exp(-\rho \lambda t)](a_F(t') - \phi)
\]

\[
= \exp(-\rho \lambda t')(a_F(t') - \phi) - \exp(-\rho \lambda t)(a_F(t') - a_F(t)).
\]

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,

\[
a_F'(t') = (\rho + \lambda)(a_F(t') - \phi) - \lambda
\]

\[
\iff
\]

\[
a_F'(t') = (\rho + \lambda)[a_F(t') - \phi - (1 - \hat{\phi})]
\]

\[
\iff
\]

\[
a_F'(t') = (\rho + \lambda)[a_F(t') - (1 - \hat{\phi} + \phi)].
\]

Thus, whenever \( a_F(t) > 1 - \hat{\phi} - \phi \), we have \( a_F'(t) > 0 \), and hence, \( a_F(\cdot) \) is increasing at \( t \). Conversely, whenever \( a_F(t) < 1 - \hat{\phi} - \phi \), we have \( a_F'(t) < 0 \), and hence, \( a_F(\cdot) \) is decreasing at \( t \). Furthermore, whenever \( a_F(t) = 1 - \hat{\phi} - \phi \), 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 an \( t \in [0, \bar{t}_A] \).

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

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

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

Step 17. We show that \( p(t) \) is differentiable and decreasing for \( t \in (\tilde{t}_A, \bar{t}_A) \) 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 (10), 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 \( t \in (\tilde{t}_A, \bar{t}_A) \).

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

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

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

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

**Proof of Proposition 5.1.** Consider a single auditing stage in which the principal randomizes between performing the test and default approval, i.e. the scenario in which \( \tilde{t}_A = 0 \) in Lemma 5.1.

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

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

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

\[
\tilde{t}_A = -\frac{\ln(1 - \sigma)}{\mu_1} = \tilde{t}_1.
\]

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

\[
u_A(t) = \int_0^t \lambda \exp(-(\rho + \lambda) s) \, ds + \exp(-(\rho + \lambda)t)(1 - \alpha p(t) - \phi).
\]
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
\[ p(t) = \frac{\hat{\phi} - \phi}{\alpha} + \kappa_3 \exp((\rho + \lambda) t). \]
Boundary condition, \( p(T_1) = 0 \) gives
\[ p(t) = \frac{\hat{\phi} - \phi}{\alpha} \left(1 - \exp(-(\rho + \lambda) (T_1 - t))\right). \]
This is a decreasing function as required. Moreover \( p(0) \leq 1 \) if \( \alpha > \hat{\phi} - \phi \), or if \( \alpha < \hat{\phi} - \phi \) and \( T_1 \leq \delta_A \), as stipulated.

Payoffs. The strategic agent is indifferent about submitting a fake project at all times \( t \in [0, T_1] \), 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
\[ V_A = \int_0^{T_1} \exp(- (\rho + \lambda) t) (\lambda (1 - \sigma F(t)) (1 - \theta) - \sigma f(t) \theta) \, dt + (1 - \sigma) (1 - \theta) \int_{T_1}^{\infty} \lambda \exp(-(\rho + \lambda) t) \, dt. \]
To see this, note first that for an arrival during \([0, T_1]\), 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 \( t \in [0, T_1] \) 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 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
\[ V_A = \int_0^{T_1} \exp(- (\rho + \lambda + \mu_1) t) (\lambda (1 - \theta) - \theta \mu_1) \, dt + (1 - \sigma) (1 - \theta) \int_{T_1}^{\infty} \lambda \exp(-(\rho + \lambda) t) \, dt \]
\[ = \int_0^{T_1} \exp(- (\rho + \frac{\lambda}{\theta_1}) t) (\lambda (1 - \theta) - \theta \lambda \frac{1 - \theta_1}{\theta_1}) \, dt + (1 - \sigma) (1 - \theta) \int_{T_1}^{\infty} \lambda \exp(-(\rho + \lambda) t) \, dt \]
\[ = \lambda (1 - \theta) \frac{\theta}{\theta_1} \int_0^{T_1} \exp(- (\rho + \frac{\lambda}{\theta_1}) t) \, dt + (1 - \sigma) (1 - \theta) \int_{T_1}^{\infty} \lambda \exp(-(\rho + \lambda) t) \, dt. \]
Proof of Proposition 5.2. Consider an equilibrium in which the principal first randomizes between performing the test and default rejection and switches at time $\hat{t}_A$ to randomizing between performing the test and default approval, as in Lemma 5.1 with $\tilde{t}_A > 0$. Expressions (A.1) and (A.2) 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, $\hat{t}_A, \tilde{t}_A$, combined with the boundary conditions, $p(\hat{t}_A) = 1$ and $p(\tilde{t}_A) = 0$, established in Lemma 5.1.

Strategies and Phase Transitions. For the principal to be indifferent between auditing and rejecting we must have $g(t) = \theta_0$ for all $t \in [0, \hat{t}_A)$. The agent therefore mixes with constant arrival rate $\mu_0$. Furthermore, since the equilibrium begins with this phase, we cannot have a mass point at zero. It follows that for $t \in [0, \hat{t}_A)$,

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

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

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

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

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

Solving yields

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

To construct the equilibrium we must paste the two phases together ensuring continuity of the CDF, and the boundary conditions $p(\hat{t}_A) = 0$ and $p(\tilde{t}_A) = 1$. We have the following five equations in the five unknowns $\kappa_1, \kappa_2, \kappa_3, \hat{t}$ and $\tilde{t}_A$:

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

$$\Rightarrow \quad \frac{1}{\sigma}(1 - \kappa_1 \exp(-\mu_1 \tilde{t}_A)) = 1,$$

