Honest equilibria in reputation games: 
The role of time preferences

Melis Kartal*

Vienna University of Economics and Business  
Welthandelsplatz 1, 1020 Vienna, Austria.  
E-mail: melis.kartal@wu.ac.at

January 6, 2017

Abstract

New relationships are often plagued with uncertainty because one of the players has some private information about her “type”. The reputation literature has shown that equilibria which reveal this information typically involve breach of trust and conflict. But are these inevitable for equilibrium learning? I analyze self-enforcing relationships where one party is privately-informed about her time preferences. I show that there always exist honest reputation equilibria, which fully reveal information and support cooperation without breach or conflict. I compare these to dishonest reputation equilibria from several perspectives. My results are applicable to a broad class of repeated games.

* I would like to thank two anonymous referees, Alessandro Lizzeri, Debraj Ray, Ariel Rubinstein, Karl Schlag, Matan Tsur, and Kemal Yildiz for helpful comments and suggestions.
1 Introduction

How can two agents build cooperation when one of them does not fully know the other’s motives and preferences? Consider a repeated relationship (e.g., a buyer-seller, principal-agent or lender-borrower relationship), where incentives are informal because actions and outcomes are not verifiable or contractible. In such a relationship that lacks formal enforcement, trust is essential. However, a buyer may be uncertain about the trustworthiness of a seller, a lender may have doubts about the repayment incentives of a borrower, and an agent may not know the quality of a prospective principal’s work environment.

Whenever there are “better” types in a market, the “worse” types have an opportunistic incentive to imitate them. Starting with the seminal works by Kreps and Wilson (1982), Milgrom and Roberts (1982), and Sobel (1985), the reputation literature has shown that the “bad” type exploits the informational asymmetry and accepts short-run losses in order to build a “good” reputation and obtain a high profit in the long-run. This profit is eventually attained by breaching trust: A bad type borrower defaults, a bad type principal withholds the agent’s bonus, and a bad type seller delivers a defective product. Such deviation from the implicit agreement generates conflict, and the relationship likely breaks down. So, a natural question follows. Are such imitation and breach of trust inevitable for equilibrium learning? For example, young firms are often advised to invest in a “name” via advertising, hiring reputable executives or board members and engaging in social responsibility acts in order to signal a good type. How effective are “costly signals” when incentives are only informal? Is there still an effective signaling mechanism which reveals information and supports cooperation without breach and conflict, and if yes, under what conditions?

These questions are important because a risk of breach may be (i) inefficient as it is correctly anticipated by the other party in equilibrium, and (ii) costlier than is presumed in standard models because individuals may exhibit an aversion to disappointment or betrayal. I study these questions in the context of relational contracts with persistent private information—my results also generalize to other settings. In the model, the principal is privately informed about her time preferences. This represents a situation in which the agent does not fully know the preferences and the commitment of the principal at the beginning of the relationship. I characterize the “honest” and “dishonest” information revelation regimes

---

1See Fombrun and Shanley (1990) and Deutsch and Ross (2003).
2One important exception in contract theory is the seminal work by Hart and Moore (2008), which assumes that people care about what they receive relative to what they are “entitled” to, and getting less than what one is entitled to causes a disutility in the form of “aggrievement” and even leads to retaliation. Also, see Bohnet et al. (2008), Abeler et al. (2011), and Gill and Pworse (2012) for experimental studies on betrayal aversion and disappointment aversion, and Gul (1991), and Kószegi and Rabin (2006, 2007) for theoretical models of disappointment aversion and loss aversion. In these models, agents are sensitive to downside deviations from their expectations and incur a psychological disutility when they receive less than what they expected or deemed “fair”.
of the game and compare them from several perspectives.\textsuperscript{3} I show that there always exists an equilibrium which fully separates the principal types in an honest manner via costly signaling. Differential time preferences of types have a unique role in this result because a separating equilibrium does not generally exist with other types of private information. Hence, differential discounting is analogous to a “monotonicity condition” for repeated private information games. Moreover, I find that there exist parameters under which separating equilibria dominate other types of equilibria. These results have a simple, intuitive structure and generalize to other settings where reputations matter.

I now outline the model and its assumptions. I develop a relational contracting model in which the agent’s discount factor is fixed and known, whereas the principal’s discount factor is her private information. The principal is one of two types: high or low. The high type has a higher discount factor than the low type. It is well known from the theory of repeated games that a high discount factor is associated with more cooperative behavior whereas a low discount factor typically results in opportunism. Similarly, in the context of relational contracts, the higher the discount factor, the easier it is to honor promises because it increases the value of the future trade between the two parties. Thus, discount factor is a simple proxy for trustworthiness and commitment in my model. Traits such as commitment and trustworthiness can be difficult to observe. This is why, for example, Forbes has been announcing “America’s most trustworthy companies” every year. Indeed, some high-profile cases that fall on the opposite ends of the trustworthiness scale indicate that firms differ in the value that they give to their workforce. Southwest Airlines famously declined to quit its no-layoff policy and kept the employee morale high even in difficult times, whereas IBM and Credit Suisse First Boston chose to renege on their promises regarding bonus payments or layoff policies in order to cut costs (Stewart, 1993; Conlin, 2001).

At the beginning of each period, the principal makes a compensation offer to the agent, which the agent either accepts or rejects. The offer consists of a legally enforceable fixed wage and a performance-contingent bonus transfer, which is not legally enforceable. If the agent accepts the offer, then he chooses an effort level. Output is strictly increasing in effort, and exerting effort is costly. Output is observed by both the principal and the agent but cannot be verified by a third party. At the end of each period, the party responsible for making the bonus transfer decides whether or not to honor it.

Nonverifiability of output limits the scope of cooperation even in the absence of private information because incentives are only informal. Introducing private information to the standard relational setting exacerbates the incentive problem further. How does this informational asymmetry make things worse? Since a bonus promise is not legally enforceable, the willingness of the principal to honor a bonus promise depends on her future profit from

\textsuperscript{3}I refer to an equilibrium as “dishonest” if it involves breach of trust with a positive probability and as “honest” otherwise.
the relationship, which is higher for the high type principal because she is more patient. Thus, the high type principal can fulfill higher bonus promises than the low type. The low type then wants to mimic the high type so that she can obtain high effort from the agent with a high bonus promise which she then defaults on. Therefore, the high type principal needs a credible signaling mechanism in order to separate herself from the low type.

I show that there always exists a separating equilibrium; the two types separate immediately by offering different contracts. Although learning is immediate in such an equilibrium, the high type must distort her behavior for an extended period of time, and there is delay in full cooperation with many parameter values.

A separating equilibrium is honest and free of conflict as it involves no imitation, breach of trust or punishment. This result contrasts with the standard reputation literature, where equilibrium information revelation involves dishonesty and conflict; in these “dishonest reputation equilibria,” there is a strictly positive probability that a bonus is unpaid, a rip-off price is charged for a low-quality product, the loan is not repaid, etc. My framework identifies the conditions under which equilibrium information revelation and cooperation do not involve these. In particular, it shows that honest information revelation is possible in any repeated private information game as long as (i) types differ in their time preferences; and (ii) the game allows for costly signals, such as monetary transfers between parties or “money-burning” activities.

How can one reconcile the result above with the defaults at IBM and Credit Suisse First Boston? Earle and Sabirianova Peter (2009) note that “there is no systematic data collection about breaches of the wage contract in most economies—perhaps because they are rare”. This may not be so surprising in light of the recent findings regarding disappointment and betrayal aversion. Many people will presumably reject an offer which exposes them to a nontrivial risk of exploitation. Moreover, dishonest reputation equilibria can be inefficient, and they are generally not robust to a dynamic version of the Intuitive Criterion as I discuss below. Thus, in both theory and practice, an honest and conflict-free arrangement may be a plausible norm in economic relationships even if individuals enter a relationship with private information.

---

\[4\] I define a separating equilibrium as a perfect Bayesian equilibrium in which the agent’s belief about the principal’s type is degenerate after the initial contract offer as well as any equilibrium path history.

\[5\] In addition to the classic references mentioned before, see Diamond (1989), Watson (1999; 2002), Mailath and Samuelson (2001), Halac (2012) and Jullien and Park (2014). Private information is not about time preferences in any of these models.

\[6\] I show explicitly that the main results go through in a buyer-seller relationship with moral hazard and asymmetric information à la Mailath and Samuelson (2001).

\[7\] There is empirical evidence for this from labor markets, in addition to the experimental evidence from studies mentioned in Footnote 2. For example, Mas (2006) shows that police performance declines sharply when police officers lose arbitrations and argues that considerations of fairness and disappointment affect workplace behavior.
Next, I investigate the properties of the optimal separating contract. The low type principal offers her optimal symmetric-information contract immediately, and the game continues as in the symmetric-information setting for her. The high type, however, engages in dynamic signaling that evolves until it reaches the symmetric-information benchmark in finite time. Along the high type’s equilibrium path, the effort exerted by the agent and the surplus in the relationship increase gradually. This is intuitive because discounting affects the prospective payoffs of the two types differently; the high type must forgo earlier profits and delay full cooperation in order to make imitation less tempting for the low type. This gradualism result also emerged in previous studies with asymmetric information, albeit in the context of dishonest reputation equilibria. As it turns out, a gradualist pattern of cooperation is superior even in settings where learning is immediate because it optimally spreads costly signaling over time.\textsuperscript{8}

I then characterize other equilibria of the game; namely, pooling equilibria in which no information is revealed and hybrid equilibria which are dishonest as they involve imitation and breach by the low type. I find that the optimal separating contract always dominates pooling equilibria, and I characterize the conditions under which it also dominates hybrid equilibria. Next, I develop a dynamic version of the Intuitive Criterion (Cho and Kreps, 1987) and show that it typically selects the optimal separating equilibrium as the unique reasonable equilibrium.

To reiterate, differential time preferences of types have a unique role in repeated private information games. A separating equilibrium does not generally exist if the principal’s private information is uncorrelated with her time preferences. This role of discounting connects my paper to the model of Becker and Mulligan (1997) in which an individual is endowed with a baseline level of discount rate and can choose to increase this level at some cost. One of the main insights of their model is that an increase in future payoffs boosts the incentive to invest in one’s patience; therefore, they argue, higher income individuals are more patient. If a “good” type expects a higher future payoff than the “bad” type in a private information game (as is often the case), there is yet another incentive to invest in a higher discount factor; separation from the bad type is surely possible and also cheaper. The upshot is that if individuals can invest in their time preferences, and if the good type has a higher lifetime payoff, then a separating equilibrium always exists.

2 Related Literature

This paper draws from the study of relational contracts and the literature on reputation and dynamic signaling. Below, I discuss the two literatures separately.

\textsuperscript{8}Even after information is fully revealed, costly signaling may continue (see the discussion in Section 2.1).
2.1 Reputation and Dynamic Signaling

My work relates to the literature on reputation building, and in particular to Sobel (1985), Diamond (1989), Ghosh and Ray (1996), Kranton (1996), Watson (1999, 2002), Tadelis (1999), Mailath and Samuelson (2001), Halac (2012) and Jullien and Parks (2014). In these models, separating equilibria do not exist, and information revelation is dishonest: In equilibrium, there is a nontrivial probability that the bonus is unpaid, the buyer pays a rip-off price for a low-quality product, the loan is not repaid, etc. Furthermore, such events often take place at high stakes if relationships can build up over time (see, for example, Sobel (1985), Watson (1999, 2002) and Halac (2012)). In contrast with these models, I identify conditions under which separating equilibria always exist, and thus, equilibrium learning is honest and partnerships are long-lasting. I analyze the efficiency properties of separating equilibria and compare them to other types of equilibria. I also show that in the optimal separating equilibrium, the relationship becomes more cooperative and valuable over time provided that the informed party is a high type. While I obtain these results in a principal-agent model, they generalize to other settings. Section 7.1 presents an application of the results to a buyer-seller relationship à la Mailath and Samuelson (2001).

My work also relates to a subset of the dynamic signalling literature in which out-of-equilibrium degenerate beliefs are allowed to change. This property implies that costly signaling can continue even after beliefs have become degenerate. This property has been applied in numerous models such as Admati and Perry (1987), Noldeke and Van Damme (1990), Cramton (1992), Kremer and Skrzypacz (2007), and Kaya (2009). Admati and Perry (1987) and Cramton (1992) analyze a bargaining game, whereas Noldeke and Van Damme (1990) extend Spence’s job market signaling model to a dynamic environment. With the exception of Kaya (2009), (i) these games are dynamic but not repeated; and (ii) the range of the signaling variable is too small, and therefore, distorting behavior in only one period is not sufficient to achieve separation. Kaya (2009) extends the concept of dynamic signaling to a class of repeated games where the stage game has a separating equilibrium by construction. She then characterizes the least cost separating equilibrium of the repeated game. My model differs from the aforementioned papers in two ways. Among them, only

\[9\] It may be argued that once the agent attaches probability one to the principal being a certain type, the subsequent game should essentially be one of symmetric information. Yet, this requires a restriction on beliefs because it is not implied by the definition of perfect Bayesian equilibrium. Moreover, this type of restriction sometimes leads to non-existence of a perfect Bayesian equilibrium.

\[10\] In the basic job market signaling model by Spence, the game is static, and the more productive type separates from the less productive type simply by investing in education. Even though there is a multiplicity of equilibria, only the Riley outcome (i.e., the Pareto-best separating equilibrium) survives the Intuitive Criterion. Could the Riley outcome survive in a continuous time setting? If it is to survive, then costly education must continue after beliefs have become degenerate. This example, presented in Noldeke and Van Damme (1990), illustrates why it is plausible to have costly signaling after beliefs have become degenerate.
Kaya (2009) analyzes a repeated game. But the stage game of her model has a separating equilibrium by assumption, whereas different types would exhibit identical behavior if the stage game in my setting is repeated only once because incentives are only informal. Thus, these papers do not address the issues that I study. Moreover, the signal space is essentially endogenous in my model (as it depends on the surplus generated within the relationship), and it is not at all clear whether or not a separating equilibrium exists.

2.2 Relational Contracting

A relational contract sustains trade between a principal and an agent if the performance measure is nonverifiable. Numerous relational contracting models focused on environments with symmetric information (see, for example, MacLeod and Malcolmson (1989, 1998), and Board and Meyer-Ter-Vehn (2013)). Asymmetric information has also been incorporated into relational contracting. Shapiro and Stiglitz (1984), and Baker, Gibbons and Murphy (2002) consider relational contracts with moral hazard, whereas Levin (2003) analyzes two distinct scenarios. He assumes that either the agent’s effort or the agent’s time-specific cost parameter cannot be observed by the principal. MacLeod (2003) and Fuchs (2007) consider asymmetric information about output realization, whereas Li and Matouschek (2013) assume that there is asymmetric information about the state of the world which affects the opportunity cost of paying a bonus to the agent. Asymmetric information has no persistence in these papers.

To my knowledge, there are only two papers that analyze relational contracts with persistent asymmetric information. Halac (2012) analyzes relational contracts in a setting where the principal has persistent private information regarding her outside option. As discussed above, separating equilibria do not exist, and information revelation is dishonest in Halac (2012): If the principal is the less cooperative type, then the relationship breaks down after a while because the principal defaults on a payment promise. Yang (2013) assumes that each agent has persistent private information about his ability but his setting differs from mine as he considers relational contracting in a repeated matching market where matches are constantly reshuffled, and the informed party (i.e., the agent) is protected by limited liability.

In Section 7.1, I re-analyze the model of Mailath and Samuelson (2001) in a relational environment and show that my main results apply to this setting. In Mailath and Samuelson, a seller can repeatedly sell a product to a long-lived buyer or to a sequence of short-lived buyers. In each period, the quality of the product is stochastic. The quality depends on the effort chosen by the seller in that period, and exerting effort is costly. Unlike Mailath and Samuelson (2001), I allow the seller to send a non-binding message regarding the realized quality (see Jullien and Park (2014) for a similar approach). The quality announcement is cheap-talk just like a principal’s bonus announcement, but truthful announcements may be enforceable, and the seller may be induced to exert effort perpetually if the future profit from
the relationship is high enough.\textsuperscript{11} If the seller types differ in their discount factors, then my main results hold, whereas separation is not generally possible if the private information is on another parameter, such as the seller’s cost of effort or ability.\textsuperscript{12}

\section{The Model}

Two risk-neutral parties, a principal (she) and an agent (he) interact repeatedly in periods $t = 0, 1, \ldots$. The agent’s discount factor is $\delta$, which is fixed and known, whereas the principal’s discount factor is $\delta_\theta$, where $\theta \in \{l, h\}$ is the principal’s private information and $\delta_l < \delta_h$. The principal learns her type at the beginning of the initial period, and this type remains the same in all subsequent periods.

At the beginning of period $t \geq 0$, the principal makes a contract offer to the agent. The agent either accepts this offer or rejects it: $d_t \in \{0, 1\}$ denotes the agent’s decision, where $d_t = 1$ if the agent accepts the offer, and $d_t = 0$ otherwise. If the agent accepts the offer, then he chooses effort $e_t \in [0, \bar{e}]$ and incurs a cost $c(e_t)$, where $c$ is strictly increasing, differentiable and convex with $c(0) = 0$ and $c'(\bar{e}) = \infty$. Agent’s effort $e_t$ generates the output $y_t = y(e_t)$, where $y$ is strictly increasing, differentiable and concave. The term $s(e) \equiv y(e) - c(e)$ represents the expected surplus given the effort level $e$. Output is observed by both the principal and the agent but cannot be verified by a third party.\textsuperscript{13} Contract offer at the beginning of period $t$ consists of a fixed wage $w_t$ and a bonus transfer $b_t$ contingent on performance. This contract offer is denoted by $C_t = \{w_t, b_t\}$. The fixed wage $w_t$ is legally enforceable, whereas the bonus payment $b_t$ is not. After the output realization in period $t$, the party that is responsible for making the bonus payment $b_t$ decides whether or not to honor the payment. If $b_t > 0$, then the decision belongs to the principal, whereas the agent makes the decision if $b_t < 0$. Total payment from the principal to the agent is denoted by $P_t$, where $P_t = w_t + b_t$ if the promised payment is honored, and $P_t = w_t$ otherwise. Thus, the agent’s per-period payoff is $P_t - c(e_t)$ and the principal’s is $y_t - P_t$.

If the agent rejects the principal’s offer, both parties receive their outside options in the current period: $\bar{\pi}$ for the principal and $\bar{u}$ for the agent. There exists an effort level $e$ such that $s(e) > \bar{\pi} + \bar{u} \geq s(0)$.

\textsuperscript{11}Such cheap-talk allows for equilibria in which the seller can be motivated to exert effort even in a symmetric information setting—unlike the case in Mailath and Samuelson (2001). A similar idea still applies if the seller observes the realized quality only after the sale. If the value of the repeated relationship is sufficiently high for the seller, then the seller gives a refund to the buyer in case the quality is observed to be lower than the expected quality given equilibrium seller effort.

\textsuperscript{12}My results also extend to the case in which sellers enter and exit the market in every period in a stochastic fashion, and names are tradable assets, as in Tadelis (1999) and Mailath and Samuelson (2001).

\textsuperscript{13}Since output is a perfect measure of effort, whether or not effort is observable is inconsequential. Therefore, I assume without loss of generality that effort is unobservable.
Parties care about their discounted payoff stream. As of period \( t \), the respective payoffs for the type-\( \theta \) principal and the agent can be written as
\[
\pi_{\theta,t} = \sum_{\tau=t}^{\infty} \delta^{\tau-t} [d_\tau (y_\tau - P_\tau) + (1 - d_\tau)\overline{p}],
\]
and
\[
u_t = \sum_{\tau=t}^{\infty} \delta^{\tau-t} [d_\tau (P_\tau - c(e_\tau)) + (1 - d_\tau)\overline{u}].
\]

Let \( \mu_0 \) denote the prior probability that the principal is high type. Within period \( t \geq 0 \), the agent updates his beliefs twice: first, after \( C_t \) (but before \( P_t \)) is observed, and second, after \( P_t \) is observed. The term \( \mu^1_t \) denotes the posterior belief of the agent at \( t \) after only \( C_t \) is observed, and \( \mu^2_t \) denotes the posterior belief of the agent at \( t \) after \( P_t \) is observed.

### 3.1 Equilibrium Concept

Let \( h_t = \langle C_{t-1}, d_{t-1}, y_{t-1}, P_{t-1} \rangle \) denote the public outcome at the end of period \( t - 1 \), and let \( h^t = \langle h_0, ..., h_t \rangle \in \mathcal{H}^t \) denote the history up to the beginning of \( t \), where \( \mathcal{H}^t \) represents the set of all possible \( h^t \) realizations with \( h^0 = \mathcal{H}^0 = \emptyset \). A relational contract describes a complete plan for the relationship (Levin, 2003). That is, for every \( h^t \in \mathcal{H}^t \), a relational contract must specify (i) the contract that a principal of type \( \theta \in \{h, l\} \) offers at \( t \); (ii) the posterior belief \( \mu^1_t \) given the contract offer at \( t \); (iii) whether the agent accepts or rejects the offer; (iv) the effort choice in case the agent accepts the contract offer; (v) the bonus payment decision given the output realization; and (vi) the posterior belief \( \mu^2_t \) given the total payment \( P_t \). Such a contract is self-enforcing if it describes a perfect public Bayesian equilibrium (PPBE) of the repeated game. A PPBE is a set of public strategies and posterior beliefs such that strategies form a Bayesian Nash equilibrium in every continuation game given the posterior beliefs, and beliefs are updated according to Bayes’ rule whenever possible. A public strategy depends only on the publicly observed history of play and the player’s payoff-relevant private information. More specifically, the agent conditions his strategy only on the public history whereas the principal conditions her strategy on her discount factor and the public history.

First, I study separating equilibria. A separating equilibrium is a PPBE in which the agent’s belief about the principal’s type is degenerate after the first contract offer at \( t = 0 \) as well as any equilibrium path history; that is, \( \mu^i_t \in \{0, 1\}, i \in \{1, 2\} \) for every \( t \geq 0 \) on the equilibrium path. Second, I analyze pooling contracts.\(^{14}\) Third, I analyze hybrid equilibria.

\(^{14}\)There can be equilibria which are separating after an initial pooling phase during which both types’ contract offers and actions are identical. Eventually, such contracts will prove to be irrelevant for my findings because, as Proposition 7 shows, pooling contracts are dominated contracts, which implies that a contract that is separating after an initial pooling phase is also dominated (see the discussion in Section 5).
in which both types start by offering the same contract, and separation is induced through
default by the low type (or the absence thereof). As a reminder, I refer to an equilibrium
as “dishonest” if it involves breach of trust with a positive probability and as “honest”
otherwise. Thus, separating and pooling equilibria are honest, whereas hybrid equilibria are
dishonest.

3.2 Symmetric-information model

First, I analyze the benchmark setting in which the agent knows the discount factor of the
principal. The analysis for the symmetric-information model consists of two cases: $\delta < 
\delta_0$ and $\delta > \delta_0$. The case with $\delta = \delta_0$ is analyzed in Levin (2003).

I define an optimal contract as the contract that maximizes the profit of the principal
and is not pareto-dominated by another contract. Lemma 1 implies that if $\delta < \delta_0$, then the
optimal symmetric-information contract with a type-$\theta$ principal is stationary; i.e., $e_t = e,$
$b_t = b,$ and $w_t = w$ for all $t.$ Thus, the optimal contract is unique, and there exists no
optimal contract that is nonstationary on the equilibrium path. These contrast with Levin
(2003). In Levin, there are both stationary and nonstationary optimal contracts. It turns
out that the indeterminacy of the optimal contract stems from the knife-edge case in which
the principal and the agent have the same discount factor.\textsuperscript{\textsuperscript{16}}

\textbf{Lemma 1} \textit{In the optimal symmetric-information contract with $\delta < \delta_0$, $u_t = \frac{u}{1-\delta}$ for every
t $\geq 0$ on the equilibrium path. In the optimal contract with $\delta > \delta_0$, $\pi_{\theta,t} = \frac{s}{1-\delta}$ for every $t \geq 1$
on the equilibrium path.}

Lemma 1 implies that if $\delta < \delta_0$, then variations in continuation payoffs are never used
to discipline the agent in the optimal contract. As a result, it is optimal for the principal to
offer the same contract to the agent in every period: She offers a fixed wage $w$ and a bonus
reward $b > 0$ that is contingent on output $y$. The respective life-time payoffs for the type-$\theta$
principal and the agent are given by

\[ \pi_{\theta} = \frac{y - w - b}{1 - \delta_0} = \frac{s - \bar{u}}{1 - \delta_0}, \]

and

\[ u = \frac{w + b - c(e)}{1 - \delta} = \frac{\bar{u}}{1 - \delta}. \]

\textsuperscript{\textsuperscript{15}}The only remaining type of equilibrium is an equilibrium which is hybrid with probability $\alpha \geq 0,$
separating with probability $\beta \geq 0,$ and pooling with probability $(1 - \alpha - \beta)$ at $t = 0.$ See Footnote
27 for details.

\textsuperscript{\textsuperscript{16}}My definition of an optimal contract is more restrictive than that of Levin. However, this is
inconsequential for my claims. If $\delta = \delta_0,$ then there are both stationary and nonstationary optimal
contracts according to my definition, just as in Levin.
In a self-enforcing contract, the principal does not default on a bonus promise. So, there is no loss in assuming that default is punished by ending the relationship, which is the worst possible punishment. Hence, I obtain the enforcement constraint for the type-$\theta$ principal:

\[
\frac{\delta_{\theta}}{1 - \delta_{\theta}}(s - \bar{u} - \bar{\pi}) \geq b.
\]

The optimal symmetric-information contract maximizes the expected surplus \(s(e) = y(e) - c(e)\) subject to the incentive compatibility constraint for the agent’s effort choice \(b \geq c(e)\), the agent’s participation constraint \(w + b - c(e) \geq \bar{u}\), and the enforcement constraint for the type-$\theta$ principal. The terms \(e_{\theta}\) and \(b_{\theta}\) give the solution to the maximization problem, and the maximized surplus is denoted by \(s_{\theta}\). If \(\delta_{\theta}\) is too low, then no contract can be self-enforcing. If, however, \(\delta_{\theta}\) is sufficiently close to one, then even a relational contract achieves the first-best effort. Consistent with the previous literature on self-enforcing contracts, I focus on environments in which trade is feasible but the enforcement constraint is binding, and the first-best outcome cannot be attained. Following,

\[
b_{\theta} = \frac{\delta_{\theta}}{1 - \delta_{\theta}}(s_{\theta} - \bar{u} - \bar{\pi}) = c(e_{\theta})
\]

in the optimal contract, and the fixed wage is given by \(w_{\theta} = \bar{u}\). The contract \(C_{\theta} = \{w_{\theta}, b_{\theta}\}\) implements \(e_{\theta}\) and is called the optimal symmetric-information contract of type $\theta$. Note that since \(\delta_{h} > \delta_{l}\), it follows that \(b_{h} > b_{l}\), and \(s_{h} > s_{l}\).

If \(\delta > \delta_{\theta}\), then the optimal contract is independent of $\theta$, unique, and nonstationary as a result of Lemma 1. However, surplus and effort are stationary. Moreover, the contract becomes fully stationary after the initial period. The principal receives a frontloaded transfer at \(t = 0\) via \(w_{0}\) (the fixed wage at \(t = 0\)), the agent gets the excess of the surplus over \(\bar{\pi}\) in every \(t \geq 1\) via \(w_{t}\), and a bonus is never paid out; that is, \(b = 0\) in equilibrium. Thus, in the optimal contract parties trade payoffs across time if the agent is more patient than the principal. To see why \(b = 0\), first note that \(b \leq 0\) since \(\pi_{\theta,t} = \frac{\pi_{t}}{1 - \delta}\) for every \(t > 0\) by Lemma 1. But if it were the case that \(b < 0\) in equilibrium, then the enforcement constraint for the agent would become

\[
\frac{\delta}{1 - \delta}(s(e) - \bar{u} - \bar{\pi}) \geq c(e) - b,
\]

rather than

\[
\frac{\delta}{1 - \delta}(s(e) - \bar{u} - \bar{\pi}) \geq c(e).
\]

As before, I focus on environments in which trade is feasible but the enforcement constraint is binding. Note that if \(b < 0\), then the former inequality poses a tighter constraint and generates lower equilibrium surplus than the latter inequality. Therefore, the principal chooses \(e \in (0,1)\) to maximize \(s(e) = y(e) - c(e)\) subject to the constraint \(\frac{\delta}{1 - \delta}(s(e) - \bar{u} - \bar{\pi}) = c(e)\).
The term $e^*$ gives the solution to this maximization problem, and $s^*$ denotes the maximized surplus. In the optimal contract, $w_0 = (y(e^*) - \bar{\pi}) - \frac{s^* - \bar{u} - \bar{\pi}}{1 - \delta}$ and $w_t = w = y(e^*) - \bar{\pi}$ for every $t \geq 1$. The contracts $C_0 = \{w_0, 0\}$ and $C_t = C = \{w, 0\}$ for $t > 0$ implement $e^*$ and denote the optimal symmetric-information equilibrium with either principal type.

The symmetric information benchmark relates to the work by Lehrer and Pauzner (1999). Lehrer and Pauzner analyze a class of repeated games in which players have different discount factors. They show that players can mutually benefit from trading payoffs across time; i.e., it is efficient to reward the patient player later and the impatient player earlier. The outcome in Lemma 1 is consistent with their result.

4 Separating Contracts

A one-shot game has a separating equilibrium provided that it satisfies some monotonicity condition, such as the single-crossing property. In infinitely repeated games, it may not be straightforward to find a corresponding property, and the problem is further mitigated in games where incentives are only informal (as in my model) because different types would exhibit identical behavior if the game is played only once. This section shows that a separating equilibrium always exists, and that differential time preferences of types is essential for this result. I characterize the optimal separating contract in Section 4.1 and show when it dominates other equilibria and when it is dominated in Section 5.

On a related note, several studies have documented that individuals may exhibit an aversion to disappointment and betrayal. Several others have explicitly modeled disappointment aversion and loss aversion, and the insights from these studies have also been incorporated into contract theory.\textsuperscript{17} I discuss the implications of a similar approach for my model in Section 7.

I focus on the case in which $\delta < \delta_t$ throughout the remainder of the paper. This is due to the following:

- If $\delta > \delta_\theta$, then the optimal symmetric-information contract of the type-$\theta$ principal depends on $\delta$ but not on $\delta_\theta$, as discussed in the previous section. Hence, if $\delta \geq \delta_h$, then private information does not distort incentives, and symmetric-information contracts can still be implemented. So, I do not pursue this case further.

- If $\delta_h > \delta \geq \delta_t$, then all the forthcoming results that I derive under the assumption that $\delta < \delta_t$ are still valid.

The previous section has shown that the trade between a high type principal and the agent generates a surplus of $s_h$ via the optimal symmetric-information contract $C_h = \{w_h, b_h\}$

\textsuperscript{17}See the references in Footnotes 2 and 7, and Section 7.3.
if the principal’s discount factor is common knowledge. However, $C_h$ is not a credible offer if the principal is privately informed about her discount factor because a low type principal would like to imitate a high type offering $C_h$. To see why, note that $b_h > b_l$ and $s_h > s_l$ since $\delta_h > \delta_l$. Thus, a low type principal who successfully imitates the high type can obtain at least $\frac{s_h - \bar{u}}{1 - \delta_h}$, which is greater than her symmetric-information contract payoff. Moreover, the low type chooses to default immediately because

$$b_h > \frac{\delta_l}{1 - \delta_l} \left( s_h - \bar{u} - \bar{\pi} \right),$$

which further increases her imitation payoff. It follows that the high type principal must use a credible signaling mechanism in order to separate herself from the low type.

The characterization of separating equilibria involves two sets of incentive compatibility constraints. One set of constraints ensures that the low type is deterred from imitating the high type, whereas the second set ensures that the high type is willing to separate and signal her type. In order to simplify the exposition of the incentive constraints below, I assume that the low type offers her optimal symmetric-information contract $C_l$ in every period and obtains $\frac{s_l - \bar{u}}{1 - \delta_l}$ in every separating equilibrium. This assumption is without loss of generality for Proposition 2 and Proposition 3. As a matter of fact, optimality dictates that the low type offer $C_l$ in every period of a separating equilibrium: Inspecting the incentive compatibility constraints below reveal that the low type is strictly worse off, but the high type cannot be made better-off otherwise.\textsuperscript{18}

The low type principal who imitates the high type deviates from the separating contract at some $t$ either refusing to pay $b_t > 0$ or offering a contract different than the equilibrium prescription. I assume that $b_t \geq 0$ in order to simplify the exposition of the incentive compatibility constraints. None of the upcoming results rely on this simplification. Moreover, $b_t > 0$ must hold in the optimal separating contract, as implied by Lemma 4. The sequence $\{w_t, b_t\}_{t=0}^\infty$ represents a separating equilibrium provided that it satisfies conditions (1)-(4); if these conditions are satisfied, then the low type principal is deterred from imitation, the high type is willing to signal her type, and $\{w_t, b_t\}_{t=0}^\infty$ is enforceable for the high type principal.\textsuperscript{19}

\begin{equation}
\forall t \geq 0 : \sum_{k=0}^{t-1} \delta_t^k (y(e_k) - w_k - b_k) + \delta_t^t (y(e_t) - w_t) + \frac{\delta_{t+1}}{1 - \delta_t} \bar{\pi} \leq \frac{s_l - \bar{u}}{1 - \delta_l},
\end{equation}

\textsuperscript{18}I define the optimal contract as the equilibrium contract that maximizes a weighted average of the two principal type’s equilibrium payoffs. Section 4.1 provides a detailed description.

\textsuperscript{19}It is without loss of generality to assume that there is no default in a separating equilibrium and that the worst punishment is imposed in case of a default. Of course, the optimal separating contract cannot involve default on the equilibrium path.
\[
\sum_{k=0}^{\infty} \delta^k (y(e_k) - w_k - b_k) \geq \frac{s_l - \bar{u}}{1 - \delta_h} \quad \text{and} \quad \forall t \geq 0 : \sum_{k=t}^{\infty} \delta^{k-t} (y(e_k) - w_k - b_k) \geq \frac{\bar{\pi}}{1 - \delta_h}, \tag{2}
\]

where \(e_k\) is implemented by \(\{w_k, b_k\}\) and satisfies the agent’s incentive compatibility constraint for effort as well as his participation constraint at every \(k \geq 0\):

\[
\forall k \geq 0 : \quad w_k + b_k - c(e_k) + \sum_{t=k+1}^{\infty} \delta^{t-k} (w_t + b_t - c(e_t)) \geq \max\{w_k + \frac{\delta}{1 - \delta} \bar{u}, \frac{\bar{u}}{1 - \delta}\}. \tag{3}
\]

Finally, the high type’s enforcement constraint holds at every \(t \geq 0\) if the following is satisfied:

\[
\forall t \geq 0 : \quad b_t + \frac{\delta_h}{1 - \delta_h} \bar{\pi} \leq \sum_{k=t+1}^{\infty} \delta^{k-t}_h (y(e_k) - w_k - b_k). \tag{4}
\]

Since \(C_h\) is the optimal symmetric-information contract of the high type principal, a contract offer different than \(C_h\) generates a surplus that is lower than \(s_h\) and reduces the profit of the principal. Thus, the high type’s behavior is “distorted” (put differently, the high type engages in costly signaling) in a separating equilibrium whenever her contract offer differs from \(C_h\). It is sometimes necessary that costly signaling lasts multiple periods because, as Proposition 2 shows, separation may be impossible otherwise. To be more precise, there exists a nontrivial set of parameters such that if the high type offers a contract different from \(C_h\) only at \(t = 0\), then separation is impossible.

**Proposition 2** Separation is not generally possible if the high type’s behavior is distorted only at \(t = 0\). To be more precise, there exists a \(\Lambda > 0\) such that if \(\delta_h - \delta_l < \Lambda\), and \(\delta_h > \delta_l > \delta_l\) (or, if \(\delta_h - \delta < \Lambda\), and \(\delta_h > \delta \geq \delta_l\)) then there exists no separating equilibrium in which the high type’s behavior is distorted in only the initial period.

The formal proof of Proposition 2 is relegated to the Appendix. Note that \(\Lambda\) need not be small. As the discussion below shows, if for example \(\delta_l = 2/3\) and \(\bar{\pi} = \bar{u} = 0\), then \(\delta_h\) values that satisfy \(3s_l > s_h\) also satisfy \(\delta_h - \delta_l < \Lambda\).