(A.3)
continuity of $F(\cdot)$ at $\tilde{t}_A$

$$\implies \frac{1}{\sigma}(1 - \exp(-\mu_0 \tilde{t}_A)) = \frac{1}{\sigma}(1 - \kappa_1 \exp(-\mu_1 \tilde{t}_A)),$$  \hspace{1em} (A.4)

$$p(\tilde{t}_A) = 1$$

$$\implies \frac{\phi}{\phi - \alpha} + \kappa_2 \exp((\rho + \lambda)(\tilde{\phi} - \alpha)\tilde{t}_A) = 1,$$  \hspace{1em} (A.5)

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

$$\implies \frac{\hat{\phi} - \phi}{\alpha} + \kappa_3 \exp((\rho + \lambda)\tilde{t}_A) = 1,$$  \hspace{1em} (A.6)

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

$$\implies \frac{\hat{\phi} - \phi}{\alpha} + \kappa_3 \exp((\rho + \lambda)\tilde{t}_A) = 0.$$  \hspace{1em} (A.7)

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.7) we get

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

Substituting this into (A.6) gives

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

$$\implies \exp(-(\rho + \lambda)(\tilde{t}_A - \tilde{t}_A)) = 1 - \frac{\alpha}{\hat{\phi} - \phi}$$

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

$$\tilde{t}_A - \tilde{t}_A = -\frac{\ln(1 - \frac{\alpha}{\hat{\phi} - \phi})}{\rho + \lambda} = \delta_A.$$  \hspace{1em} (A.8)

We will return to this equation later. From (A.5) we get

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

Observe that $\alpha < \hat{\phi} - \phi$ implies $\kappa_2 > 0$. From (A.4) we get

$$\kappa_1 = \exp(-(\mu_0 - \mu_1)\tilde{t}_A).$$
Note that $\kappa_1 > 0$. Furthermore, $\mu_0 > \mu_1$ implies $\kappa_1 < 1$ whenever $\tilde{t}_A > 0$. Substituting $\kappa_1$ into (A.3), rearranging, and taking logs gives

\[(\mu_0 - \mu_1)\tilde{t}_A + \mu_1 \tilde{t}_A = -\ln(1 - \sigma).\]  

(A.9)

Solving (A.8) and (A.9) for $\tilde{t}_A$ and $\tilde{t}_A$ yields the desired expressions.

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

\[\tilde{t}_A > 0 \iff \tilde{t}_0 - \frac{\mu_1}{\mu_0} \delta_A > 0 \iff \frac{\mu_1}{\mu_0}(\tilde{t}_1 - \delta_A) > 0 \iff \tilde{t}_1 > \delta_A,
\]

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

We show that $\tilde{t}_A > 0 \implies \tilde{t}_A < \tilde{t}_1$. Note that

\[-\ln(1 - \sigma) \frac{\mu_0 - \mu_1}{\mu_1 \mu_0} > (1 - \frac{\mu_1}{\mu_0})\delta_A \implies -\ln(1 - \sigma)\left(1 - \frac{1}{\mu_1 \mu_0}\right) > (1 - \frac{\mu_1}{\mu_0})\delta_A \implies \tilde{t}_1 - \tilde{t}_0 > (1 - \frac{\mu_1}{\mu_0})\delta_A \implies \tilde{t}_1 > \tilde{t}_0 + (1 - \frac{\mu_1}{\mu_0})\delta_A = \tilde{t}_A.
\]

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

Following similar logic to Proposition 5.1, given the agent’s mixed strategy $F(\cdot)$, the principal’s payoff is

\[V_A = \int_{\tilde{t}_A}^{\tilde{t}_A} \exp(-(\rho + \lambda)t)(\lambda(1 - \sigma F(t))(1 - \theta - \sigma f(t)\theta) dt + (1 - \sigma)(1 - \theta) \int_{\tilde{t}_A}^{\infty} \lambda \exp(-(\rho + \lambda)t) dt.
\]

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

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

which implies that the principal’s payoff is

\[\exp(-(\mu_0 - \mu_1)\tilde{t}_A) \int_{\tilde{t}_A}^{\tilde{t}_A} \exp(-(\rho + \lambda + \mu_1)t)(\lambda(1 - \theta - \theta \mu_1) dt + (1 - \theta)(1 - \sigma) \int_{\tilde{t}_A}^{\infty} \lambda \exp(-(\rho + \lambda)t) dt.
\]

Substituting $\mu_1$ inside the integrand and simplifying yields

\[V_A = \exp(-(\mu_0 - \mu_1)\tilde{t}_A)(1 - \frac{\theta}{\theta_1}) \int_{\tilde{t}_A}^{\tilde{t}_A} \exp(-(\rho + \lambda)\frac{t}{\theta_1}) dt + (1 - \sigma)(1 - \theta) \int_{\tilde{t}_A}^{\infty} \lambda \exp(-(\rho + \lambda)\frac{t}{\theta_1}) dt
\]

\[\square\]
Proof of Proposition 5.3. We consider the principal’s, ethical agent’s, and strategic agent’s payoffs in the one phase and two stage auditing equilibrium.

**Principal, One Phase Equilibrium.** From Proposition 5.1, we have that the principal’s payoff in the one stage auditing equilibrium is

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

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 \( \theta_1 \), we have

\[ \frac{dV_A}{d\theta_1} = \lambda \theta \int_0^{\bar{T}_1} \exp(-\rho + \frac{\lambda}{\theta_1}) t \, dt \\
+ \lambda (1 - \frac{\theta}{\theta_1}) \left[ \exp(-\rho + \frac{\lambda}{\theta_1}) \bar{T}_1 \frac{d\bar{T}_1}{d\theta_1} + \int_{\theta_1}^{\bar{T}_1} \exp(-\rho + \frac{\lambda}{\theta_1}) t \frac{\lambda}{\theta_1} dt \right] - (1 - \sigma)(1 - \theta) \lambda \exp(-\rho + \lambda) \bar{T}_1 \frac{d\bar{T}_1}{d\theta_1}. \]

Note that

\[(1 - \sigma) \exp(-\rho + \frac{\lambda}{\theta_1}) \bar{T}_1 = \exp(1 - (\rho + \lambda) \frac{\theta_1}{\lambda(1 - \theta_1)} t (1 - \sigma)) \]

\[= \exp(-\rho + \frac{\lambda}{\theta_1})(\frac{\theta_1}{\lambda(1 - \theta_1)}) t (1 - \sigma)) = \exp(-\rho + \frac{\lambda}{\theta_1} \bar{T}_1), \]

and

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

Substituting and simplifying, we have

\[ \frac{dV_A}{d\theta_1} = \lambda \theta \int_0^{\bar{T}_1} \exp(-\rho + \frac{\lambda}{\theta_1}) t \, dt - \lambda \theta \bar{T}_1 \frac{\lambda}{\theta_1} + \lambda \frac{(1 - \theta_1)}{\theta_1} \int_{\theta_1}^{\bar{T}_1} \exp(-\rho + \frac{\lambda}{\theta_1}) t \frac{\lambda}{\theta_1} dt. \]

Noting that the last integral is strictly positive, we have

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

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 \( \theta_1 \), (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 there exists $\tilde{\sigma}$ such that the two stage auditing equilibrium obtains iff $\sigma > \tilde{\sigma}$ and that the principal’s payoff in the two equilibrium approaches her payoff in the 1-phase equilibrium as $\sigma \downarrow \tilde{\sigma}$. 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 $\sigma \in (\tilde{\sigma}, \tilde{\sigma} + \epsilon)$.

In Step 2 we show that if the principal’s payoff is higher in the 2-stage auditing equilibrium than in the baseline model for any value of $\sigma$, 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 $\sigma$ such that the two stage auditing equilibrium exists, there exists $\sigma' < \sigma$ such that (i) the two stage auditing equilibrium exists at $\sigma'$, and (ii) at $\sigma'$ 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 5.2, we have $\alpha < \hat{\phi} - \phi$ and $\tilde{t}_1 > \delta_A$. Note that

$$\tilde{t}_1 > \delta_A \iff \ln(1 - \sigma) > \mu_1 \delta \iff \sigma > \tilde{\sigma} \equiv 1 - \exp(-\mu_1 \delta_A).$$

By implication, if $\sigma > \tilde{\sigma}$, then the two stage auditing equilibrium exists for all $\sigma \in (\tilde{\sigma}, \sigma)$.

Next, we argue that as $\sigma \downarrow \tilde{\sigma}$, 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

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

As $\sigma \downarrow \tilde{\sigma}$, we have $\tilde{t}_1 \downarrow \delta_A$. By implication,

$$\sigma \downarrow \tilde{\sigma} \Rightarrow \tilde{t}_A = \tilde{t}_0 - \frac{\mu_1}{\mu_0} \delta_A = \frac{\mu_1}{\mu_0} (\tilde{t}_1 - \delta_A) \downarrow 0.$$ 

By routine simplification, the $\sigma \downarrow \tilde{\sigma}$, 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 phase equilibrium that obtains at $\tilde{\sigma}$. 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 $\sigma' \in (\tilde{\sigma}, \tilde{\sigma} + \epsilon)$, the principal’s payoff in the
two stage equilibrium at $\sigma'$ strictly exceeds her payoff in the baseline model. Hence, for any $\sigma > \tilde{\sigma}$, there exists $\sigma' \in (\tilde{\sigma}, \sigma)$ at which the principal’s payoff in the two stage equilibrium is higher than her payoff in the baseline model.

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

Consider the payoff difference between the auditing equilibrium and the baseline model,

$$
\exp(-(\mu_0 - \mu_1)\tilde{t}_A)(1 - \frac{\theta}{\theta_1})\int_{\tilde{t}_A}^{T_A} \exp(-(\rho + \frac{\lambda}{\theta_1})t) dt
+ (1 - \theta)(1 - \sigma)\left[\int_{\tilde{t}_A}^{\infty} \lambda \exp(-(\rho + \lambda)t) dt - \int_{\tilde{t}_A}^{\infty} \lambda \exp(-(\rho + \lambda)t) dt \right].
$$

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

We simplify the first line.

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

Note that

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

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

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

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

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

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

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

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

$$
\kappa_1(1 - \sigma)^{(\rho + \frac{\lambda}{\theta_0})\frac{\mu_1}{\mu_0}},
$$

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

\[
(1 - \theta)(1 - \sigma) \int_{\bar{t}A}^{\infty} \lambda \exp(-(\rho + \lambda)t) \, dt - \int_{\bar{t}}^{\infty} \lambda \exp(-(\rho + \lambda)t) \, dt =
\]

\[
(1 - \theta)(1 - \sigma) \int_{\bar{t}A}^{\infty} \lambda \exp(-(\rho + \lambda)t) \, dt = (1 - \theta)(1 - \sigma) \frac{\lambda}{\lambda + \rho} [\exp(-(\rho + \lambda)\bar{t}_A) - \exp(-(\rho + \lambda)\bar{t})].
\]

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

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

Note that

\[
(1 - \sigma) \exp(-(\rho + \lambda)\bar{t}_0) = \exp(1 - (\rho + \lambda)\frac{\theta_0}{\lambda(1 - \theta_0)}) \ln(1 - \sigma)) = \exp(-\rho + \frac{\lambda}{\theta_0} \theta_0)(-\frac{\theta_0}{\lambda(1 - \theta_0)}) \ln(1 - \sigma)) = \exp(-\rho + \frac{\lambda}{\theta_0} \bar{t}_0),
\]