I will now spend some time on explaining the intuition behind this proposition since it is instrumental for understanding as to why there always exists a separating equilibrium if the private information is on time preferences, and as to why a separating equilibrium may not

\[20\text{The second incentive compatibility constraint for the high type could also be written as}\]

\[
\forall t \geq 0 : \quad \sum_{k=t}^{\infty} \delta^{k-t} (y(e_k) - P_k) \geq \frac{s_l - \bar{u}}{1 - \delta_h},
\]

in case she wants to stop costly signalling at \(t\) and offer \(C_l\) from then on—assuming that she has never defaulted on a bonus payment and a future payoff greater than \(\bar{u}/(1 - \delta)\) has not been promised to the agent. It might be reasonable to assume that the agent does not punish this because \(C_l\) is enforceable with both types. However, this doesn’t affect any of my results.
exist with other types of private information. To see the intuition behind Proposition 2, first note that if the high type’s behavior is not distorted at all, then the benefit of separation for the high type principal is strictly lower than the benefit of imitation for the low type. For simplicity, assume that $\bar{\pi} = \bar{u} = 0,$ and note that the benefit of separation for the high type equals

$$\beta_h = \frac{s_h}{1 - \delta_h} - \frac{s_l}{1 - \delta_l} = \frac{s_h - s_l}{1 - \delta_h},$$

whereas the benefit of imitation for the low type equals

$$\beta_l = s_h + b_h - \frac{s_l}{1 - \delta_l} = \frac{s_h}{1 - \delta_l} - \frac{s_l}{1 - \delta_l}.$$

Since $\delta_h > \delta_l,$ it follows that $\beta_l > \beta_h.$ The left-hand side of the second equality above follows because an imitator promises $b_h$ and obtains output $y(e_h)$ but defaults on $b_h$ subsequently. The right-hand side follows because the enforcement constraint of the high type is binding in the contract $C_h,$ and thus $b_h = \delta_h \frac{s_h}{1 - \delta_h}$ with $\bar{\pi} = \bar{u} = 0.$

Now, assume that the behavior of the high type principal is distorted only at $t = 0$ in a separating equilibrium. This implies that a low type principal who chooses to imitate the high type optimally defaults either at $t = 0$ or at $t = 1.$ A key observation is that the equilibrium benefit of separation for the high type must exceed the equilibrium benefit of imitation for the low type in a separating contract. For simplicity, I assume that the imitator chooses to imitate the high type fully at $t = 0$ honoring the bonus promise and offers $C_h$ at $t = 1.$ Thus, $t = 0$ can be ignored in the comparison of the equilibrium benefit of separation and the equilibrium benefit of imitation because there is no discounting at $t = 0,$ and the payoff prospects are identical for both types. From $t = 1$ onwards, the low type obtains the benefit of imitation $\beta_l$ as described above, and the high type obtains the benefit of separation $\beta_h.$ Recall that $\beta_l > \beta_h.$ The “discounted” benefit of imitation $\delta_l \beta_l$ is also strictly higher than the “discounted” benefit of separation $\delta_h \beta_h$ if $\frac{s_l}{1 - \delta_l} > s_h$ holds. But this condition can easily hold if $\delta_l$ and $\delta_h$ are not far from each other (as an example, if $\delta_l = 2/3,$ then the condition is satisfied with $\delta_h$ values such that $3s_l > s_h$). In that case, separating contracts in which the high type’s behavior is distorted at only $t = 0$ require excessively costly signaling (because the low type benefits more from imitation than does the high type from separation), and therefore, the high type prefers imitating the low type and offering $C_l$ instead.

Importantly, this problem is fully resolved if the high type continues to distort her behavior at $t > 0$ and delays $C_h$ as well as other high-surplus contracts to a sufficiently distant future. Such delay is effective in separating the principal types because they differ in their patience. Note that Proposition 2 is not specific to the case where the principal’s private

\[21\] Since the bonus equals $b_h$ at every $t \geq 1$ by hypothesis, defaulting at $t > 1$ is strictly dominated.

\[22\] The argument becomes only stronger if the low type strictly prefers defaulting at $t = 0.$
information is on time preferences. Similar or stronger versions of this result obtain with other types of private information, such as productivity or outside option; that is, separation is not generally possible if the high type’s behavior is distorted only at $t = 0$. But, as discussed in more detail below, delaying high cooperation is typically ineffective if the principal types have the same discount factor, and therefore, a result similar to Proposition 2 has general implications when the private information is on another parameter; in particular, a separating equilibrium may not exist.

Delaying $C_h$ and other high-surplus contracts enables separation in my model because the cost of waiting is “monotonic” in type due to the difference in the time preferences of types. More precisely, the equilibrium benefit of separation for the high type exceeds the equilibrium benefit of imitation for the low type with a sufficiently long delay because $\delta_h > \delta_l$, and thus, $\delta_h^T \beta_h > \delta_l^T \beta_l$ must hold with a sufficiently large (but finite) $T$. That is,

$$\delta_h^T \left( \frac{s_h - s_l}{1 - \delta_h} \right) > \delta_l^T \left( \frac{s_h}{1 - \delta_h} - \frac{s_l}{1 - \delta_l} \right). \tag{5}$$

As a result, the high type principal who delays $C_h$ until period $T$, limits the surplus of the relationship to a low level (for example, $s_l$) in the first $T - 1$ periods, and spends the amount

$$\delta_h^T \beta_h = \delta_l^T \left( \frac{s_h}{1 - \delta_h} - \frac{s_l}{1 - \delta_l} \right)$$

(this is precisely the discounted benefit of imitation for the low type) on costly signaling at $t = 0$ is able to separate herself from the low type and strictly prefers doing so. This is the key insight behind the upcoming result.$^{23}$

**Proposition 3** There always exists a separating equilibrium.

**Proof.** I outline a separating mechanism based on the intuition above. The low type’s separating equilibrium contract is always $C_l$. At $t = 0$, the high type principal offers the contract $\{w_h + \Delta, b_l\}$, where $w_h$ is the fixed wage and $b_l$ is the bonus payment contingent on output $y(e_l)$ as specified in $C_l$. Thus, the high type’s contract at $t = 0$ differs from $C_l$ in that the fixed wage offer exceeds $w_l$ by $\Delta$. The amount $\Delta$ represents the initial cost of signaling and will be determined below endogenously ($\Delta$ can be interpreted as a lump-sum signing bonus for the agent). At every $t \in \{1, \ldots, T - 1\}$, the high type offers $C_l = \{w_l, b_l\}$, where $T$ is to be determined endogenously, just like $\Delta.$$^{24}$ From period $T$ onwards, the high type

---

$^{23}$In order to allow for a delay in the contract offer $C_h$ in separating equilibria, I allow out-of-equilibrium degenerate beliefs to change. In other words, even if beliefs put zero probability on a type of principal at some $t \geq 0$, off-the-equilibrium path beliefs can still attach positive probability on that type after $t$. As a result, it is possible for the high type to delay $C_h$ offer for multiple periods and deter the low type from imitation.

$^{24}$If it turns out that $T = 1$, then the high type offers $C_h$ from period 1 onwards.
offers $C_h$. In other words, the high type delays $C_h$ until period $T$ and limits the surplus of the relationship to $s_l$ in the first $T - 1$ periods. A low type principal who imitates the high type honors $b_l$ at every $t < T$ and optimally defaults on $b_h$ in period $t = T$, by construction. Therefore, it suffices to focus on the following incentive compatibility constraint for the low type:

$$\frac{s_l - \bar{u}}{1 - \delta_l} \geq y(c_l) - (w_l + \Delta) - b_l + \sum_{t=1}^{T-1} \delta_l^t (s_l - \bar{u}) + \delta_l^T V_l,$$  \hspace{1cm} (6)

where $V_l$ represents the undiscounted imitation payoff of the low type from $T$ onwards; i.e., $V_l = s_h - \bar{u} + b_h + \frac{\delta_l}{1 - \delta_l}$. The only relevant incentive compatibility constraint for the high type is

$$y(c_l) - (w_l + \Delta) - b_l + \sum_{t=1}^{T-1} \delta_h^t (s_l - \bar{u}) + \frac{\delta_h^T}{1 - \delta_h} (s_h - \bar{u}) \geq \frac{s_l - \bar{u}}{1 - \delta_h},$$  \hspace{1cm} (7)

because the high type would never deviate from costly signaling at $t > 0$ and prefers honoring the bonus promise at every $t \geq 0$ by construction. Thus, (6) and (7) are necessary and sufficient conditions for this construction to result in a separating equilibrium. Period $T$ is determined in a way that the discounted benefit of separation for the high type exceeds the discounted benefit of imitation for the low type, just like in (5). Let $T$ be the lowest $t$ such that

$$\delta_h^t \left( \frac{s_h - \bar{u}}{1 - \delta_h} - \frac{s_l - \bar{u}}{1 - \delta_l} \right) > \delta_l^t \left( V_l - \frac{s_l - \bar{u}}{1 - \delta_l} \right)$$

holds. Once $T$ is found, $\Delta$ is determined in a way that the low type is deterred from imitation in the least costly way; that is, $\Delta$ is equal to $\delta_l^T \left( V_l - \frac{s_l - \bar{u}}{1 - \delta_l} \right)$, which is the discounted benefit of imitation for the low type. Using this, it can easily be verified that (6) holds and hence, the low type is indifferent between mimicking and revealing her type. Finally, it can be checked that (7) holds with strict inequality, and thus, the high type strictly prefers separating. As a result, a separating equilibrium always exists.\(^{25}\)

The separation mechanism described above is fairly simple as it involves only two costly signals, $\Delta$ and $T$. The amount $\Delta$ is the initial cost of separation (to reiterate, this can be interpreted as a signing bonus paid upfront to the agent), and $T$ refers to the delay in offering $C_h$. Although this separating equilibrium need not be optimal, its simplicity is instructive. In general, the key features of any separating equilibrium are (i) an initial transfer to the agent (or some form of money-burning); and (ii) delaying high-surplus contracts to a sufficiently distant future. These two deter the low type principal from imitating the high type in a

\(^{25}\)If $\delta_l < \delta < \delta_h$, then only a slight modification is needed in the construction above. The idea of the proof remains the same but the incentive compatibility constraints end up being slightly different than those presented above. This is because if $\delta_l < \delta$, then the symmetric-information contract of the low type depends on $\delta$, not on $\delta_l$. However, this difference is inessential for the construction above.
cost-efficient way due to the following:

- Delaying high-surplus contracts implies that imitating the high type and defaulting early on does not pay off given the initial cost of signaling $\Delta$.

- Waiting for high-surplus contracts (such as $C_h$) is costlier for the low type—recall that the cost of delay is monotonic in type, which allows me to obtain an inequality such as (5) above. If the delay is sufficiently long, then the benefit of separation for the high type exceeds the benefit of imitation for the low type in discounted terms. The initial costly signaling $\Delta$ is then simply set equal to the discounted benefit of imitation for the low type.

Thus, a separating equilibrium always exists: The low type is deterred from imitation, and the high type prefers revealing her type. This simple idea can be applied to any infinitely repeated game where the private information is on time preferences.

To reiterate, differential time preferences of types is essential for a general separation result. As I stated before, similar or stronger versions of Proposition 2 hold with other types of private information; that is, separation is not generally possible if the high type’s behavior is distorted only at $t = 0$. But such a result has broader implications with other types of private information because if the principal types have the same discount factor, then delaying better contracts is not effective as in my model. For example, it is not possible to come up with an inequality similar to (5) and ensure that there exists an equilibrium in which the benefit of separation for the high type exceeds the benefit of imitation for the low type. As a result, a separating equilibrium may not exist with other types of private information. As Halac (2012) has shown, there exists no separating contract if the principal types differ in their outside options. Consider next the case where the high type and the low type principals differ in (and are privately-informed regarding) their productivity and are identical in other respects. Let $y_\theta$ represent the production function of type-$\theta$ principal, where $\theta \in \{h, l\}$. If the high type principal has a productivity advantage of $\eta > 0$ over the low type such that $\eta \equiv y_h(e) - y_l(e) > 0$ given effort $e$, then a separating contract does not exist. Furthermore, it is not generally sufficient for separation if $y_h(e) > y_l(e)$ and $y_h(e) - y_l(e)$ is increasing in $e$. I formalize these claims in the Online Appendix. For an intuition, consider the simple case where $y_h(e) - y_l(e) = \eta > 0$ for every $e$. A separating equilibrium does not exist in this case because a strong version of Proposition 2 obtains: Separation is impossible if the high type’s behavior is distorted only at $t = 0$. Moreover, delaying high-surplus contracts is ineffective. It follows that a separating equilibrium does not exist, and there is no analogue of Proposition 3. These also hold in Halac (2012): Separation is impossible if the high type’s behavior is distorted only at $t = 0$, and delay is also ineffective. As a result, there exists no separating equilibrium.
In Section 7.1, I show that my analysis and results are applicable to a reputation model à la Mailath and Samuelson (2001). In particular, I show that a separating equilibrium always exists if the seller’s private information is about her time preferences, whereas a separating contract does not generally exist if the seller is privately informed regarding another parameter, such as her ability or cost of effort.

4.1 The Optimal Separating Contract

I define the optimal contract as the equilibrium contract that maximizes a weighted average of the two principal type’s equilibrium payoffs, where the weight for the high type’s payoff, denoted by $\gamma$, is arbitrary and $\gamma \in (0, 1)$. In the optimal separating contract, the low type principal always offers $C_l$. A separating equilibrium in which the low type offers a different contract is suboptimal since changing the low type’s contract offer to $C_l$ can never make the high type worse off but makes the low type strictly better off. Thus, the optimal separating contract boils down to the contract that maximizes the payoff of the high type principal among all separating contracts in which the low type’s contract offer is $C_l$. Formally, the optimal separating contract $\{w_t, b_t\}_{t=0}^\infty$ maximizes $\sum_{t=0}^\infty \delta^t (y(e_t) - w_t - b_t)$ subject to (1)-(4).\(^{26}\)

The first step towards characterizing the optimal separating contract is to show that the high type principal does not use future rewards (i.e., variations in continuation payoffs) as a discipline device. The reason for this is as follows. In order to motivate the agent, the high type principal can use either a bonus payment (i.e., an immediate reward) or a future reward (or both). However, using a future reward scheme is costlier than a bonus scheme because the agent is less patient than the high type principal, and a future reward is not legally enforceable just like a bonus reward. Thus, the high type principal strictly prefers using a bonus scheme in order to motivate the agent, and the agent is rewarded immediately conditional on performance.

**Lemma 4** In the optimal separating contract of the high type, $u_t = \frac{a}{1-\delta}$ for every $t > 0$.

The next lemma shows that costly signaling stops at a finite period in the optimal separating contract of the high type; that is, the high type offers $C_h$ at all sufficiently large $t$. This is because the optimal separating contract is such that a low type principal who chose to imitate the high type would strictly prefer defaulting at some finite period $T$. Otherwise, the high type’s optimal separating contract is a contract that can be implemented with a low type in a symmetric-information setting, which is a contradiction. This implies that the high type will start offering $C_h$ from a sufficiently large $t \geq T$ onwards, rather than distort her behavior indefinitely.

\(^{26}\)The optimal contract is also pareto-optimal in the usual sense; if there are two contracts that maximize this weighted average, then the one that gives the agent a higher payoff is optimal.
Lemma 5  Costly signaling stops at a finite date; i.e., the high type offers $C_h$ at all sufficiently large $t$.

Finally, I characterize the optimal separating contract. Proposition 6 shows that in the optimal separating contract of the high type, $b_t > b_l$ and $e_t > e_l$ at every $t \geq 0$. This implies that from the very beginning the relationship with a high type principal generates a higher surplus than the low type’s contract. In turn, the high type must pay a relatively high fixed wage at $t = 0$ so that the low type does not imitate. In other words, a high type principal offers an initial lump-sum signing bonus and provides stronger performance incentives than a low type principal in every period. Moreover, $b_t$ is strictly increasing until it reaches $b_h$, and $b_t = c(e_t)$ at every $t \geq 0$ in the optimal contract. Thus, the strength of the performance incentives increases over time, and the percentage of the agent compensation that comes from performance bonus increases progressively. As incentives become stronger, the surplus increases, and the relationship becomes more valuable over time. This “gradualism” result is reminiscent of the results obtained in Sobel (1985), Diamond (1989), Ghosh and Ray (1996), Kranton (1996), Watson (1999; 2002), and Halac (2012). However, I obtain gradualism in honest separating equilibria, which is a novel result. As it turns out, a gradualist pattern of cooperation is optimal even in settings where learning is immediate because separation is cheaper for the high type if costly signaling is spread over time in a gradualist manner.

Proposition 6  In the optimal separating contract of the high type, $b_t < b_{t+1}$ for every $t \geq 0$ until $b_t = b_h$, which takes place in finite time. Similarly, the effort schedule and the surplus are strictly increasing until they reach $e_h$ and $s_h$, respectively. Moreover, $b_t > b_l$ and $s_t > s_l$ at every $t \geq 0$. Finally, $w_0 > \bar{u}$ and $w_t = \bar{u}$ at $t > 0$.

The intuition for Proposition 6 is as follows. Offering a lump-sum signing bonus, delaying the contract offer $C_h$, and using relatively weak incentives at the early stages of the relationship are very effective as costly signals. These generate low surplus and low principal profits early in the relationship, and thus a low imitation payoff because the low type is relatively impatient. As the relationship builds up over time, incentives become stronger, and the profit level increases progressively; such an arrangement is optimal because the initial signing bonus together with the gradually increasing profit levels deter an impatient low type from imitation in the least costly way.

To make the intuition regarding the monotonicity of the bonus schedule more transparent, consider a scenario in which $T = 2$ according to the construction that I used in order to prove Proposition 3. Since $T = 2$, the high type principal offers $b_t$ in the first two periods followed by $b_h$, thereafter. The corresponding effort levels are $e_0 = e_1 = e_l$, and $e_t = e_h$ for $t \geq 2$. The high type can do better than this contract by using a strictly increasing bonus schedule. She can (i) change the bonus at $t = 1$ so that $b_t' = b_t + \epsilon$ instead of $b_t$, where $\epsilon > 0$.  

19
is arbitrarily small; (ii) make $b_l + \epsilon$ contingent on $y(e'_1)$ such that $c(e'_1) = b_l + \epsilon$ (thus, $e'_1 > e_l$); and (iii) increase the initial fixed-wage $w_0$ by $\delta_l \left[ s(e'_1) - s(e_l) \right]$. Since $\epsilon$ is arbitrarily small, a low type who chose to imitate a high type would still optimally default at $T = 2$. Thus, the changes in $e_1$ and $w_0$ are such that the low type’s imitation payoff is exactly the same as before. However, by virtue of her higher discount factor, the high type is strictly better off. This illustrates the key mechanism behind Proposition 6. If a separating contract involves bonuses such that $b_{t+1} \leq b_t < b_h$ for some $t$, then one can always construct a separating contract that strictly dominates it.