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

\[
(1 - \theta) \frac{\lambda}{\lambda + \rho} [\exp(-(\rho + \lambda)(1 - \frac{\mu_1}{\mu_0})\delta) \exp(-(\rho + \frac{\lambda}{\theta_0} \theta_0)\bar{t}_0) - \exp(-(\rho + \frac{\lambda}{\bar{t}} \bar{t})]] =
\]

\[
(1 - \theta) \frac{\lambda}{\lambda + \rho} [\exp(-(\rho + \lambda)(1 - \frac{\mu_1}{\mu_0})\delta)(1 - \sigma) \frac{\rho \theta_0 + \lambda}{\lambda(1 - \theta_0)} - (1 - \sigma) \frac{\rho \theta_0 + \lambda}{\lambda(1 - \theta_0)}] =
\]

\[
\kappa_2(1 - \sigma) \frac{\rho \theta_0 + \lambda}{\lambda(1 - \theta_0)} - \kappa_3(1 - \sigma) \frac{\rho \theta_0 + \lambda}{\lambda(1 - \theta_0)},
\]

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

\[
(\kappa_1 + \kappa_2)(1 - \sigma) \frac{\rho \theta_0 + \lambda}{\lambda(1 - \theta_0)} - \kappa_3(1 - \sigma) \frac{\rho \theta_0 + \lambda}{\lambda(1 - \theta_0)}.
\]

Therefore,

\[
(\kappa_1 + \kappa_2)(1 - \sigma) \frac{\rho \theta_0 + \lambda}{\lambda(1 - \theta_0)} - \kappa_3(1 - \sigma) \frac{\rho \theta_0 + \lambda}{\lambda(1 - \theta_0)} > 0 \iff
\]

\[
(\kappa_1 + \kappa_2) - \kappa_3(1 - \sigma) \frac{\rho \theta_0 + \lambda}{\lambda(1 - \theta_0)} > 0 \iff
\]

\[
\frac{\kappa_1 + \kappa_2}{\kappa_3} > (1 - \sigma) \frac{(\rho \theta_0 + \lambda)(\theta_0 - \theta)}{\lambda(1 - \theta)(1 - \theta_0)}.
\]

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

**Step 3.** We show that the principal’s payoff is higher in the two stage auditing equilibrium than in the baseline model. Consider \( \sigma > \bar{\sigma} \). From Step 1, there exists \( \sigma' \in (\bar{\sigma}, \sigma) \) such
that the principal’s payoff in the two stage auditing equilibrium at $\sigma’$ 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 Phase Equilibrium.** The ethical agent never submits a fake project and all real projects are approved in the one phase equilibrium. Therefore he receives his first-best payoff

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

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

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

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

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

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

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 $t \leq \tilde{t}_A$:

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

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

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

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

Combining this with the last line above gives

$$\frac{\phi}{\phi - \alpha} + \left( 1 - \frac{\phi}{\phi - \alpha} \right) \exp \left( -\frac{\phi - \alpha}{1 - \alpha} (\rho + \lambda) (\tilde{t}_A - t) \right) > \frac{\phi}{\phi} + \left( 1 - \frac{\phi}{\phi} \right) \exp (-\rho (\tilde{t} - t)),
$$
or $p(t) > a(t)$ as desired.

**Strategic Agent, One Phase Equilibrium.** The one phase equilibrium exists when $\alpha > \hat{\phi} - \phi$, for all $\sigma \in [0, 1]$. In the one phase equilibrium, the strategic agent’s payoff is $1 - \alpha p(0) - \phi$. As $\sigma \to 1$ both $\tilde{t}_A$ and $\tilde{t}$ diverge so that

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

and

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

Note that

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

**Strategic Agent, Two Stage Equilibrium.** Note that for $\alpha < \hat{\phi} - \phi$ and $\sigma > \tilde{\sigma}$, the auditing equilibrium is in two phases. Since we assume that $\sigma$ is large, the strategic agent’s payoff is $p(0) - \phi$. As $\sigma \to 1$ both $\tilde{t}_A$ and $\bar{t}_A$ diverge so that

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

and

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

Thus, the two stage auditing equilibrium payoff is higher whenever

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

and, thus, $\hat{\phi} < 1$ implies that the first limit is larger than the second. \qed
A.3 Proofs for Opaque Standards

Proof of Lemma 6.1. Proof of (i). This point follows exactly from the arguments in the proof of Lemma 4.3, replacing \(a(\cdot)\) with \(a_U(\cdot)\).

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

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

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

Step 1. We show that for any \(t \geq 0\),

(a) if \(a_H(t) > 0\), then \(a_L(t) = 1\)

(b) if \(a_L(t) < 1\), then \(a_H(t) = 0\).

(a) and (b) follow immediately from each type of principal’s sequentially rational acceptance strategy, coupled with \(\theta_H > \theta_L\).

Step 2. We show that for any \(t \in [0, \bar{t}_U)\), exactly one of the following conditions must hold

(A) \(a_H(t) \in (0, 1)\) and \(a_L(t) = 1\)

(B) \(a_L(t) \in (0, 1)\) and \(a_H(t) = 0\)

(C) \(a_L(t) = 1\) and \(a_H(t) = 0\).

From the definition of \(\bar{t}_U\), we know that \(a_U(t) < 1\) for \(t < \bar{t}_U\). Hence, for such \(t\), at least one of \(a_i(t) < 1\) for \(i \in \{H, L\}\).

If \(a_L(t) < 1\), then \(a_H(t) = 0\) from Step 1 (b). Furthermore, from Step 0, we know that \(a_U(t) > \phi\). Coupled with \(a_H(t) = 0\), this implies \(a_L(t) > 0\). Hence, we have (B).
If $a_H(t) < 1$, then there are two possibilities. If $a_H(t) > 0$, then from Step 1 (a), we have $a_L(t) = 1$, case (A). If $a_H(t) = 0$, then we must have $a_L(t) > 0$ (otherwise $a_U(t) = 0$, contradicting Step 0). If $a_L(t) < 1$, then case (B). If $a_L(t) = 1$ then case (C).

**Step 3.** Consider $0 \leq t < t' < \tilde{t}_U$. We show that

(a) If $a_H(t) \in (0, 1)$ then $a_H(t) < a_H(t') < 1$ and $a_L(t') = 1$.

(b) If $a_L(t') \in (0, 1)$ then $a_L(t) < a_L(t') < 1$ and $a_H(t) = 0$.

From Step 0, we have $a_U(t') > a_U(t)$. Thus,

$$\nu a_H(t') + (1 - \nu)a_L(t') > \nu a_H(t) + (1 - \nu)a_L(t). \tag{A.10}$$

To prove claim (a), suppose that $a_H(t) \in (0, 1)$. By Step 1 (a), we have $a_L(t) = 1$. Hence, (A.10) implies

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

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

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

To prove claim (b), suppose that $a_L(t') \in (0, 1)$. By Step 1 (b) we have $a_H(t') = 0$. Hence, (A.10) implies

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Because \( a_L(\cdot) \leq 1 \) and \( a_H(\cdot) \geq 0 \) the result follows.

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

**Proof of Proposition 6.1.** We construct a one phase equilibrium, consistent with Lemma 6.1, showing that such an equilibrium exists if and only if \( \nu > (1 - \frac{\phi}{\sigma})(1 - \exp(-\rho \tilde{t}_H)) \), and that the only such equilibrium is the one characterized in the statement of the proposition. To this end, consider the one phase structure, characterized by Lemma 6.1 with \( \tilde{t}_U = 0 \).

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

\[
g(t) = \theta_H \Rightarrow R(t) = \theta_H \Rightarrow F(t) = \frac{1}{\sigma}(1 - \exp(-\lambda \frac{1 - \theta_H}{\theta_H} t)),
\]

64
where the last step uses point (i) of Lemma 6.1 to rule out a mass point in the agent’s mixed strategy, thereby identifying an integration constant. Using point (i) of Lemma 6.1, we have $F(\widetilde{t}_U) = 1$, implying $\widetilde{t}_U = \tilde{t}_H$ as stated in the proposition.

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

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

where we have used Lemma 6.1 (iii) to establish $a_H(t) \in (0, 1)$ and $a_L(t) = 1$. Furthermore, since $a_H(\cdot)$ is differentiable for $t < \tilde{t}_U$. It follows that

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

Solving, we have

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

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

$$a_H(t) = \frac{1}{\nu} \left( \frac{\phi}{\phi} - (1 - \nu) \right) + \left( 1 - \frac{1}{\nu} \left( \frac{\phi}{\phi} - (1 - \nu) \right) \right) \exp\{-\rho(\tilde{t}_H - t)\}$$

$$= \frac{1}{\nu} \left( \frac{\phi}{\phi} - (1 - \nu) \right) + \left( 1 - \frac{1}{\nu} \frac{\phi}{\phi} + \frac{1}{\nu} (1 - \nu) \right) \exp\{-\rho(\tilde{t}_H - t)\}$$

$$= \frac{1}{\nu} \left( \frac{\phi}{\phi} - (1 - \nu) \right) + \frac{1}{\nu} \exp\{-\rho(\tilde{t}_H - t)\}$$

$$= \frac{1}{\nu} \left( \frac{\phi}{\phi} + (1 - \frac{\phi}{\phi}) \exp\{-\rho(\tilde{t}_H - t)\} - (1 - \nu) \right).$$

Therefore, such an equilibrium exists provided two additional conditions,

$$a_H'(t) > 0 \iff \frac{1}{\nu} \left( \frac{\phi}{\phi} - (1 - \nu) \right) < 1 \iff \phi < \hat{\phi}$$

and

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

The first of these holds throughout the paper, the second is the condition given in the proposition.

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

$$a_U(0) - \phi = \nu a_H(t) + (1 - \nu) - \phi = \frac{\phi}{\phi} + (1 - \frac{\phi}{\phi}) \exp\{-\rho\tilde{t}_H\} - \phi,$$
which is identical to the strategic agent’s payoff in the main model, when facing a principal with known standard $\theta_H$. The high type principal gets payoff zero from any arrival inside the phase of doubt, and a payoff of one if the phase of credibility is reached. Thus, the high type principal’s payoff is

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

exactly as in the main model where the principal’s payoff is $\theta_H$. The low type principal’s payoff, in this equilibrium is

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

In this equilibrium,

$$1 - \sigma F(t) = \exp(-r_H t) \quad \sigma f(t) = r_H \exp(-r_H t),$$

where $r_H \equiv \lambda_{1-\theta_H}$. and hence, the low type principal’s payoff is

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

This is exactly the principal’s payoff in the one stage auditing equilibrium, where $\theta_1 = \theta_H$ and $\theta = \theta_L$. As shown in Proposition 5.3, this payoff is larger than the payoff of the low type in the main model, where she is known to have standard $\theta_L$.