5 Other contracts

In this section, I investigate the remaining types of equilibria, namely, pooling equilibria and hybrid equilibria.\footnote{As discussed in Footnote 14, there can be equilibria which are separating after an initial pooling phase, but such contracts are not relevant for my main results because they are inefficient as implied by Proposition 7. Finally, there is one more type of contract that is possible: a contract which is hybrid with probability $\alpha \geq 0$, separating with probability $\beta \geq 0$, and pooling with probability $(1 - \alpha - \beta)$ at $t = 0$. However, none of the upcoming results are affected by this possibility.} There is no information revelation in a pooling equilibrium. In a hybrid equilibrium, both types start by offering the same contract, and information is revealed over time through default (or the absence thereof). To be more precise, the high type honors implicit bonus or future contract promises in the hybrid contract, whereas the low type eventually reneges on a promise.

The optimal pooling equilibrium is stationary and implements $C_l$ in every period. To see why, first note that in every period of a pooling equilibrium, the principal types must honor the bonus promise or default with identical probability so that there is no revelation of information regarding the principal’s type. However, a pooling equilibrium that involves default in some period is inefficient. Therefore, the principal always fulfills her promises in the optimal pooling contract, regardless of her type. Put differently, the optimal pooling contract is enforceable with either type in a symmetric-information setting.\footnote{In a pooling equilibrium which involves default (that is, both types default in some period with identical probability), incentive provision must involve a default risk premium so that the agent is willing to participate and exert effort despite the default risk. But this is entirely wasteful just as it would be in a symmetric-information setting: Given an equilibrium in which the two types default in some period with the same probability, there exists a strictly better pooling equilibrium in which both types honor their promises with probability one, the output implemented is always the same as in the original pooling contract, and the agent compensation is strictly lower because there is no need to pay the default risk premium.} But a contract which is enforceable with the low type principal is surely enforceable with the high type principal since the high type is more patient. Thus, the optimal pooling equilibrium implements $C_l$ in every period because this is the best contract which is enforceable with the low type.

In the optimal separating equilibrium, the low type principal always offers $C_l$, whereas
the high type principal strictly prefers revealing her type to imitating the low type as the equilibrium construction in the proof of Proposition 3 indicates. Hence, the following result obtains.

**Proposition 7** *The optimal separating equilibrium strictly dominates pooling equilibria.*

Next, I analyze the optimal hybrid contract. One factor that complicates the analysis is that the low type principal defaults on the equilibrium path, unlike the case in the separating contract. Thus, the form of the punishment is an important feature of the equilibrium. One useful observation is that after the posterior belief becomes degenerate, imposing the worst punishment on a principal who defaults is without loss of generality. Put differently, if there exists a hybrid equilibrium such that a default that takes place after the posterior belief has become degenerate is not punished in the worst possible way, then there exists a payoff-equivalent hybrid equilibrium which imposes the worst punishment instead.

While it is possible to obtain some intermediate results regarding the optimal hybrid contract without making any assumption about the form of the punishment, my main interest lies in comparing different types of equilibria, and this requires some restriction on punishments given the complexity of the problem. I impose Assumption 1 (A1) in order to be able to obtain a precise characterization of the optimal hybrid contract and compare efficiency under different types of contracts. After I present assumption A1, I discuss in detail this assumption and the conditions under which it is optimal (or at least plausible) and relate it to the recent literature.

**Assumption 1 (A1)** *If the principal defaults on a payment promise, then the worst punishment is imposed and the agent terminates the relationship.*

As discussed above, A1 is without loss of generality after the posterior belief becomes degenerate. With nondegenerate posteriors, A1 may in principal be restrictive. However, there are various conditions under which A1 is optimal even with nondegenerate beliefs. For example, if \( C_h \) is sufficiently far from the first-best contract (that is, if \( y'(e_h)/c'(e_h) \) is sufficiently larger than 1), then it is optimal to impose the worst punishment following a default. Imposing the worst punishment is also optimal if \( \delta_l \) and \( \delta_h \) are not far from each other or if \( \mu_0 \) is sufficiently high.\(^{29}\) A1′ below is one possible alternative for A1.

**Assumption 1′ (A1′)** *Parameters are such that \( y'(e_h)/c'(e_h) \) is sufficiently larger than 1, and therefore, imposing the worst punishment following a default is optimal.*

---

\(^{29}\)I prove in the Appendix that if \( y'(e)/c'(e) \) is sufficiently far from 1, then the optimal contract implements the worst punishment following default. I also prove that the same is true if \( \delta_l \) and \( \delta_h \) are not far from each other or if \( \mu_0 \) is sufficiently high. Note that, in order to obtain sharper results regarding the optimality of A1, I disregard certain hybrid equilibria that are strictly dominated regardless of the form of the punishment; this is without loss of generality as it has no consequence for the results stated in Proposition 8. Details are relegated to the Appendix (see in particular the discussion preceding Claim 6, on page 49).
Apart from optimality, $A_1$ may depict those cases where defaulting on payment promises may require the firm to shut down the business or change its name and location. This is reportedly a common form of wage violation.\footnote{http://www.labor.ucla.edu/wage-theft/} Moreover, adopting an assumption such as $A_1$ or $A_1'$ has a precedent in the literature. Halac (2012) focuses on parameters such that imposing the worst punishment after default is optimal, and this allows for characterizing the evolution of equilibrium bonuses and effort levels in the optimal hybrid contract.

While the optimal separating equilibrium strictly dominates the optimal pooling equilibrium, the comparison of the optimal separating equilibrium and the optimal hybrid equilibrium depends on the magnitude of $\mu_0$ and $\gamma$ as Proposition 8 shows.

**Proposition 8** Assume $A_1$ or $A_1'$, and fix $\gamma \in (0, 1)$.

(i) There exists an $\varepsilon > 0$ such that if $\mu_0 \leq \frac{b_l}{b_h} + \varepsilon$, then hybrid contracts are strictly dominated, whereas if $\mu_0$ is sufficiently close to 1, then separating contracts are strictly dominated. More generally, given $\gamma \in (0, 1)$ there exists a $\mu_{\gamma} > \frac{b_l}{b_h}$ such that the optimal contract is separating if $\mu_0 \leq \mu_{\gamma}$, and hybrid otherwise. Moreover, $\mu_{\gamma}$ is increasing in $\gamma$.\footnote{Note that it is possible to set $\gamma = \mu_0$. Then, the result would be as follows: There exists a $\tilde{\mu}_0 > \frac{b_l}{b_h}$ such that the optimal contract is hybrid if $\mu_0 > \tilde{\mu}_0$ and separating otherwise.}

(ii) If the optimal contract is hybrid, then it fully reveals information in finite time and exhibits gradualism; that is, there exists a $T < \infty$ such that as long as the principal honors the promised payments, $b_t$ and $s_t$ are strictly increasing until period $T$ when they reach $b_h$ and $s_h$, respectively.

For an intuition of part (i), first note that if the prior probability $\mu_0$ is sufficiently high and close to 1, then the optimal contract is hybrid regardless of $\gamma$ because the chances that the principal is a low type is sufficiently small, and it is possible to construct a hybrid equilibrium that approximates $C_h$ at $t = 0$ and is identical to $C_h$ at every $t > 0$ provided that the bonus promise is honored at every $\tau < t$. Such an equilibrium makes both types strictly better off than they would be in a separating equilibrium. If however $\mu_0$ is sufficiently low (for example, lower than $b_l/b_h$) then hybrid equilibria are wasteful regardless of $\gamma$ because equilibrium information revelation involves a relatively high probability of default for several periods, and thus, incentivizing the agent is too costly relative to other types of equilibria. To be more precise, incentive provision in a hybrid equilibrium requires insuring the agent against default to some extent so that the agent is still willing to participate and exert effort, and this requirement becomes too costly when the prior is relatively low and the probability of default is relatively high. Note that the lower bound on $\mu_0$ stated in Proposition 8 may be quite stringent. For example, if $\delta_l$ and $\delta_h$ are close, then $b_l$ and $b_h$ are also close, and hybrid equilibria are inefficient unless the prior $\mu_0$ is close to 1.

If $\mu_0$ is an intermediate value bounded away from both $b_l/b_h$ and 1, then whether the
optimal contract is hybrid or separating depends on both the magnitude of $\mu_0$ and the welfare weights of the principal types. While the payoff of the optimal separating equilibrium does not depend on $\mu_0$, the payoff of the optimal hybrid contract is strictly increasing in $\mu_0$. Therefore, given fixed $\gamma \in (0, 1)$ there exists a unique threshold value of $\mu_0$ such that the payoff of the optimal separating equilibrium and the payoff of the optimal hybrid equilibrium are identical. Hence, the optimal contract is hybrid if the prior exceeds this threshold and separating otherwise. This threshold value, denoted by $\mu_0$, is monotone increasing in $\gamma$ since the high type prefers separating equilibria over hybrid equilibria for a wider range of $\mu_0$ values than the low type.

I now focus on part (ii) and discuss the structure of the optimal hybrid equilibrium assuming that the optimal contract is hybrid. The low type is initially indifferent between defaulting and honoring the bonus promise but eventually defaults at some $t \geq 0$ with probability one. Once the principal defaults, the agent learns that the principal is low type with probability one. As long as the principal keeps honoring her bonus promises, trust is gradually established, higher bonus promises become more credible, and the surplus of the relationship increases progressively until it becomes stationary at the high type’s symmetric-information surplus level.

6 Equilibrium Selection: Dynamic Intuitive Criterion

As discussed in the previous section, hybrid equilibria are dominated unless $\mu_0$ is above a certain threshold. I now appeal to a dynamic version of the Intuitive Criterion and show that most hybrid equilibria are “unreasonable.”

In what follows, let $\{C_t\}_{t=0}^\infty$ denote an equilibrium set of contracts. Assume that the principal deviates and announces an out-of-equilibrium set of contracts $\{D_t\}_{t\geq k}$ at an arbitrary period $k \geq 0$. $\{D_t\}_{t\geq k}$ is defined to be equilibrium-dominated for type-$\theta$ if the equilibrium payoff of a type-$\theta$ principal is strictly higher than the highest possible payoff she gets if she honors all the payments until period $k$ and deviates to $\{D_t\}_{t\geq k}$ (the highest possible payoff may involve default, for example reneging on a bonus promise in some $D_t$).\(^{32}\)

I define the Dynamic Intuitive Criterion as follows. An equilibrium $\{C_t\}_{t=0}^\infty$ fails to satisfy the Dynamic Intuitive Criterion (DIC) if there exists a $k \geq 0$ and an out-of-equilibrium set of contracts $\{D_t\}_{t\geq k}$ such that:

(i) $\{D_t\}_{t\geq k}$ is equilibrium-dominated for the low type principal, and
(ii) $\{D_t\}_{t\geq k}$ is enforceable for the high type (i.e., the high type would never default) and strictly profitable if the agent best-responds according to the belief $\mu_t = 1$ at every $t \geq k$.

\(^{32}\)For $\{D_t\}_{t\geq k}$ to be meaningful as out-of-equilibrium behavior, the game must have proceeded to period $k \geq 0$ without default. Otherwise, the agent is certain that the principal is low type.
DIC eliminates every separating equilibrium other than the optimal separating equilibrium; the optimal separating equilibrium is always robust to DIC. DIC also eliminates every pooling or hybrid equilibrium—with only one possible exception. There is only one hybrid equilibrium that might be immune to DIC, and this is the one in which the low type defaults with probability one at \( t = 0 \), \( b_0 \) in the period-0 hybrid contract is such that

\[
b_0 = \delta_l \left( s_h - \bar{u} + b_h + \frac{\delta_l}{1 - \delta_l} \bar{u} \right),
\]

and the high type offers \( C_h \) from \( t = 1 \) onwards (the low type defaults with probability one at \( t = 0 \)). Note that such an equilibrium can be robust to DIC only if \( \mu_0 > \frac{b_h}{b_0} \). Also, note that this type of lower bound is more stringent than \( \frac{b_h}{b_0} \), the lower bound discussed in Proposition 8, because \( b_0 < b_h \). However, even when \( \mu_0 > \frac{b_h}{b_0} \) holds, it is still not obvious whether this equilibrium is robust to DIC; the high type may find a deviation \( \{D_t\}_{t \geq 0} \) that is equilibrium dominated for the low type and makes her better off unless \( \mu_0 \) is sufficiently close to 1. An extensive discussion regarding the Dynamic Intuitive Criterion and the results of this part are relegated to the Online Appendix.

7 Extensions

7.1 Buyer-Seller Relationships

In this section, I analyze an infinitely repeated game à la Mailath and Samuelson (2001) and show that my main results regarding separating contracts extend to other settings. In this model, a long-lived seller sells one product to a long-lived buyer or to a sequence of short-lived buyers in each period \( t = 0, 1, \ldots \) The quality of the product for sale, denoted by \( q \), is either low (i.e., \( q = q_L \)), or medium (i.e., \( q = q_M \)), or high (i.e., \( q = q_H \)).\footnote{I extend the setting in Mailath and Samuelson (2001) and allow for more than two quality levels. I also assume that both seller types can be strategic, as in my main model.} The terms \( u_H \), \( u_M \) and \( u_L \) denote the utility from consuming a product with \( q = q_H \), \( q = q_M \), and \( q = q_L \), respectively. I assume that \( u_H > u_M > u_L \geq 0 \).

In each period, \( q \) is stochastically determined according to the effort the seller chooses. There are three possible effort levels: low effort (i.e., \( e = L \)), medium effort (i.e., \( e = M \)) and high effort (i.e., \( e = H \)). The disutility of effort increases in the effort level. I normalize the cost of \( L \) to 0. Since higher effort levels are costlier, \( c_H > c_M \), where \( c_M \) and \( c_H \) denote the cost of \( M \) and \( H \), respectively. If the seller chooses \( e = H \), then the product is high-quality...
(q = q_H) with probability Φ and medium quality (q = q_M) with probability (1 − Φ):

\[ q_{e=H} = \begin{cases} q_H & \text{with probability } \Phi \\ q_M & \text{with probability } 1 - \Phi \end{cases} \]

If the seller chooses e = M, then the product is medium-quality (q = q_M) with probability Φ and low quality (q = q_L) with probability (1 − Φ):

\[ q_{e=M} = \begin{cases} q_M & \text{with probability } \Phi \\ q_L & \text{with probability } 1 - \Phi \end{cases} \]

Finally, if the seller chooses e = L, then the product is low-quality (q = q_L) with probability one. High effort is socially efficient whereas low effort generates the lowest social surplus; i.e.,

\[ \Phi u_H + (1 - \Phi)u_M - c_H > \Phi u_M + (1 - \Phi)u_L - c_M > u_L. \]

I now assume that there are two seller types. I consider three scenarios. In the first scenario, I consider a “good” type (type-g) and a “bad” type (type-b) such that type-g is more able and has a lower cost of effort than type-b; i.e., \( c^b_e > c^g_e \), where \( c^i_e \) denotes the cost of effort \( e \in \{M, H\} \) for type-\( i \) seller, \( i \in \{g, b\} \). In the second scenario, type-g has a higher Φ value than type-b. Finally, in the third scenario type-g has a higher discount factor than type-b. In the first two scenarios, I assume that types have a common discount factor \( δ \).

I focus on parameter values such that type-g has an incentive to separate himself from type-b in order to avoid uninteresting cases (equivalently, type-b has an incentive to imitate type-g). In particular, I focus on the case where type-g can be induced to exert high effort in a symmetric-information setting whereas type-b can only be induced to exert medium effort (thus, type-g is similar to the high type principal and type-b is similar to the low type principal in my main model).34

A seller may choose to burn money (via, for example, advertising) in order to signal his type. It is reasonable to assume that the cost of such money-burning is identical for the two types.

34 As discussed in Section 2.2, I analyze this model in a relational setting. As a result, the seller may be motivated to exert effort in a symmetric-information setting. If the seller can observe the quality realization before sale, then the seller may send a message regarding the realized quality of the product. Similar to the bonus promise of a principal, the message regarding the quality is only cheap-talk and nonbinding; however, it can be truthful and credible (just like a self-enforcing bonus scheme in a relational contract) provided that the future profit of the seller from “honest” trade is high enough. See the details in the Online Appendix. If the seller can observe the quality realization only after the sale, then the seller may accept to give a refund to the buyer in case the product quality is observed to be lower than some benchmark level. Again, this can be sustained in equilibrium if the future profit from honest trade is sufficiently high.
I first show that a separating equilibrium does not exist in the first case if $c_H^b - c_H^0 = c_M^b - c_M^0$ or $c_H^b - c_H^0 < c_M^b - c_M^0$ holds. Intuitively, this is because in either case, the benefit of separation for the high type is strictly lower than the benefit of imitation for the low type. Thus, equilibrium information revelation must involve dishonest behavior by type-$b$ seller if $c_H^b - c_H^0 < c_M^b - c_M^0$. Moreover, I show that the condition $c_H^b - c_H^0 > c_M^b - c_M^0$ is not sufficient for the existence of a separating equilibrium. Details of this model and the analysis can be found in the Online Appendix. In a similar vein, separation is not generally possible in the second case, where type-$g$ has a higher value than type-$b$.

However, a separating equilibrium always exists in the third case where seller types differ in their time preferences. The construction in the proof of Proposition 3 can be directly applied to show this result (see the details in the Online Appendix). Thus, there exist equilibria which reveal information fully and without breach; consumers know the type of the seller and are never deceived about the quality of the product that they purchase in a separating equilibrium. Other main results also apply in the third case; the optimal separating equilibrium is gradualist (i.e., it involves rising quality over time and increasing profits for the informed seller that is a good-type) and hybrid equilibria are inefficient unless the prior probability that the seller is a good type is sufficiently high.

Finally, my results extend to the market setting where sellers enter and exit the economy stochastically, and names can be traded, as modeled in Tadelis (1999) and Mailath and Samuelson (2001). In addition to the trading of names, I allow for name changes; for example, an existing type-$b$ firm that has a bad reputation can try to erase the public memory about his type by choosing a new name. I also maintain the assumption in Tadelis (1999) and Mailath and Samuelson (2001) that changes in names’ ownership are unobservable. I show that my main result holds also in this setting provided that type-$g$ sellers have a sufficiently high discount factor. In particular, there exists a separating equilibrium such that a good name never goes bad because good names are too expensive for bad sellers. This result contrasts with Tadelis (1999) and Mailath and Samuelson (2001) in which the seller types have identical discount factors.