**Normative Analysis.** Point (i) is obvious, since the high type principal’s payoff is the same as under transparency, where she is known to be the high type. (ii) follows immediately, since the low type principal’s payoff is higher than in the main model when she is known to be the low type. (iii) Note that under opacity, the strategic agent’s expected payoff is $\frac{\hat{\phi}}{\phi} + (1 - \frac{\hat{\phi}}{\phi}) \exp(-\rho \tilde{t}_H) - \phi$, as if she faces the high standards principal with probability 1. With transparency, his payoff is

$$\nu\left(\frac{\phi}{\phi} + (1 - \frac{\phi}{\phi}) \exp(-\rho \bar{t}_H) - \phi\right) + (1 - \nu)\left(\frac{\phi}{\phi} + (1 - \frac{\phi}{\phi}) \exp(-\rho \bar{t}_L) - \phi\right),$$

where $\bar{t}_L \equiv -\ln(1 - \sigma)\theta_L/(\lambda(1 - \theta_L))$. Because $\bar{t}_L < \bar{t}_H$, the agent’s payoff is larger with transparency. (iv) Analogous to (iii), with the additional observation that Corollary 4.1 guarantees that the ethical agent’s payoff is decreasing in $\theta$. 

**Proof of Proposition 6.2.** We construct a two stage equilibrium, consistent with Lemma 6.1, showing that such an equilibrium exists if and only if $\nu < (1 - \frac{\phi}{\phi})(1 - \exp(-\rho \tilde{t}_H))$, and that
the only such equilibrium is the one characterized in the statement of the proposition. To this end, consider the two phase structure, characterized by Lemma 6.1 with $\tilde{t}_U > 0$.

**Strategies and Phase Transitions.** From Lemma 6.1, in phase 1, we have $a_L(t) \in (0, 1)$, and hence

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

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

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

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

Solving, we have

$$a_L(t) = \left(\frac{1}{1 - \nu}\right)\frac{\hat{\phi}}{\phi} + \kappa_1 \exp\{\rho t\},$$

where $\kappa_1$ is an integration constant. Using boundary condition $a_L(\tilde{t}_U) = 1$ we have

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

$$= \left(\frac{1}{1 - \nu}\right)\left[\frac{\phi}{\hat{\phi}} + (1 - \frac{\phi}{\hat{\phi}} - \nu) \exp\{-\rho(\tilde{t}_U - t)\}\right].$$

From Lemma 6.1, in phase 2, we have $a_H(t) \in (0, 1)$, and hence,

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

where $\kappa_2$ is an integration constant. From the boundary condition $F(\tilde{t}_U) = 1$ we have

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

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

The differential equation for the agent’s indifference condition in phase 2 is identical to the differential equation for the indifference condition in the one phase equilibrium, characterized in Proposition 6.1 case. Following the same argument with boundary condition $a_H(\tilde{t}_U) = 1$, we have

$$a_H(t) = \frac{1}{\nu}\left(\frac{\phi}{\hat{\phi}} + (1 - \frac{\phi}{\hat{\phi}}) \exp\{-\rho(\tilde{t}_U - t)\} - (1 - \nu)\right).$$
To determine the phase transitions, $\tilde{t}_U, \tilde{t}_H$, we use continuity of $F(\cdot)$ at $\tilde{t}_U$, and the boundary condition $a_H(\tilde{t}_U) = 0$, both of which come from Lemma 6.1.

\[
\frac{1}{\sigma}(1 - \exp\{-\mu_L \tilde{t}_U\}) = \frac{1}{\sigma}(1 - (1 - \sigma) \exp\{\mu_H(\tilde{t}_U - \tilde{t}_U)\})
\]
\[
\frac{1}{\nu} \left(\frac{\phi}{\hat{\phi}} + (1 - \frac{\phi}{\hat{\phi}}) \exp\{-\rho(\tilde{t}_U - \tilde{t}_U)\} - (1 - \nu)\right) = 0
\]

Simplifying both equations, we have

\[
\exp\{-\mu_L \tilde{t}_U\} = (1 - \sigma) \exp\{\mu_H(\tilde{t}_U - \tilde{t}_U)\}
\]
\[
\exp\{-\rho(\tilde{t}_U - \tilde{t}_U)\} = 1 - \frac{\nu \hat{\phi}}{\hat{\phi} - \phi}
\]

Taking logs and rearranging, we have

\[
(\mu_L - \mu_H) \tilde{t}_U + \mu_H \tilde{t}_U = -\ln(1 - \sigma)
\]
\[
\rho(\tilde{t}_U - \tilde{t}_U) = -\ln(1 - \frac{\nu \hat{\phi}}{\hat{\phi} - \phi}).
\]

Note that for the second equation to have any solution, we must have $\nu < 1 - \phi/\hat{\phi}$. Solving, we have

\[
\tilde{t}_U = \tilde{t}_L - \frac{\mu_H}{\mu_L} \delta_U
\]
\[
\tilde{t}_U = \tilde{t}_L + (1 - \frac{\mu_H}{\mu_L}) \delta_U,
\]

where $\delta_U = -\ln(1 - \frac{\nu \hat{\phi}}{\hat{\phi} - \phi})/\rho$.