### 7.2 Multiple Types

If there are more than two principal types, there still exists a separating equilibrium—the mechanism in Proposition 3 can be extended to show the existence of a separating equilibrium with multiple types. The precise characterization of the optimal separating contract is tedious, but I conjecture that for the better types, the monotonicity of the bonus schedule and the effort schedule is preserved.

---

35To be more precise, if $q_H - q_M \leq q_M - q_L$, then a separating equilibrium does not exist. Moreover, the condition $q_H - q_M > q_M - q_L$ is not sufficient for separation.
In the presence of multiple types, hybrid equilibria with full information revelation will be costlier. As the number of types increases, the number of types with an incentive to imitate naturally rises. As a result, the measure of the best type typically decreases. This in turn restricts the surplus from hybrid equilibria that fully reveal information because the uninformed party will anticipate default with a high probability for a lengthy period and behave accordingly. As a result, the optimal contract is more likely to be separating in the presence of multiple types.

### 7.3 A Behavioral Approach

As discussed before, individuals may be sensitive to downside deviations from their expectations and incur a psychological cost when they receive less than what they expect or are entitled to. This is the main theme in models of disappointment aversion and loss aversion, such as Gul (1991), and Köszegi and Rabin (2006, 2007). This idea has also been incorporated into contract theory in several papers, such as Hart and Moore (2008) and Herweg et al. (2010).

Disappointment aversion and loss aversion do not affect separating contracts since separating contracts are honest. However, hybrid contracts become costlier; in particular, the set of priors with which the optimal contract is hybrid become smaller. This is because hybrid equilibria involve breach with positive probability, and the agent faces a lottery in each period of a hybrid contract: Until information is fully revealed, there is a risk that the agent is not compensated for his effort, which entails a psychological cost. Let $\lambda_t$ denote the probability that the bonus is paid at $t$ in a hybrid equilibrium—from the perspective of the agent—provided that there was no default until $t$. In the optimal hybrid contract with standard, risk-neutral agent preferences, $w_t$, $\lambda_t$, $b_t$ and $e_t$ are such that if $\mu_t^1 \in (0, 1)$, then $\lambda_t < 1$ and $w_t + \lambda_t b_t - c(e_t) = \bar{u}$ for at least some $t \geq 0$. I now adopt a simple formulation of disappointment-averse preferences and assume that disappointment-averse expected utility of the agent for time $t$ is given by

$$w_t + \lambda_t b_t - c(e_t) - \theta (1 - \lambda_t) \lambda_t b_t$$

where $\theta$ is the disappointment parameter, and material utility is linear for simplicity. This formulation is now standard in the literature with two-outcome games or lotteries. With probability $1 - \lambda_t$ the agent does not receive $b_t$ and suffers a disutility of $\theta \lambda_t b_t$ since the material utility falls below the expected material utility by $\lambda_t b_t$. If $w_t + \lambda_t b_t - c(e_t) = \bar{u}$, then the agent will not agree to participate as $\theta > 0$. Let $\theta = 1$ for simplicity. If the bonus

---

36Hybrid equilibria with full information revelation require all principal types but the best type to default.

37See Gill and Prowse (2012) and the discussion therein.
is \( \frac{b_t}{\lambda_t} \) in every \( t \geq 0 \) instead of \( b_t \), and everything else is kept the same in the contract, then the agent’s per period expected utility is \( \bar{u} \) and the agent agrees to participate, as before. What about agent’s effort incentives? Conditional on the agent’s participation and the same \( \lambda_t \) as before, the required bonus \( b'_t \) for the agent to still exert effort \( e_t \) is such that \( c(e_t) = \lambda_t (b'_t - (1 - \lambda_t)b'_t) = \lambda_t b_t \) holds. Thus, if \( b'_t = \frac{b_t}{\lambda_t} \), the agent will agree to participate and exert the same effort as before, but the contract is now costlier as bonus payments must be higher. Put differently, the optimal hybrid contract is no longer feasible, and the profit of the principal (of either type) will be lower. Obviously, increasing \( \lambda_t \) cannot alleviate the problem: If \( \lambda_t \) is increased, the bonus will still have to be higher than \( b_t \) (assuming that \( \lambda_t < 1 \) still holds) and \( \lambda_{t+k} \) will have to be lower for some \( k > 0 \) in order to make up for the increase in \( \lambda_t \). But then, \( b'_{t+k} \) must increase (i.e., \( b'_t > \frac{b_{t+k}}{\lambda_{t+k}} \)) in order to compensate the agent for betrayal risk. To conclude, disappointment-aversion extends the set of priors such that hybrid equilibria are inefficient.

7.4 Stochastic Output

Many of my main results in Sections 4 and 5 go through in a richer model where the agent’s unobservable effort gives rise to a stochastic output. More specifically, assume that the agent’s effort in period \( t \) generates stochastic output \( y_t \), where \( y_t \) is either “high” or “low” (i.e., \( y_t \in \{L, H\} \)). The probability that \( y_t = H \) is equal to \( e_t \). In this case, Propositions 1-3, 7, and Lemmas 4 and 5 go through without additional assumptions. A version of Proposition 6 also holds; however, it requires assuming that the bonus reward is contingent on calendar time, the current output and the agent’s beliefs, but not on the whole history of output. This constraint does not seem to be extreme, thanks to (a version of) Lemma 4: A history-dependent bonus scheme is never used to motivate the agent in the optimal contract. Such a scheme could still be the efficient arrangement in order to deter the low type principal from imitating, I cannot rule out this possibility. Yet, history-dependent bonus schemes are very complicated, and therefore, I focus on a constrained set of contracts in which the bonus reward is contingent on calendar time, the current output and the agent’s beliefs. Let \( b_t^* \) denote the bonus at \( t \) in the constrained optimal contract. Then, Proposition 8 follows.

Proposition 8 Assume that \( ec''(e) \) is weakly increasing. In the constrained optimal separating contract of the high type principal, \( b'_t < b'_{t+1} \) for every \( t \geq 0 \) until \( b'_t = b_h \), which takes place in finite time. Similarly, the effort schedule is strictly increasing until it reaches \( e_h \). Finally, the fixed wage is strictly decreasing over time until it reaches \( w_h \).

The proof of Proposition 9 is relegated to the online Appendix.\textsuperscript{38} Finally, note that a version of Proposition 8 also holds; however, it requires additional assumptions, such as imposing a

\textsuperscript{38}The rest of the proofs for this section are available upon request. They are omitted since they are similar to the current proofs (albeit, more tedious).
Markovian restriction on the principal’s strategy space.

8 Conclusion

This paper sheds light on the dynamics of cooperation and information revelation in repeated games with private information and informal incentives. In the previous reputation literature, information revelation has typically involved dishonesty and breach of trust. The aim of this paper is to unify the analysis of information revelation mechanisms and identify when breach of trust is necessary for information revelation and when it is not necessary. My results indicate that costly signals, such as signing bonuses, advertising, hiring reputable executives, are effective, and a separating equilibrium always exists if the private information is on time preferences; thus, breach is not necessary for information revelation in this case. In fact, this type of private information is essential for a “general” separation result; with other types of private information, separation via costly signaling is not generally possible. Finally, note that if the “good” type expects a higher future payoff than the “bad” type (as is often the case), then the good type can “endogenously” be more patient than the bad type in view of the Becker-Mulligan theory of discounting. But if this holds, then a separating equilibrium exists with any type of private information.

REFERENCES


A Appendix

Proof of Proposition 1. First, note that $u_0 = \bar{u} = \frac{\bar{u}}{1-\delta}$ holds in the optimal contract. Otherwise, a small reduction in $w_0$ is strictly profitable for the principal because this change doesn’t affect the participation and the incentive constraints of the agent in any period. Next, assume that $\delta_\theta > \delta$. Let $\{w_t, b_t\}_{t=0}^{\infty}$ denote the optimal equilibrium set of contracts, and let $e_t$ denote the effort implemented at $t$ in this equilibrium. Assume towards a contradiction that the optimal contract specifies $u_{t+1} > \frac{\bar{u}}{1-\delta}$ on-the-equilibrium path for some $t \geq 0$. Consider the following changes in the contract at $t$ and $t+1$: $b_t$ is increased by a small amount $\delta \varepsilon$, whereas $w_{t+1}$ is decreased by $\varepsilon$ (this implies that $u_{t+1}$ decreases and $\pi_{\theta,t+1}$ increases by $\varepsilon$). The bonus reward $b_t + \delta \varepsilon$ is still contingent on the effort level $e_t$ (that is, output level $y(e_t)$) as in the original contract. This modified contract strictly increases the payoff of the principal, and the agent is unaffected. To see why, first consider the case in which $b_t \geq 0$. The increase in $b_t$ coupled with the increase in $\pi_{\theta,t+1}$ is enforceable since

$$b_t + \frac{\delta_\theta}{1-\delta_\theta} \bar{\pi} \leq \delta_\theta \pi_{\theta,t+1}$$

and $\delta_\theta > \delta$ imply that

$$b_t + \delta \varepsilon + \frac{\delta_\theta}{1-\delta_\theta} \bar{\pi} < \delta_\theta (\pi_{\theta,t+1} + \varepsilon).$$

Thus, $b_t + \delta \varepsilon$ is enforceable. Moreover, the agent’s participation constraint is satisfied at every $t$ (since $\varepsilon$ is small), and the agent’s incentive-compatibility constraint for choosing effort level $e_t$ is still satisfied by construction; that is,

$$b_t + \delta \varepsilon - c(e_t) + \delta (u_{t+1} - \varepsilon) \geq \frac{\delta}{1-\delta} \bar{u}.$$ 

The same is also true for $e_{t+1}$ and the subsequent effort levels since only $w_{t+1}$ is reduced. Thus, the principal gains $(\delta_\theta - \delta) \varepsilon > 0$, a contradiction. Second, consider the case in which $b_t < 0$. Again, increasing $b_t$ by a small amount $\delta \varepsilon$ and decreasing $w_{t+1}$ by $\varepsilon$ while keeping the effort requirement the same satisfies the participation constraint of the agent at every $t$ (since $\varepsilon$ is small). Moreover, these changes satisfy the enforcement and the incentive-compatibility
constraints of the agent since

\[ b_t - c(e_t) + \delta u_{t+1} \geq \frac{\delta}{1-\delta} \bar{u} \]

implies that

\[ (b_t + \delta \varepsilon) - c(e_t) + \delta (u_{t+1} - \varepsilon) \geq \frac{\delta}{1-\delta} \bar{u} . \]

This ensures that the agent optimally chooses effort level \( e_t \), and the enforcement constraint is satisfied as well. As a result, the principal gains \((\delta_\theta - \delta)\varepsilon > 0\). Hence, \( u_t = \frac{\bar{u}}{1-\delta} \) must hold in the optimal contract at every \( t \geq 0 \) if \( \delta_\theta > \delta \).

Now, assume that \( \delta_\theta < \delta \). Assume towards a contradiction that \( \pi_{\theta,t+1} > \frac{\delta_\theta}{1-\delta_\theta} \bar{\pi} \) for some \( t \geq 0 \). First, consider the case in which \( b_t > 0 \). Consider the following changes in the contract: \( b_t \) is decreased by a small amount \( \delta \varepsilon \), whereas \( u_{t+1} \) is increased by \( \varepsilon \) (this implies that \( u_{t+1} \) increases and \( \pi_{\theta,t+1} \) decreases by \( \varepsilon \)). The bonus reward \( b_t - \delta \varepsilon \) is contingent on the effort level \( e_t \) as in the original contract. The decrease in \( b_t \) coupled with the decrease in \( \pi_{\theta,t+1} \) is enforceable since

\[ b_t + \frac{\delta_\theta}{1-\delta_\theta} \bar{\pi} \leq \delta_\theta \pi_{\theta,t+1} \]

and \( \delta_\theta < \delta \) imply that

\[ b_t - \delta \varepsilon + \frac{\delta_\theta}{1-\delta_\theta} \bar{\pi} < \delta_\theta (\pi_{\theta,t+1} - \varepsilon) . \]

Thus, \( b_t - \delta \varepsilon \) is enforceable. Moreover, the agent’s participation constraint is still satisfied at every \( t \) (since \( \varepsilon \) is small), and the agent’s incentive-compatibility constraint for choosing effort level \( e_t \) is still satisfied by construction. It follows that the principal gains \((\delta - \delta_\theta)\varepsilon > 0\), a contradiction. Thus, \( \pi_{\theta,t} = \frac{\delta_\theta}{1-\delta_\theta} \bar{\pi} \) for every \( t \geq 0 \). The proof for the case in which \( b_t \leq 0 \) is similar and therefore omitted.

**Proof of Proposition 2.** I provide the proof only for the case in which \( \delta_h > \delta_l > \delta \). The proof for the case in which \( \delta_h > \delta \geq \delta_l \) is almost identical, and, therefore, omitted. Now, assume that the two types separate offering different contracts at \( t = 0 \), and that the continuation play following separation consists of the optimal symmetric-information contract. Let \( C_0 = \{w_0,b_0\} \) be the contract offer of the high type at \( t = 0 \), where the bonus reward \( b_0 \) is contingent on the effort level \( e_0 \) (that is, output level \( y(e_0) \)). Of course, \( w_0 + b_0 - c(e_0) \geq \bar{u} \) must hold.\(^{39}\) Note that the agent must accept the offer \( C_0 \); it can easily be checked that a contract offer at \( t = 0 \) which the agent rejects cannot be part of this separating equilibrium given that the continuation play consists of \( C_h \). Given these, the

\(^{39}\) I assume without loss of generality that the high type always honors \( b_0 \), and that \( w_0 + b_0 - c(e_0) \geq \bar{u} \). Refusing to pay \( b_0 \) or offering a contract such that \( w_0 + b_0 - c(e_0) < \bar{u} \) cannot be credible costly signals for the high type.
relevant incentive compatibility constraints boil down to

\[ y(e_0) - w_0 - b_0 + \frac{\delta_h}{1 - \delta_h} (s_h - \bar{u}) \geq \frac{s_l - \bar{u}}{1 - \delta_l} \]  

(8)

for the high type and

\[ \frac{s_l - \bar{u}}{1 - \delta_l} \geq y(e_0) - w_0 - b_0 + \max\{\delta_l V_l, b_0 + \frac{\delta_l}{1 - \delta_l} \tilde{\pi}\} \]  

(9)

for the low type, where

\[ V_l = s_h - \bar{u} + b_h + \frac{\delta_l}{1 - \delta_l} \tilde{\pi}. \]

The term \( V_l \) denotes the continuation payoff for an imitator low type principal at \( t \geq 1 \) given that she has honored \( b_0 \) and the continuation play is the symmetric-information contract of the high type. Note that (9) follows because an imitator reneges either on \( b_0 \) at \( t = 0 \), or on \( b_h \) at \( t = 1 \) depending on the magnitude of \( b_0 \) relative to \( b_h \). Combining inequalities (8) and (9) implies that

\[ \frac{\delta_h}{1 - \delta_h} (s_h - \bar{u}) + \frac{\delta_l}{1 - \delta_l} (s_l - \bar{u}) \geq \frac{\delta_h}{1 - \delta_h} (s_l - \bar{u}) + \delta_l V_l \]  

(10)

is a necessary condition for separation. Recall that

\[ b_h = \frac{\delta_h}{1 - \delta_h} (s_h - \bar{u} - \bar{\pi}), \]

due to the binding enforcement constraint of the high type. Thus,

\[ V_l = s_h - \bar{u} + b_h + \frac{\delta_h - \delta_l}{(1 - \delta_h)(1 - \delta_l)} \tilde{\pi}. \]

It follows that

\[ \delta_l V_l - \frac{\delta_h}{1 - \delta_h} (s_h - \bar{u}) = \frac{\delta_l - \delta_h}{1 - \delta_h} \left[ (s_h - \bar{u}) + \delta_l \frac{\tilde{\pi}}{1 - \delta_l} \right]. \]

As a result,

\[ \left( \delta_l V_l - \frac{\delta_h (s_h - \bar{u})}{1 - \delta_h} \right) + \left( \frac{\delta_h}{1 - \delta_h} - \frac{\delta_l}{1 - \delta_l} \right) (s_l - \bar{u}) = \frac{\delta_h - \delta_l}{1 - \delta_h} \left[ \frac{s_l - \bar{u}}{1 - \delta_l} - (s_h - \bar{u}) - \delta_l \frac{\tilde{\pi}}{1 - \delta_l} \right] \]

\[ = \frac{\delta_h - \delta_l}{1 - \delta_h} \left[ \frac{s_l - \bar{u} - \tilde{\pi}}{1 - \delta_l} - (s_h - \bar{u} - \tilde{\pi}) \right]. \]

The term inside the square brackets on the right hand side of the equality is strictly positive if \( s_l - \bar{u} - \tilde{\pi} \geq (1 - \delta_l)(s_h - \bar{u} - \tilde{\pi}) \). But this will be the case if, for example, \( \delta_h \) and \( \delta_l \) are
sufficiently close (note that \( s_h \) and \( s_l \) are close if \( \delta_h \) and \( \delta_l \) are close, by the Theorem of the Maximum). Thus, (10) cannot hold if \( \delta_h \) and \( \delta_l \) are close. This implies that separation is impossible if \( \delta_h \) and \( \delta_l \) are sufficiently close, and the high type’s behavior is distorted at only \( t = 0 \).

**Proof of Lemma 4.** Let \( \{w_t, b_t\}_{t=0}^{\infty} \) denote the optimal separating set of contracts, and let \( e_t \) denote the effort implemented at \( t \). Assume towards a contradiction that the optimal separating contract specifies \( u_{t+1} > \frac{\bar{u}}{1 - \delta} \) for some \( t \geq 0 \) on-the-equilibrium path. In period \( t \) of a separating equilibrium, either (i) \( b_t > \delta_l \pi_{t,t+1}^i - \frac{\delta_l}{1 - \delta_l} \bar{\pi} \); or (ii) \( b_t \leq \delta_l \pi_{t,t+1}^i - \frac{\delta_l}{1 - \delta_l} \bar{\pi} \), where \( \pi_{t,t+1}^i \) represents the imitation payoff of a low type principal who has not defaulted until \( t + 1 \) from \( t + 1 \) onwards. First, I show that if (i) holds, then \( u_{t+1} > \frac{\bar{u}}{1 - \delta} \) cannot be optimal. The reason is as follows. Consider the following changes in the separating contract at \( t \) and \( t + 1 \): \( b_t \) is increased by a small amount \( \delta \varepsilon \), whereas \( w_{t+1} \) is reduced by \( \varepsilon \); this implies that \( u_{t+1} \) decreases by \( \varepsilon \), whereas \( \pi_{h,t+1} \) and \( \pi_{l,t+1} \) increase by \( \varepsilon \). The bonus reward \( b_t + \delta \varepsilon \) is still contingent on effort level \( e_t \) (that is, output level \( y(e_t) \)) as in the original contract. This modified separating contract strictly increases the payoff of the high type, the low type is still deterred from imitation, and the agent is unaffected. To see why, first note that the agent’s participation constraint is still satisfied at every \( t \) (since \( \varepsilon \) is small), and the agent’s incentive-compatibility constraint for choosing effort level \( e_t \) is still satisfied at every \( t \geq 0 \) by construction provided that the increased bonus payment is enforceable with the high type principal. But this is indeed the case since \( \pi_{h,t+1} \) increases by \( \varepsilon \), and

\[
 b_t + \frac{\delta_h}{1 - \delta_h} \bar{\pi} \leq \delta_h \pi_{h,t+1}
\]

together with \( \delta_h > \delta \) implies that

\[
 b_t + \delta \varepsilon + \frac{\delta_h}{1 - \delta_h} \bar{\pi} < \delta_h (\pi_{h,t+1} + \varepsilon)
\]

Thus, \( b_t + \delta \varepsilon \) is enforceable with the high type. Assuming that \( \varepsilon \) is sufficiently small, \( b_t + \delta \varepsilon > \delta_l (\pi_{t,t+1}^i + \varepsilon) - \frac{\delta_l}{1 - \delta_l} \bar{\pi} \) also holds. Thus, the low type’s strategy is unaffected, and therefore, there is no change in the low type’s imitation payoff. But the high type’s payoff increases by \( \delta_{h}^{-1}(\delta_h - \delta)\varepsilon \) in the modified separating contract, a contradiction.

If \( b_t \leq \delta_l \pi_{t,t+1}^i - \frac{\delta_l}{1 - \delta_l} \bar{\pi} \), then again, \( u_{t+1} > \frac{\bar{u}}{1 - \delta} \) is suboptimal. Consider the following changes to the contract: \( b_t \) is increased by a small amount \( \delta \varepsilon \), \( w_{t+1} \) is reduced by \( \varepsilon \) (i.e., \( u_{t+1} \) is reduced by \( \varepsilon \)), and \( w_0 \) is increased by \( \delta_l (\delta_l - \delta)\varepsilon \). This modified separating contract strictly increases the payoff of the high type, the low type is still deterred from imitation, and the agent is unaffected. To see why, first assume that \( b_t \geq 0 \). Note that the agent’s participation constraint is still satisfied at every \( t \) (since \( \varepsilon \) is small), and the agent’s incentive-compatibility constraint for choosing effort level \( e_t \) is still satisfied at every \( t \) by construction because the
increased bonus payment is enforceable with the high type principal (note that the decrease in $w_{t+1}$ increases $\pi_{h,t+1}$ by $\varepsilon$). Since $b_t \leq \delta t \pi_{t,t+1}^{i} - \frac{\delta_t}{1-\delta_t} \bar{\pi}$, it follows that

$$b_t + \delta \varepsilon < \delta t \left( \pi_{t,t+1}^{i} + \varepsilon \right) - \frac{\delta_t}{1-\delta_t} \bar{\pi}.$$ 

This implies that $\pi_{t,t}^{i}$ increases by $(\delta - \delta) \varepsilon$. But since $w_0$ is increased by $\delta t (\delta - \delta) \varepsilon$ the imitation payoff of the low type $\pi_{t,0}^{i}$ is either the same as before or even lower than before, depending on when it would be optimal to default for an imitator. However, the high type’s payoff increases by $[\delta_t (\delta_h - \delta) - \delta_t (\delta_l - \delta)] \varepsilon > 0$, a contradiction. The argument is similar if $b_t < 0$—this time, the agent’s enforcement constraint matters. Hence, $u_{t+1} = \frac{\bar{g}}{1-\delta}$ must hold in the optimal contract for every $t \geq 0$.

**Proof of Lemma 5.** First, note that the optimal separating contract $\{w_t, b_t\}_{t=0}^{\infty}$ must be such that a low type principal who imitates the high type strictly prefers defaulting at some $t \geq 0$. Suppose not. Then, $\{w_t, b_t\}_{t=0}^{\infty}$ is an equilibrium that can be implemented with the low type in a symmetric-information setting. Thus, either $\{w_t, b_t\}_{t=0}^{\infty}$ is such that $\{w_t, b_t\} = C_l$ and $e_l$ is implemented at every $t$, which is a contradiction, or the imitation payoff of the low type from $\{w_t, b_t\}_{t=0}^{\infty}$ is strictly lower than $\frac{y(e_l) - w_l - b_l}{1-\delta}$. This latter is also a contradiction because then $\{w_t, b_t\}_{t=0}^{\infty}$ can be strictly improved upon as follows. There exists a large enough (but finite) $T$ such that if the high type starts offering $C_h$ from $t$ onwards (without any change prior to $T$), then the imitation payoff of the low type increases by a very small amount—i.e., the imitation payoff is still lower than $\frac{y(e_l) - w_l - b_l}{1-\delta}$, and the low type is still deterred from imitation. Moreover, the agent’s incentives are unaffected given Lemma 4. But the high type is strictly better off, a contradiction. Thus, the optimal separating contract $\{w_t, b_t\}_{t=0}^{\infty}$ is such that the low type strictly prefers defaulting at some $t \geq 0$. Let $T$ be the first period such that

$$b_T > \delta t \pi_{t,T+1}^{i} - \frac{\delta_t}{1-\delta_t} \bar{\pi}$$

in the optimal separating contract (recall that $\pi_{t,t}^{i}$ represents the imitation payoff of a low type principal who has not defaulted until $t$ from $t$ onwards). Either $b_T = b_h$, or $b_{T+1} = b_h$, or there exists a $t > T$ such that the high type will offer $C_h$ from $t$ onwards. In the first two cases, the desired result is obtained. Next, assume that $b_T < b_h$, and that $b_{T+1} < b_h$. Given that the low type strictly prefers defaulting at $T$, there must exist a $t > T$ such that the high type starts offering $C_h$ from $t$ onwards. Suppose not. But there exists a large enough $T' > T$ such that if the high type offers $C_h$ from $T'$ onwards, then $\pi_{t,T+1}^{i}$ increases by a very small amount, and the imitation payoff of the low type is unaffected as the imitator still strictly prefers defaulting at $T$. But then the high type is strictly better off, a contradiction.
Proof of Proposition 6. First, I show that \( b_t < b_{t+1} \) for every \( t \geq 0 \) until \( b_t = b_h \), which takes place in finite time. Let \( T = \min\{t \in \mathbb{N}|b_t = b_h\} \). Hence, \( b_t = b_h \) for all \( t \geq T \). Given Lemma 5, costly signaling ends at a finite period, and thus, \( T < \infty \). First, note that \( b_{T-1} < b_T = b_h \) by the definition of \( T \). Otherwise, either \( b_{T-1} > b_T \), in which case the high type principal defaults or \( b_{T-1} = b_h \), which is impossible due to the definition of \( T \). Next, I show that \( b_{T-2} < b_{T-1} \). Suppose towards a contradiction that \( b_{T-2} \geq b_{T-1} \). I will make use of the following lemma.

**Lemma 10** Let \( T = \min\{t \in \mathbb{N}|b_t = b_h\} \). In the optimal contract,

\[
 b_t + \frac{\delta_t}{1 - \delta_t} \bar{\pi} \leq \delta_t \pi_{t,t+1}^i
\]

for every \( t < T - 1 \).

**Proof.** Suppose not, so that \( b_t + (\delta_t \bar{\pi})/(1 - \delta_t) > \delta_t \pi_{\tau,\tau+1}^i \) for some \( \tau < T - 1 \). Thus, a low type who imitates the high type would strictly prefer defaulting at \( \tau \). Since \( b_{T-1} < b_h \), it follows that \( b_{T-1} < \frac{\delta_{T-1}}{1 - \delta_h}(s_h - \bar{\pi} - \bar{\pi}) \). So, the high type can increase \( b_{T-1} \) by a small amount \( \varepsilon \), which is still enforceable for the high type. This enables the high type to demand an output level \( y(c_{T-1}') \) such that \( c(c_{T-1}') = b_{T-1} + \varepsilon \), which strictly increases the surplus and the high type’s payoff at \( T - 1 \). Observe that this change in the contract does not affect the imitation payoff of the low type provided that \( \varepsilon \) is small enough. But the high type is strictly better off, a contradiction. Thus, in the optimal contract \( b_t + (\delta_t \bar{\pi})/(1 - \delta_t) \leq \delta_t \pi_{t,t+1}^i \) must hold for every \( t < T - 1 \), where \( T = \min\{t \in \mathbb{N}|b_t = b_h\} \). □

Since \( b_{T-1} < b_T = b_h \) it follows that \( \pi_{t,T-1}^i < \pi_{t,T}^i = V_t \). From Lemma 10,

\[
 b_{T-2} \leq \delta_t \pi_{T-1}^i \leq \frac{\delta_t}{1 - \delta_t} \bar{\pi}.
\]

Since \( \pi_{T-1}^i < \pi_{T}^i \), it follows that

\[
 b_{T-1} \leq b_{T-2} \leq \delta_t \pi_{T}^i \leq \frac{\delta_t}{1 - \delta_t} \bar{\pi}.
\]

However,

\[
 b_{T-1} < \delta_t \pi_{T}^i - \frac{\delta_t}{1 - \delta_t} \bar{\pi} \quad (11)
\]

\(^{40}\)In the optimal contract, \( b_t = c(e_t) \) for every \( t \geq 0 \). To see why, first note that future rewards are not used in the optimal separating contract, and thus, \( b_t \geq c(e_t) \). To see why \( b_t = c(e_t) \), suppose towards a contradiction that \( b_t > c(e_t) \) for some \( t \geq 0 \). Then, \( e_t \) and \( w_t \) can be changed to \( e_t' \) and \( w_t' \), respectively, such that \( c(e_t') = b_t \), and \( e_t' = \bar{u} \) and \( w_t \) would be increased to \( w_t' \) such that \( w_t' = w_t + \delta_t (s(e_t') - s(e_t)) \) assuming that \( t > 0 \). As a result, the imitation payoff of the low type is unaffected, and the high type is strictly better off due to her higher discount factor, a contradiction. If, however, \( t = 0 \), then the same arrangement makes the agent strictly better off while the high type and the low type are unaffected.

37
cannot hold in the optimal contract. To see why, first note that an arbitrarily small increase in $b_{T-1}$ is enforceable for the high type since the surplus becomes $s_h$ after $T-1$, and $b_{T-1} < b_h$. Second, note that since (11) holds, a very small increase in $b_{T-1}$ increases the payoff of the high type by strictly more than the imitation payoff of the low type. This is because the discount factors of the two types are different, and a low type who imitates until $T-1$ still prefers honoring at $T-1$ as the increase in $b_{T-1}$ is arbitrarily small and (11) holds. Thus, the high type can (i) increase $b_{T-1}$ by a small $\epsilon > 0$, (ii) increase the required output level to $y(e'_{T-1})$ in a way that $c(e'_{T-1}) = b_{T-1} + \epsilon$ so that the surplus at $T-1$ increases by $(s(e'_{T-1}) - s(e_{T-1}))$, and (iii) increase the initial fixed wage $w_0$ by $\delta_l^{-1} s(e_{T-1}) - s(e_{T-1})$ in order to deter the low type from imitating. This modified contract is still separating, and the high type’s payoff increases by $(\delta_{h}^{T-1} - \delta_l^{T-1}) (s(e'_{T-1}) - s(e_{T-1})) > 0$. Hence, a contradiction. As a result, (11) cannot hold, and $b_{T-2} < b_{T-1}$.

Next, assume that $b_{\tau} < b_{\tau+1}$ holds for all $\tau \in \{t, t+1, \ldots, T-1\}$ by the induction hypothesis. I now show that $b_{t-1} < b_t$ must also hold. The proof of this is very similar to the proof above for the claim that $b_{T-2} < b_{T-1}$. First, one needs to verify that $\pi_{t, \tau}^i < \pi_{l, \tau+1}^i$ and $\pi_{h, \tau} < \pi_{h, \tau+1}$ for all $T-1 \geq \tau \geq t$. But this is true due to the hypothesis that $b_{\tau}$ is monotone increasing for $\tau \geq t$, $b_t = c(e_t)$ for every $t \geq 0$ and due to the fact that continuation payoffs are not used to motivate the agent by Lemma 4. From Lemma 10, it follows that

$$b_{t-1} \leq \delta_l \pi_{l, t}^i - \frac{\delta_l}{1 - \delta_l} \pi.$$

Now, assume towards a contradiction that $b_{t-1} \geq b_t$. From

$$b_t \leq b_{t-1} \leq \delta_l \pi_{l, t}^i - \frac{\delta_l}{1 - \delta_l} \pi < \delta_l \pi_{l, t+1}^i - \frac{\delta_l}{1 - \delta_l} \pi,$$

it follows that

$$b_t < \delta_l \pi_{l, t+1}^i - \frac{\delta_l}{1 - \delta_l} \pi.$$

But this implies that

$$b_t = \delta_h \pi_{h, t+1} - \frac{\delta_h}{1 - \delta_h} \pi.$$

Otherwise, the high type could (i) increase $b_t$ by a small $\epsilon > 0$, (ii) increase the required output level to $y(e_t')$ in a way that $c(e_t') = b_t + \epsilon$ so that surplus increases by $(s(e_t') - s(e_t))$, and (iii) increase the initial fixed wage $w_0$ by $\delta_l^{-1} (s(e_t') - s(e_t))$, and make a positive gain, as argued before. But then,

$$b_t = \delta_h \pi_{h, t+1} - \frac{\delta_h}{1 - \delta_h} \pi > \delta_h \pi_{h, t} - \frac{\delta_h}{1 - \delta_h} \pi \geq b_{t-1}$$

---

41 Recall that in the optimal contract $b_t = c(e_t)$ for every $t \geq 0$ (see Footnote 41).
implies that $b_{t-1} < b_t$, a contradiction. Since $b_t < b_{t+1}$ and $b_t = c(e_t)$ for every $t \leq T - 1$ it follows that $e_t < e_{t+1}$ and $s_t < s_{t+1}$ for every $t \leq T - 1$. Next, I will show that $b_t > b_{t+1}$ at every $t \geq 0$. Suppose not. Then, there exists a $t \geq 0$ such that $b_t \leq b_{t+1}$ given what I proved above. Note that in that case,

$$b_t < \delta_l \pi_{l,t+1}^i - \frac{\delta_l}{1 - \delta_l} \bar{\pi}.$$ 

Otherwise, a contradiction would follow because $\pi_{l,t+1}^i > \frac{s_t - u_t}{1 - \delta_l}$ must hold given that bonuses are monotone increasing and $b_\tau > b_t$ for all $\tau \geq t + 1$ given the way $t$ is determined. This implies that

$$b_t = \delta_h \pi_{h,t+1}^i - \frac{\delta_h}{1 - \delta_h} \bar{\pi}$$

must hold, similar to what I argued above. Yet, this is a contradiction since $b_t \leq b_{t+1}$ and $\pi_{h,t+1} > \frac{s_t - u_t}{1 - \delta_h}$ (this latter holds since $b_\tau > b_t$ at every $\tau \geq t + 1$, and $b_t = c(e_t)$ at every $t \geq 0$ in the optimal separating contract). As a result, $b_t > b_{t+1}$ at every $t \geq 0$. It follows that $e_t > e_{t+1}$ and $s_t > s_{t+1}$ at every $t \geq 0$ since $b_t = c(e_t)$ at every $t \geq 0$. Recall that in the optimal contract, $u_{t+1} = \frac{u}{1 - \delta}$ for $t \geq 0$. This implies that $w_t = \bar{u}$ at every $t > 0$ because $b_t = c(e_t)$ for every $t \geq 0$ in the optimal separating contract. Moreover, $w_0 > \bar{u}$: Since $b_t > b_{t+1}$, $s_t > s_{t+1}$, and $u_{t+1} = \frac{u}{1 - \delta}$ at every $t \geq 0$, the high type must offer a high enough $w_0$ so that the low type is deterred from imitating the high type.

**Proof of Proposition 8.** In order to prove Proposition 8, I state and prove a series of Claims. Note that Claims 1-3 do not rely on the assumptions A1 or A1’. The first claim below implies that future rewards are never used in the optimal hybrid contract, and hence, default takes place through refusing to pay the bonus.

**Claim 1** In the optimal hybrid equilibrium, $u_t = \frac{u}{1 - \delta}$ at every $t \geq 0$. Thus, only bonus payments are used to discipline the agent.

**Proof.** Let $\{w_t, b_t\}_{t=0}^\infty$ denote the set of contracts in the optimal hybrid equilibrium, and let $e_t$ denote the effort implemented at $t$ (note that $\{w_t, b_t\}$ is offered and accepted provided that the principal has not defaulted at any $\tau < t$). Obviously, $u_0 = \frac{u}{1 - \delta}$ must hold, otherwise $w_0$ can be reduced by a small amount and make both types better off. Next, suppose towards a contradiction that $u_{t+1} > \frac{u}{1 - \delta}$ at some $t \geq 0$. Note that either (i) $b_t > \delta_l \pi_{l,t+1}^i - \delta_l \pi_{l,t+1}^d$; or (ii) $b_t \leq \delta_l \pi_{l,t+1}^i - \delta_l \pi_{l,t+1}^d$, where $\pi_{l,t+1}$ represents the payoff of a low type principal who has never defaulted until $t + 1$ (by an abuse of notation), and $\pi_{l,t+1}^d$ represents the punishment payoff of a type-$\theta$ principal who defaulted at $t$. If (i) holds, then $u_{t+1} > \frac{u}{1 - \delta}$ cannot be optimal for any $t \geq 0$. The reason is as follows. In this case, the low type strictly prefers defaulting at $t$. Consider the modified hybrid contract: $b_t$ is increased by a small amount $\delta \varepsilon > 0$, and $w_{t+1}$ is reduced by $\varepsilon$; thus, $u_{t+1}$ reduces by $\varepsilon$ whereas $\pi_{h,t+1}$ and $\pi_{l,t+1}$ both increase by $\varepsilon$. The bonus reward $b_t + \delta \varepsilon$ is still contingent on $e_t$ as in the original contract, and everything else...
remains the same. This modified hybrid contract strictly increases the payoff of the high type, whereas the low type principal and the agent are unaffected. To see why, first note that the agent’s participation constraint is still satisfied at every $t$ (since $\varepsilon$ is small enough), and the agent’s incentive-compatibility constraint for choosing effort level $e_t$ is satisfied at every $t$ by construction as long as that the increased bonus payment is enforceable with the high type principal. But this is indeed the case since $\pi_{h,t+1}$ increases by $\varepsilon$, and

$$b_t + \delta_h \pi_{h,t+1}^d \leq \delta_h \pi_{h,t+1}$$

together with $\delta_h > \delta$ implies that

$$b_t + \delta \varepsilon + \delta_h \pi_{h,t+1}^d < \delta_h (\pi_{h,t+1} + \varepsilon).$$

Thus, $b_t + \delta \varepsilon$ is enforceable with the high type principle. With $\varepsilon$ sufficiently small, $b_t + \delta \varepsilon + \pi_{l,t+1}^d > \delta_l (\pi_{l,t+1}^l + \varepsilon)$ also holds. Thus, the low type still strictly prefers defaulting on the bonus promise at $t$, and therefore, there is no change in the low type’s imitation payoff. But the high type’s payoff increases by $\delta_h^{-1} (\delta_h - \delta) \varepsilon > 0$ in the modified separating contract, a contradiction.