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

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

\[
\tilde{t}_U > 0 \iff \tilde{t}_L > \frac{\mu_H}{\mu_L} \delta_U \iff \tilde{t}_U > \delta_U \iff
\]
\[
-\rho \tilde{t}_H < \ln(1 - \frac{\nu \hat{\phi}}{\hat{\phi} - \phi}) \iff
\]
\[
\exp\{-\rho \tilde{t}_H\} < 1 - \frac{\phi \nu}{\hat{\phi} - \phi} \iff \nu < (1 - \frac{\phi}{\hat{\phi}})(1 - \exp\{-\rho \tilde{t}_H\}).
\]

It follows that the two stage equilibrium exists when $\nu < (1 - \frac{\phi}{\hat{\phi}})(1 - \exp\{-\rho \tilde{t}_H\})$, and in this equilibrium the strategies and phase transitions are the ones given in the proposition. One additional claim is made in the phase transitions part of the proposition, which we verify next.
**Claim 2.** We show that if \( \nu < (1 - \frac{\phi}{\phi})(1 - \exp\{-\rho \bar{t}_H\}) \), then \( \bar{t}_U < \bar{t}_H \). From claim 1, for such \( \nu \) we have

\[
\bar{t}_U > 0 \Rightarrow \bar{t}_L > \frac{\mu_H}{\mu_L} \delta_U \Rightarrow \bar{t}_H > \delta_U \Rightarrow \bar{t}_H(1 - \frac{\mu_H}{\mu_L}) > (1 - \frac{\mu_H}{\mu_L}) \delta_U \Rightarrow
\]

\[
- \ln(1 - \sigma)(\frac{\mu_L - \mu_H}{\mu_H \mu_L}) > (1 - \frac{\mu_H}{\mu_L}) \delta_U \Rightarrow - \ln(1 - \sigma)(\frac{1}{\mu_H} - \frac{1}{\mu_L}) > (1 - \frac{\mu_H}{\mu_L}) \delta_U \Rightarrow
\]

\[
\bar{t}_H - \bar{t}_L > (1 - \frac{\mu_H}{\mu_L}) \delta_U \Rightarrow \bar{t}_H > \bar{t}_L + (1 - \frac{\mu_H}{\mu_L}) \delta_U = \bar{t}_U
\]

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

\[
\mu_H \bar{t}_U + \ln(1 - \sigma) = \mu_H (- \ln(1 - \sigma) + (1 - \frac{\mu_H}{\mu_L}) \delta_U) + \ln(1 - \sigma) =
\]

\[
(1 - \frac{\mu_H}{\mu_L}) \ln(1 - \sigma) + \mu_H (1 - \frac{\mu_H}{\mu_L}) \delta_U =
\]

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

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

**Beliefs.** Follows immediately from Lemma 6.1 and the principal’s sequentially rational acceptance decision.

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

\[
a_U(0) - \phi = (1 - \nu)a_L(0) - \phi = \frac{\phi}{\phi} + (1 - \frac{\phi}{\phi} - \nu) \exp(-\rho \bar{t}_U) - \phi.
\]

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

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

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

\[
V_L = \int_{\bar{t}_U}^{\infty} \exp(-(\rho + \lambda)t)(\lambda(1-\sigma F(t))(1-\theta_L)-\sigma f(t)\theta_L)dt+(1-\theta_L)(1-\sigma) \int_{\bar{t}_U}^{\infty} \lambda \exp(-(\rho + \lambda)t) dt
\]
where $F(\cdot)$ and $f(\cdot)$ are the agent’s CDF and PDF of the agent’s mixed strategy. From the equilibrium characterization, we have

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

which implies that the low type principal’s payoff is

$$(1 - \sigma)\exp(\mu_H t_I) \int_{t_I}^{t_U} \exp(-(\rho + \lambda t))(\lambda(1 - \theta_L) - \mu_H \theta_L) dt + (1 - \theta_L)(1 - \sigma) \int_{t_I}^{\infty} \lambda \exp(-(\rho + \lambda)t) dt =$$

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

Using Claim 3 to simplify the leading term, we have

$$\exp(-\mu_L - \mu_H)(t_I - t) \int_{t_I}^{t_U} \exp(-(\rho + \lambda t)) dt + (1 - \theta_L)(1 - \sigma) \int_{t_I}^{\infty} \lambda \exp(-(\rho + \lambda)t) dt.$$

**Normative Comparison.** That the high type principal strictly benefits from opaque standards in this equilibrium follows immediately from $t_U < t_H$, proved in Claim 2. The low type principal’s payoff is identical in form to the principal’s payoff in the two stage auditing equilibrium, with $\delta_U$ replacing $\delta_A$, $\theta_L$ replacing $\theta_0$ and $\theta$, $\theta_H$ replacing $\theta_1$, $\mu_L$ replacing $\mu_0$ and $\mu_H$ replacing $\mu_1$. A similar argument establishes that this payoff is higher than with transparency.

Opacity is worse than transparency for the strategic agent if

$$\frac{\phi}{\phi} + (1 - \frac{\phi}{\phi}) - \nu \exp(-\rho t_I) < (1 - \nu)\left[\frac{\phi}{\phi} + (1 - \frac{\phi}{\phi}) \exp(-\rho t_L) + \nu\left(\frac{\phi}{\phi} + (1 - \frac{\phi}{\phi}) \exp(-\rho t_H)\right)\right]$$

$$(1 - \frac{\phi}{\phi}) - \nu \exp(-\rho t_I) < (1 - \nu)(1 - \frac{\phi}{\phi}) \exp(-\rho t_L) + \nu\left(\frac{\phi}{\phi} + (1 - \frac{\phi}{\phi}) \exp(-\rho t_H)\right)$$

$$\frac{(1 - \frac{\phi}{\phi}) - \nu}{(1 - \frac{\phi}{\phi})} \exp(-\rho t_I) < (1 - \nu) \exp(-\rho t_L) + \nu \exp(-\rho t_H)$$

$$\exp(-\rho t_I) < (1 - \nu) \exp(-\rho t_L) + \nu \exp(-\rho t_H)$$

We show that for $\nu$ close to $\nu^*$, the strategic agent is worse off under opacity. Note that when $\nu = \nu^*$, $t_I = t_H = \delta_U$. Because $t_L < t_U$, the RHS is strictly bigger. By implication, when $\nu = \nu^*$, the strategic agent is worse off with opacity. By continuity, this holds for $\nu \in (\nu - \epsilon, \nu^*)$ for some $\epsilon > 0$.