Next, consider the case in which (ii) holds, and suppose towards a contradiction that $u_{t+1} > \frac{a}{1-\delta}$ for some $t \geq 0$. First, consider the case where $b_t \geq 0$. Consider the modified contract: $b_t$ is increased by a small amount $\delta_l \varepsilon$, $b_t + \delta_l \varepsilon$ implements $e_t$ as before, and $w_{t+1}$ is reduced by $\varepsilon$, which reduces $u_{t+1}$ by $\varepsilon$ and increases $\pi_{h,t+1}$ and $\pi_{l,t+1}$ by $\varepsilon$—everything else remains the same. As I explain below, the agent’s participation and incentive compatibility constraints are unaffected; therefore, this change strictly increases the payoff of the high type whereas the low type’s payoff is unaffected. To see why, note that

$$b_t + \delta_l \varepsilon \leq \delta_l (\pi_{l,t+1}^l + \varepsilon) - \delta_l \pi_{l,t+1}^d$$

holds. In particular, the low type’s strategy is exactly the same as before by construction. This is also true for the high type. Thus, the agent’s participation and incentive compatibility constraints still hold, and the high type’s payoff increases by $\lfloor \delta_h (\delta_h - \delta_l) \rfloor \varepsilon > 0$, whereas the low type’s payoff is the same as before, a contradiction. The argument is similar if $b_t < 0$—this time, the agent’s enforcement constraint matters. Hence, $u_t = \frac{a}{1-\delta}$ must hold in the optimal contract for every $t \geq 0$.

By Claim 1, default takes place only through refusing to pay the bonus, and thus, beliefs are updated only after $P_t$ is observed and $\mu_{t+1}^l = \mu_t^2$ for every $t$. Therefore, I focus on $\mu_t \equiv \mu_t^1$ for $t \geq 1$, where $\mu_t$ denotes the posterior belief of the sender at the beginning of $t \geq 1$, and $\mu_0$ denotes the prior belief at $t = 0$, as before.
Claim 2 If the optimal contract is hybrid, then there exists a $T < \infty$ such that the high type principal starts offering $C_h$ from $T$ onwards.

Proof. Let $\{w_t, b_t\}_{t=0}^{\infty}$ denote the optimal hybrid set of contracts, and let $e_t$ denote the effort implemented at $t$. To prove the claim, first I show that there exists a $\bar{T} < \infty$ such that the low type strictly prefers defaulting at period $\bar{T}$. To see why, assume towards a contradiction that this is not the case. But this would mean that $\{w_t, b_t\}_{t=0}^{\infty}$ is an equilibrium set of contracts that can be implemented with the low type in a symmetric-information setting without default, and thus, either $\{w_t, b_t\}_{t=0}^{\infty}$ is such that $\{w_t, b_t\} = C_t$ and $e_t$ is implemented at every $t$—this is a contradiction—or the imitation payoff of the low type from $\{w_t, b_t\}_{t=0}^{\infty}$ is strictly lower than $\frac{s_l - \bar{u}}{1 - \delta_l}$. This latter is also a contradiction because then either (i) the high type’s payoff from this hybrid contract is lower than $\frac{s_l - \bar{u}}{1 - \delta_l}$, and hence, a separating equilibrium is strictly better, or (ii) the high type’s payoff from the hybrid contract is higher than $\frac{s_l - \bar{u}}{1 - \delta_l}$, in which case it is possible to construct a separating contract in a way that the low type is strictly better off and the high type is indifferent. The construction is as follows. The high type offers the set of contracts $\{w_t, b_t\}_{t=0}^{\infty}$ and implements $e_t$ at every $t \geq 0$ exactly as in the original contract, whereas the low type offers $C_t$. This is indeed a separating equilibrium since the imitation payoff of the low type is strictly lower than $\frac{s_l - \bar{u}}{1 - \delta_l}$, and thus, the low type has no incentive to imitate. But this separating equilibrium generates a strictly higher payoff than the hybrid equilibrium, a contradiction. Hence, if the optimal contract is hybrid, then there exists a $\bar{T} < \infty$ such that the low type strictly prefers defaulting at $\bar{T}$. Given this, there must exist a finite $T$ such that the high type starts offering $C_h$ from $T$ onwards. The proof for this is similar to the argument in Lemma 5, and therefore omitted.

For the following claims, let $\lambda_t$ denote the equilibrium probability with which $b_t$ is honored assuming that past bonus payments have been honored. The high type honors the bonus payment at every $t \geq 0$ in the optimal hybrid contract, whereas the low type may default. Thus,

$$\lambda_t = \mu_t + (1 - \mu_t)\nu_t,$$

where $\nu_t$ denotes the equilibrium probability with which the low type principal honors $b_t$.

Claim 3 If the optimal hybrid contract is such that $\nu_0 > 0$ and the low type obtains an equilibrium payoff that is weakly lower than $\frac{s_l - \bar{u}}{1 - \delta_l}$, then it is strictly dominated by a separating contract. If the optimal contract is hybrid, then the equilibrium payoff of the low type is weakly greater than $\frac{s_l - \bar{u}}{1 - \delta_l}$.

Proof. Let $\{w_t, b_t\}_{t=0}^{\infty}$ denote the optimal hybrid set of contracts, and let $e_t$ denote the effort implemented by the contract at $t$. Suppose that $\nu_0 > 0$ and that the low type’s payoff is weakly lower than $\frac{s_l - \bar{u}}{1 - \delta_l}$ given this set of contracts. I will modify this contract to generate a separating contract that makes the high type strictly better off and the low type
weakly better off. I assume that the high type’s payoff from the optimal hybrid equilibrium is weakly greater than $\frac{s_l - \bar{u}}{1 - \delta_l}$; otherwise, the optimal hybrid contract is even worse than the optimal pooling contract. The separating contract of the high type is as follows. At $t = 0$, let

$$w'_0 = w_0 + \sum_{t=0}^{T} \delta_l^t (1 - \lambda_t) b_t,$$

where $T > 0$ is the first period such that $\nu_T = 0$ ($T > 0$ since $\nu_0 > 0$ by hypothesis), $b'_t = \lambda_t b_t$ for $t \geq 0$, and everything else is exactly the same as in the original hybrid contract for every $t \geq 0$. Note that it is without loss to assume that $T < \infty$ since the proof of Claim 2 indicates that a hybrid equilibrium with $T = \infty$ is strictly dominated. The low type, however, always offers $C_l$. This is a separating equilibrium since the imitation payoff of the low type cannot exceed $\frac{s_l - \bar{u}}{1 - \delta_l}$ by construction. Moreover, the high type is strictly better off in this equilibrium because $\delta_h > \delta_l$ and $\lambda_T < 1$. Hence, the first statement is proved. Given the first statement, I need to prove the second statement only in the case where $\nu_0 = 0$. In that case, information is fully revealed by the end of $t = 0$. Suppose towards a contradiction that the low type’s payoff is strictly lower than $\frac{s_l - \bar{u}}{1 - \delta_l}$. Then, I modify the optimal hybrid contract to generate a separating contract for the high type as follows. At $t = 0$, $w'_0 = w_0 + (1 - \mu_0) b_0 - \varepsilon$, and $b'_0 = \mu_0 b_0$ where $\varepsilon > 0$ is arbitrarily small, and everything else is the same for the high type as in the original contract. The low type offers $C_l$ in the separating contract. The low type strictly prefers doing so with sufficiently small $\varepsilon > 0$. Thus, both types are strictly better off, a contradiction.

From now on, I will assume that either $A1$ or $A1'$ holds. In the final claim of the proof (Claim 6), I will show that if $y'(e_h)/c'(e_h)$ is sufficiently larger than 1, then imposing the worst punishment is optimal, as stated in $A1'$. I will also show that a similar result obtains if $\delta_l$ and $\delta_h$ are not far from each other or if $\mu_0$ is sufficiently high.

**Claim 4** Consider the optimal hybrid equilibrium. If $\lambda_t b_t < b_l$ ($\lambda_t b_t \leq b_l$) at some $t \geq 0$ such that $\nu_t < 1$ (and $\nu_\tau > 0$ for every $\tau < t$ if $t > 0$), then the equilibrium payoff of the low type is strictly lower than $\frac{s_l - \bar{u}}{1 - \delta_l}$ (at most $\frac{s_l - \bar{u}}{1 - \delta_l}$).

**Proof.** For $t = 0$, the statement is obvious due to Claim 1. Suppose that $\lambda_t b_t < b_l$ at some $t > 0$ such that $\nu_t < 1$ and $\nu_\tau > 0$ for every $\tau < t$. Since $\lambda_t b_t < b_l$, it follows that $b_\tau < b_l$ for all $\tau < t$. To prove this, I will start by showing that $b_{t-1} < b_l$. Since $\nu_{t-1} > 0$ and $\nu_t < 1$ by hypothesis,

$$b_{t-1} \leq \delta_l \pi_{t,t} - \frac{\delta_l}{1 - \delta_l} \bar{u} = \delta_l \left( s(e(\lambda_t, b_l)) - \bar{u} \right) + \lambda_t b_t + \frac{\delta_l}{1 - \delta_l} \bar{u} - \frac{\delta_l}{1 - \delta_l} \bar{u}$$

$$< \delta_l \frac{s_l - \bar{u} - \bar{u}}{1 - \delta_l} = b_l.$$
where $\pi_{l,t}$ represents the payoff of a low type principal who has not defaulted until $t$ as described in Claim 1, and $e(\lambda, b)$ denotes the effort level implemented in the optimal contract given that $b$ is honored with probability $\lambda$ and a future reward is not used because $\lambda_t b_t = c(e(\lambda_t, b_t))$ for all $t \geq 0$ in the optimal hybrid contract.\footnote{If it were the case that $\lambda_t b_t > c(e_t)$ at some $t > 0$ in the optimal hybrid contract, then $e_t$ and $w_t$ could be increased to $e'_t$ and $w'_t$, respectively, such that $c(e'_t) = \lambda_t b'_t$, and $w'_t = w_t + s(e'_t) - s(e_t)$. As a result, $u_t > \frac{u}{1 - \delta_t}$, and $e_0$, the equilibrium effort at $t = 0$, could be increased by a small amount because $u_t > \frac{u}{1 - \delta_t}$. Thus, the equilibrium strategy of the low type is unaffected, and both principal types are better off, a contradiction. Of course, $\lambda_0 b_0 > c(e_0)$ cannot hold in the optimal contract since $e_0 < e_h$. Therefore, in the optimal contract, $\lambda_t b_t = c(e_t)$ for all $t$.} The second inequality above follows because $\lambda_t b_t < b_t$, and thus,

$$\pi_{l,t} = s(e(\lambda_t, b_t)) - \bar{u} + \lambda_t b_t + \frac{\delta_t}{1 - \delta_t} \bar{\pi} < s_l - \bar{u} \frac{\delta_t}{1 - \delta_t} \bar{\pi}.$$ 

must hold. As a result, $b_{t-1} < b_t$. Next, I assume that $b_k < b_l$ for all $k \in \{\tau, \tau + 1, \ldots, t - 1\}$ by the induction hypothesis and show that $b_{\tau - 1} < b_.$. Since $\nu_{\tau - 1} > 0$,

$$b_{\tau - 1} \leq \delta_l \pi_{l, \tau} - \frac{\delta_l}{1 - \delta_l} \bar{\pi}.$$ 

But $\pi_{l, \tau} < \frac{s_l - \bar{u}}{1 - \delta_l}$ because (i) $\pi_{l,t} < \frac{s_l - \bar{u}}{1 - \delta_l}$ as argued above, and (ii) $b_k < b_l$ and $\lambda_k b_k = c(e_k)$ imply that $s_k < s_l$ for all $k \in \{\tau, \tau + 1, \ldots, t - 1\}$. As a result, $b_{\tau - 1} < \delta_l \frac{s_l - \bar{u} - \bar{s}_l}{1 - \delta_l} = b_l$. Thus, $b_\tau < b_l$ and $t_\tau < s_l$ for all $\tau < t$. Given these and given that $\pi_{l,t} < \frac{s_l - \bar{u}}{1 - \delta_l}$, the equilibrium payoff of the low type is strictly lower than $\frac{s_l - \bar{u}}{1 - \delta_l}$. The proof for the case stated inside the parentheses is very similar and therefore, omitted.

In what follows, let $t_k$ index periods such that the posterior belief is updated from $t_k$ to $t_k + 1$; that is, $\nu_{t_k} < 1$, $\nu_t > 0$ for all $t < t_k$, and $\mu_{t_k} \neq \mu_{t_{k+1}}$. In particular, $t_0 = \min\{t \geq 0 | \mu_0 \neq \mu_{t+1}\}$, and $t_k = \min\{t > t_{k-1} | \mu_{t+1} \neq \mu_t\}$ provided that $\nu_{t_k - 1} > 0$.

Claim 5 If the optimal hybrid contract weakly dominates the optimal separating contract, then (1) $b_l < \lambda_{t_k} b_{t_k}$ at every $t < t_k$, and (2) $b_l \leq \lambda_{t_0} b_{t_0}$.

Proof. I start with part (1). If $v_0 = 0$, then $t_0 = 0$ and there is nothing to prove, so assume that $v_0 > 0$. Take an arbitrary $t_k > 0$, and suppose towards a contradiction that $b_{t_{k-1}} < \lambda_{t_k} b_{t_k}$. By the definition of $t_k$, $\nu_{t_k} < 1$ and $\nu_t > 0$ for all $t < t_k$. Thus,

$$\lambda_{t_k} b_{t_k} \leq b_{t_{k-1}} \leq \delta_l \left( s(e(\lambda_{t_k}, b_{t_k})) - \bar{u} + \lambda_{t_k} b_{t_k} + \frac{\delta_l}{1 - \delta_l} \bar{\pi} \right) - \frac{\delta_l}{1 - \delta_l} \bar{\pi}.$$ 

But this implies that $\lambda_{t_k} b_{t_k} \leq b_l$. By Claim 4, the payoff of the low type is at most $\frac{s_l - \bar{u}}{1 - \delta_l}$. But this is a contradiction given the initial hypothesis and Claim 3 because $v_0 > 0$. Hence, $b_{t_{k-1}} < \lambda_{t_k} b_{t_k}$ must hold. Next, assume that $b_l < \lambda_{t_k} b_{t_k}$ holds for all $t \in \{\tau, \tau + 1, \ldots, t_k - 1\}$
by the induction hypothesis. I now show that \( b_{r-1} < \lambda_{t_k}b_{t_k} \) also holds. Again, by the definition of \( t_k \), \( \nu_{r-1} > 0 \). Thus,

\[
b_{r-1} \leq \delta_l \left( \sum_{i=\tau}^{t_k} \delta_l^{i-\tau} (s(e(\lambda_i, b_i)) - \bar{u}) + \delta_l^{t_k-\tau} \lambda_{t_k}b_{t_k} + \frac{\delta_l}{1 - \delta_l} \right) - \frac{\delta_l}{1 - \delta_l} \bar{u}.
\]

I will now show that

\[
\sum_{i=\tau}^{t_k} \delta_l^{i-\tau} (s(e(\lambda_i, b_i)) - \bar{u}) + \delta_l^{t_k-\tau} \lambda_{t_k}b_{t_k} + \frac{\delta_l}{1 - \delta_l} \bar{u} < s(e(\lambda_{t_k}, b_{t_k})) - \bar{u} + \lambda_{t_k}b_{t_k} + \frac{\delta_l}{1 - \delta_l} \bar{u}. \tag{12}
\]

To see why this holds, first note that

\[
(\max_i s(e(\lambda_i, b_i)) - \bar{u}) \sum_{i=\tau}^{t_k} \delta_l^{i-\tau} \geq \sum_{i=\tau}^{t_k} \delta_l^{i-\tau} (s(e(\lambda_i, b_i)) - \bar{u}),
\]

and that \( s(e(\lambda_{t_k}, b_{t_k})) = \max_i s(e(\lambda_i, b_i)) \) because \( \lambda_i b_t = c(e(\lambda_i, b_t)) \) and by the induction hypothesis \( b_t < \lambda_{t_k}b_{t_k} \) for all \( t \in \{\tau, \tau + 1, \ldots, t_k - 1\} \). Therefore, it is enough to show that

\[
(s(e(\lambda_{t_k}, b_{t_k})) - \bar{u}) \sum_{i=0}^{t_k-\tau-1} \delta_l^i < (1 - \delta_l^{t_k-\tau})(s(e(\lambda_{t_k}, b_{t_k})) - \bar{u} + \lambda_{t_k}b_{t_k} + \frac{\delta_l}{1 - \delta_l} \bar{u})
\]

in order to prove that (12) holds. Note that \( \lambda_{t_k}b_{t_k} > b_t \) by Claims 3 and 4 and by the initial hypothesis that the optimal contract is hybrid. As a result, the right-hand side of the inequality above is strictly greater than

\[
(1 - \delta_l^{t_k-\tau}) \frac{s(e(\lambda_{t_k}, b_{t_k})) - \bar{u}}{1 - \delta_l}.
\]

Moreover, \( \sum_{i=0}^{t_k-\tau-1} \delta_l^i = \frac{1 - \delta_l^{t_k-\tau}}{1 - \delta_l} \). Thus, (12) holds. From (12), it follows that \( b_{r-1} < \lambda_{t_k}b_{t_k} \) must hold. Otherwise, the implication is that \( \lambda_{t_k}b_{t_k} < b_t \), but this is a contradiction by Claims 3 and 4. Finally, I show that if the optimal contract is hybrid, then \( b_t \leq \lambda_{t_0}b_{t_0} \) must hold. Suppose not. Then, \( b_t > \lambda_{t_0}b_{t_0} \). Again, this is a contradiction by Claims 3 and 4.