Next, we show that the strategic agent’s ranking as $\nu \to 0$ is ambiguous. Consider the difference,

$$\exp(-\rho t_I) - (1 - \nu) \exp(-\rho t_L) - \nu \exp(-\rho t_H).$$
As $\nu \to 0$, we have $\delta_U \to 0$, and thus, $\tilde{t}_U \to \tilde{t}_L$. By implication, the difference above approaches 0. Next, consider the derivative of this difference in $\nu$,

$$-\rho \exp(-\rho \tilde{t}_U) \frac{d\tilde{t}_U}{d\nu} + \exp(-\rho \tilde{t}_L) - \exp(-\rho \tilde{t}_H).$$

Note that

$$\frac{d\tilde{t}_U}{d\nu} = (1 - \frac{\mu_H}{\mu_L}) \frac{d\delta_U}{d\nu} = -(1 - \frac{\mu_H}{\mu_L}) \frac{-\hat{\phi}}{\delta - \phi} = (1 - \frac{\mu_H}{\mu_L}) \frac{-\hat{\phi}}{\rho(1 - \nu \frac{\hat{\phi}}{\delta - \phi}).$$

Therefore, the derivative of the difference evaluated at $\nu = 0$ is,

$$\exp(-\rho \tilde{t}_L)(1 - \frac{\mu_H}{\mu_L}) \frac{-\hat{\phi}}{\phi - \phi} - \exp(-\rho \tilde{t}_H) =

(1 - \sigma) \frac{\rho}{\mu_L} (1 - \frac{\mu_H}{\mu_L}) \frac{-\hat{\phi}}{\phi - \phi} - (1 - \sigma) \frac{\rho}{\mu_H}.$$

We seek to show that this derivative may be either positive or negative. Consider the positive case,

$$(1 - \sigma) \frac{\rho}{\mu_L} (1 - \frac{\mu_H}{\mu_L}) \frac{-\hat{\phi}}{\phi - \phi} - (1 - \sigma) \frac{\rho}{\mu_H} > 0 \iff

(1 - \sigma) \frac{\rho}{\mu_L} (1 - \frac{\mu_H}{\mu_L}) \frac{-\hat{\phi}}{\phi - \phi} - (1 - \sigma) \frac{\rho}{\mu_H} > 0 \iff

(1 - \sigma) \frac{\rho}{\mu_H} \frac{(\mu_L - \mu_H)}{\mu_L \rho_{PH}} > (1 - \frac{\hat{\phi}}{\phi}) \frac{\mu_L}{\mu_L - \mu_H}.$$

Because the exponent is negative, the left hand side approaches infinity as $\sigma \to 1$. Thus, for small $\sigma$ the derivative is positive. Furthermore, if the right hand side exceeds one, because $\mu_L - \mu_H$ is relatively small, then for $\sigma$ large the derivative in question may be negative. Thus, for small $\nu$, the effect of opacity on the payoff of the strategic agent is ambiguous. \qed

### A.4 Proofs for Fighting Fraud

**Proof of Lemma 7.1.** Consider a set of times $S$ (of non-zero measure), such that for all $t \in S$ we have $u(t) < u(\infty)$. Consider the following perturbation of $a(t)$. For all $t$, replace $a(t)$ by $\hat{a}(t) = a(t) + \delta(t)$, where

$$\delta(t) = \min\{\exp((\rho + \lambda)t)(u(\infty) - u(t)), 1 - a(t)\} \quad \text{if} \quad t \in S,$$

$$\delta(t) = 0 \quad \text{otherwise.}$$
Note that $\delta(t) \geq 0$ and that $\hat{a}(t) \leq 1$ for all $t$. Under the perturbed strategy, the agent’s payoff at time $t$ is

$$\hat{u}(t) = \int_0^t \lambda \exp\{-(\rho + \lambda)s\} (a(s) + \delta(s))\, dt + \exp\{-(\rho + \lambda)s\} (a(t) + \delta(t) - \phi)$$

$$= \int_0^t \lambda \exp\{-(\rho + \lambda)s\} \delta(s)\, dt + \exp\{-(\rho + \lambda)t\} \delta(t) + u(t).$$

**Step 1.** We show that for all $t$, incentive compatibility holds for the perturbed acceptance strategy. Incentive compatibility requires

$$\hat{u}(\infty) \geq \hat{u}(t) \iff \int_0^\infty \lambda \exp\{-(\rho + \lambda)t\} \delta(s)\, dt + u(\infty) \geq \int_0^t \lambda \exp\{-(\rho + \lambda)t\} \delta(t)\, dt + \exp\{-(\rho + \lambda)t\} \delta(t) + u(t)$$

$$\int_t^\infty \lambda \exp\{-(\rho + \lambda)t\} \delta(s)\, dt + u(\infty) - u(t) \geq \exp\{-(\rho + \lambda)t\} \delta(t).$$

From the construction of $\delta(t)$ we have $u(\infty) - u(t) \geq \exp(-(\rho + \lambda)t)\delta(t)$ and $\delta(t) \geq 0$. Thus, we have

$$\int_t^\infty \lambda \exp\{-(\rho + \lambda)t\} \delta(s)\, dt + u(\infty) - u(t) \geq u(\infty) - u(t) \geq \exp\{-(\rho + \lambda)t\} \delta(t),$$

which establishes IC.

**Step 2.** We show that the perturbation strictly increases the principal’s payoff. Note that the perturbation increases $u(\infty)$ by $\int_0^\infty \lambda \exp\{-(\rho + \lambda)s\} \delta(s) > 0$ by construction. \qed