Now, assume that the optimal hybrid contract weakly dominates separating equilibria and that \( \nu_{t_0} > 0 \). Then, Claim 5 implies that \( b_{t_{k-1}} < \lambda_{t_k}b_{t_k} \) for every \( t_k \geq 0 \) such that \( k > 0 \). Let \( K \) be such that \( \nu_{t_K} = 0 \). Thus,

\[
b_t \leq \lambda_{t_0}b_{t_0} < \prod_{k=0}^{K} \lambda_{t_k}b_{t_k} \leq b_h \prod_{k=0}^{K} \lambda_{t_k} = \mu_0b_h
\]

since \( b_{t_K} \leq b_h \) and \( \prod_{k=0}^{K} \lambda_{t_k} = \mu_0 \). As a result, \( \mu_0 > \frac{b_t}{b_h} \) must hold. Otherwise, the contract
is strictly dominated by a separating equilibrium. The condition \( \mu_0 > \frac{b_l}{b_h} \) must also hold if the optimal contract is hybrid, and \( \nu_{t_0} = 0 \). There are two cases to consider: (i) \( t_0 > 0 \) and (ii) \( t_0 = 0 \). First, note that in either case \( \mu_0 \geq \frac{b_l}{b_h} \) since \( b_l \leq \lambda_{t_0} b_{t_0} = \mu_0 b_{t_0} \leq \mu_0 b_h \) from Claim 5. If \( t_0 > 0 \), then \( \mu_0 = \frac{b_l}{b_h} \) cannot hold due to Claims 3 and 4. Next, suppose towards a contradiction that the optimal contract is hybrid but \( \mu_0 = \frac{b_l}{b_h} \) and \( t_0 = 0 \). By Claim 5, the only possibility is that \( b_0 = b_h \). While this gives the low type a payoff of \( \frac{g - \bar{u}}{1 - b_h} \), a contradiction. Hence, I showed that the optimal hybrid contract is strictly dominated if \( \mu_0 \leq \frac{b_l}{b_h} \).

I now show that there exists an \( \varepsilon > 0 \) such that the optimal hybrid contract is strictly dominated for \( \mu_0 \in \left( \frac{b_l}{b_h}, \frac{b_l}{b_h} + \varepsilon \right) \). This is because the optimal hybrid contract is strictly dominated if \( \mu_0 = \frac{b_l}{b_h} \), and the payoff of the optimal hybrid equilibrium is continuous in \( \mu_0 \), as I will now show (the payoff of the optimal separating equilibrium does not depend on \( \mu_0 \)). Take an arbitrary \( \mu_0 \) and an arbitrary sequence \( \{\mu^n_0\} \) such that \( \lim_{n \to \infty} \mu^n_0 = \mu_0 \). Let \( \pi \) and \( \pi^n \) denote the payoff of the optimal hybrid contract with \( \mu_0 \) and \( \mu^n_0 \), respectively. I will show that \( \pi^n \) converges to \( \pi \) as \( \mu^n_0 \) converges to \( \mu_0 \). First, I will first show that \( \lim_{n \to \infty} \pi^n \geq \pi \). To show this, I will construct a hybrid equilibrium with \( \mu^n_0 \) and large \( n \), as follows. Let \( t_0 \geq 0 \) denote the first period in which the low type defaults with positive probability in the optimal hybrid contract with \( \mu_0 \). Since \( \nu_{t_0} < 1 \), it follows that \( \lambda_{t_0} \in (0, 1) \). The hybrid contract that I will construct given \( \mu^n_0 \) is exactly the same as the optimal hybrid contract with \( \mu_0 \), in terms of the implemented effort level, the fixed wage, the bonus payment, and the default rate by the low type, with the following exception at period \( t_0 \). Take sufficiently large \( n \), and let \( \nu^n_{t_0} = \nu_{t_0} \frac{\mu^n_0}{1 - \mu^n_0} \frac{1 - \mu_0}{\mu_0} \) and \( \lambda^n_{t_0} = \mu_0 + (1 - \mu^n_0) \nu^n_{t_0} \) in the hybrid contract with \( \mu^n_0 \). Also, let the effort level implemented at \( t_0 \) be such that \( c(e^n_{t_0}) = \lambda^n_{t_0} b_{t_0} \), where \( b_{t_0} \) is the bonus in the original hybrid contract with \( \mu_0 \) at period \( t_0 \). Note that \( \nu^n_{t_0} < 1 \) with all sufficiently large \( n \). Moreover, as \( \mu^n_0 \) goes to \( \mu_0 \), \( \nu^n_{t_0} \) goes to \( \nu_{t_0} \) and \( \lambda^n_{t_0} b_{t_0} \) goes to \( \lambda_{t_0} b_{t_0} \). The posterior at \( t_0 + 1 \) is identical in the two contracts with \( \mu_0 \) and \( \mu^n_0 \) by construction, and everything else (in particular, the implemented effort level, the fixed wage, the bonus payment, the default rate by the low type) after period \( t_0 \) and prior to \( t_0 \) is the same. As a result, the payoff of this construction converges to \( \pi \) as \( \mu^n_0 \) converges to \( \mu_0 \). It follows that \( \lim_{n \to \infty} \pi^n \geq \pi \) must hold. Next, I will show that \( \lim_{n \to \infty} \pi^n = \pi \). Suppose towards a contradiction that there exists a sequence \( \{\mu^n_0\} \) such that \( \lim_{n \to \infty} \mu^n_0 = \mu_0 \) and \( \lim_{n \to \infty} \pi^n > \pi \). Let \( \varepsilon = \lim_{n \to \infty} \pi^n - \pi \). This time, I will construct a hybrid contract with \( \mu_0 \) given the optimal hybrid contract with \( \mu^n_0 \). Take a sufficiently large \( n \) and set \( \nu_{t_0} = \nu^n_{t_0} \frac{\mu_0}{1 - \mu_0} - \frac{1 - \mu^n_0}{\mu^n_0} < 1 \), \( \lambda_{t_0} = \mu_0 + (1 - \mu_0) \nu_{t_0} \) and \( c(e_{t_0}) = \lambda_{t_0} b_{t_0} \) where, this time, \( t_0 \geq 0 \) denotes the first period in which the low type defaults with positive probability in the optimal hybrid contract with \( \mu^n_0 \), and \( b_{t_0} \) is the bonus in the hybrid contract with \( \mu^n_0 \). Similar to the construction above, the posterior at \( t_0 + 1 \) is identical in the two contracts with \( \mu_0 \) and \( \mu^n_0 \) by construction, and everything else after period \( t_0 \) and prior to \( t_0 \) is the same. As a result, the payoff of this construction differs from
the optimal hybrid contract is strictly dominated if the very first claim in part (i) of Proposition 8 follows: There exists an $\varepsilon > 0$ such that

$$\lim_{n \to \infty} \pi^n > \pi$$

cannot hold. Hence, the proof is complete, and the very first claim in part (i) of Proposition 8 follows: There exists an $\varepsilon > 0$ such that the optimal hybrid contract is strictly dominated if $\mu_0 \leq \frac{b_h}{b_l} + \varepsilon$. For the following claim, consider $\mu_0$ such that $\mu_0 > 1 - \varepsilon$ for small $\varepsilon > 0$. I construct a hybrid equilibrium such that at $t = 0$ the fixed wage is $w_h$, and the bonus payment is $b_h$ contingent on effort level $e_0$, where $c(e_0) = \mu_0 b_h$. If the bonus payment is honored at $t = 0$, then the contract offer is $C_h$ from $t \geq 1$ onwards. The low type will default at $t = 0$, while the high type will always honor the contract at every $t \geq 0$. For sufficiently small $\varepsilon > 0$, the payoff of the high type and low type approximate $s_h - \bar{\bar{u}}_t b_l$ and $s_h - \bar{\bar{u}} + b_h + \bar{\bar{\delta}}_t \frac{n}{1-\bar{\bar{\delta}}}$, respectively. But the separating equilibrium payoff for the high type is bounded above away from $s_h - \bar{\bar{u}}$ due to a fixed cost of signaling which is independent of $\mu_0$, while the low type’s payoff is only $s_h - \bar{\bar{u}}$. Thus, if $\mu_0$ is sufficiently high and close to one, then the optimal contract is hybrid.

Next, I show that given $\gamma \in (0, 1)$, there exists a unique $\mu_0$ such that $\mu_0 \leq \mu_0 \gamma$ and the optimal contract is separating if $\mu_0 \leq \mu_0 \gamma$ and hybrid otherwise. First, I show why a single cutoff exists. This is because, while the prior belief $\mu_0$ does not affect the payoff of the optimal separating contract for fixed $\gamma$, the payoff of the optimal hybrid contract strictly increases in $\mu_0$. To see why, let $\mu_0 > \mu_0$. I will now modify the optimal contract with $\mu_0$ and generate a hybrid contract with $\mu_0'$ that gives a strictly higher payoff for both types. Let $t_k$ index the periods in which the low type defaults with strictly positive probability in the optimal hybrid equilibrium with $\mu_0$; that is, $\nu_{t_k} < 1$ and $\mu_{t_k} = \mu_{t_k + 1}$. Since $\mu_0' > \mu_0$, there exists a $t_K$ such that $\mu_{t_K} < \mu_0' \leq \mu_{t_K + 1}$ (note that $\mu_0 = \mu_0$ by the definition of $t_k$). Then, the default rate of the low type in the modified contract with $\mu_0'$ is zero at every $t < t_K$, and the implemented effort levels are compatible with this, that is, $\nu_{t_k}' = 1$ and $c(e_{t_k}') = b_t$ at every $t < t_K$, where $b_t$ represents the bonus at period $t$ in the original hybrid contract with $\mu_0$. Everything else is the same as in the original contract until $t_K$. In period $t_K$, $\nu_{t_K}'$ is such that $\nu_{t_K}' = \nu_{t_K} \frac{\mu_0'}{1-\mu_0} \frac{1-\mu_{t_K}}{\mu_{t_K}}$ holds. This construction ensures that the posterior at $t_K + 1$ is the same in both contracts; that is, $\mu_{t_K + 1} = \mu_{t_K + 1}'$. Moreover, $\nu_{t_K}' \geq \nu_{t_K}$, $\lambda_{t_K}' > \lambda_{t_K}$, and also, $c(e_{t_K}') = \lambda_{t_K}' b_{t_K}$; thus, a strictly higher effort level is implemented at $t_K$ in the modified contract. The rest of the modified contract is identical to the original contract. The modified contract with prior $\mu_0'$ generates a strictly higher payoff than the optimal hybrid contract with $\mu_0$ for both types since (i) the implemented effort in the contract with $\mu_0'$ is weakly higher in every period and strictly higher in, at least, one period until period $t_K + 1$ at no additional cost and with no change in incentive constraints, and (ii) everything is identical in the two contracts from $t_K + 1$ onwards. Hence, the desired result.

I now show that $\mu_0$ is increasing in $\gamma$ to complete the proof of part (i). To see why, suppose towards a contradiction that there exists a $\tilde{\gamma}$ such that $\mu_\gamma' < \mu_\gamma$ for some $\gamma' > \tilde{\gamma}$. 

46
At the prior $\mu_\gamma$, 
\[ \bar{\gamma}U_{sep}^h + (1 - \bar{\gamma})U_{sep}^l = \bar{\gamma}U_{hyb}^h(\mu_\gamma) + (1 - \bar{\gamma})U_{hyb}^l(\mu_\gamma) \]
where $U_{sep}^\theta$ ($U_{hyb}^\theta$) represents the optimal separating (hybrid) equilibrium payoff of type $\theta$. Note that $U_{hyb}^\theta$ depends on the prior belief since the payoff of the optimal hybrid equilibrium is strictly increasing in the prior, as I showed above. At the prior $\mu_{\gamma'}$, 
\[ \gamma'U_{sep}^h + (1 - \gamma')U_{sep}^l = \gamma'U_{hyb}^h(\mu_{\gamma'}) + (1 - \gamma')U_{hyb}^l(\mu_{\gamma'}) \]
\[ < \gamma'U_{hyb}^h(\mu_\gamma) + (1 - \gamma')U_{hyb}^l(\mu_\gamma) \]
where the strict inequality follows because $\mu_{\gamma'} < \mu_\gamma$. This implies that
\[ (\gamma' - \bar{\gamma})(U_{sep}^h - U_{sep}^l) < (\gamma' - \bar{\gamma})(U_{hyb}^h(\mu_\gamma) - U_{hyb}^l(\mu_\gamma)) \]
and $U_{sep}^h < U_{hyb}^h(\mu_\gamma)$ since $U_{sep}^l = \frac{s_l - \bar{u}}{1 - \delta_l}$ and $U_{hyb}^l(\mu_\gamma) \geq \frac{s_l - \bar{u}}{1 - \delta_l}$ (the latter holds since if $U_{hyb}^l(\mu_\gamma) < \frac{s_l - \bar{u}}{1 - \delta_l}$, then the hybrid equilibrium would be strictly dominated by a separating equilibrium due to Claim 3). But if $U_{sep}^h < U_{hyb}^h(\mu_\gamma)$, then
\[ \bar{\gamma}U_{sep}^h + (1 - \bar{\gamma})U_{sep}^l < \bar{\gamma}U_{hyb}^h(\mu_\gamma) + (1 - \bar{\gamma})U_{hyb}^l(\mu_\gamma) \]
a contradiction. Note that the steps above shows that if $\bar{\gamma}$ is such that $U_{hyb}^l(\mu_{\bar{\gamma}}) > \frac{s_l - \bar{u}}{1 - \delta_l}$, then it must be the case that $\mu_{\gamma'} > \mu_\gamma$ for all $\gamma' > \bar{\gamma}$. Hence, $\mu_{\gamma}$ is strictly increasing at sufficiently low levels of $\gamma$. This is because if $\gamma$ is low enough, then $U_{hyb}^l(\mu_{\gamma}) > \frac{s_l - \bar{u}}{1 - \delta_l}$ must hold in the optimal hybrid contract. To see why, suppose towards a contradiction that $U_{hyb}^l(\mu_{\gamma}) = \frac{s_l - \bar{u}}{1 - \delta_l}$ for all $\gamma \in (0, 1)$. By Claim 3, $U_{hyb}^l(\mu_{\gamma}) = \frac{s_l - \bar{u}}{1 - \delta_l}$ is possible only if $\nu_0 = 0$; that is, the low type defaults with probability one at $t = 0$. Thus, $b_0$ is such that $b_0\mu_{\gamma} = b_t$ so that $U_{hyb}^l(\mu_{\gamma}) = \frac{s_l - \bar{u}}{1 - \delta_l}$ can hold. Moreover, $b_0 < b_h$ since $b_0\mu_{\gamma} = b_t$ and $\mu_{\gamma} > \frac{b_0}{b_h}$, thus $b_0$ can be increased slightly and make the low type better off. If $\gamma$ is low enough (for example, lower than $\mu_{\gamma'}(c_{0_h})$), then $U_{hyb}^l(\mu_{\gamma}) = \frac{s_l - \bar{u}}{1 - \delta_l}$ cannot hold because the increase in the low type’s payoff due to a small increase in $b_0$ makes up for the decrease in the high type’s payoff (if there is a decrease at all).

Next, I prove part (ii); i.e., if the optimal contract is hybrid, then $b_t$, $e_t$ and $s_t$ are strictly increasing (as long as bonus payments are honored) until they reach the symmetric information benchmark, which takes place in finite time. To prove this, I will start with the last period in which the bonus is different from $b_h$ just as in the proof of Proposition 6. So, let $T = \min\{t \in \mathbb{N}|b_t = b_h\}$. Note that $b_T = b_h$ for all $t \geq T$ on the equilibrium path, otherwise the bonus payment would not be enforceable with the low type. By Claim 2, $T < \infty$. If $T = 0$, then $b_0 = b_h$ and there is nothing to prove (indeed, if $\mu_0$ is sufficiently high and close to one, then $b_0 = b_h$). So, assume that $T > 0$. If $T > 1$, then I will assume
without loss of generality that \( \nu_t > 0 \) at every \( t < T - 1 \), and thus equilibrium belief \( \mu_t < 1 \) for all \( t < T \). By the definition of \( T \), \( b_{T-1} < b_h \) must hold. Moreover, \( e_{T-1} < e_T \) must hold. To see why, note that either \( e_T = e_h \) or \( e_T < e_h \). In the former case, it is immediate that \( e_{T-1} < e_T \). Next, assume that \( e_T < e_h \). Given that \( \lambda_T b_h = c(e_T) < c(e_h) \) in the optimal contract, it follows that \( \lambda_T < 1 \). Thus, in equilibrium \( \nu_{T-1} > 0 \) and \( \nu_T = 0 \). Rewriting \( e_T = e(\lambda_T, b_T) \), it follows that

\[
b_{T-1} + \frac{\delta_l}{1 - \delta_l} \bar{\pi} \leq \delta_l \pi_{l,T} = \delta_l \left( s(e(\lambda_T, b_T)) - \bar{u} + \lambda_T b_T + \frac{\delta_l}{1 - \delta_l} \bar{\pi} \right).
\]

As before, \( \pi_{l,t} \) denotes the continuation payoff of a low type principal who has not defaulted until period \( t \) (by an abuse of notation). If it were the case that \( e_{T-1} \geq e(\lambda_T, b_T) \), then \( b_{T-1} \geq \lambda_T b_T \) would follow, and the inequality above would imply that \( \lambda_T b_T \leq b_l \), a contradiction by Claims 3 and 4. Thus, \( e_{T-1} < e_T \) Next, I show that \( b_{T-2} < b_{T-1} \) assuming that \( T \geq 2 \). Suppose not, so that \( b_{T-2} \geq b_{T-1} \). Since \( \nu_{T-2} > 0 \),

\[
b_{T-1} \leq b_{T-2} \leq \delta_l \pi_{l,T-1} - \frac{\delta_l}{1 - \delta_l} \bar{\pi}.
\]

I now argue that \( b_{T-1} < \delta_l \pi_{l,T} - \frac{\delta_l}{1 - \delta_l} \bar{\pi} \) must hold if \( b_{T-2} \geq b_{T-1} \). Otherwise, the inequality above implies that \( \pi_{l,T-1} \geq \pi_{l,T} \). But this cannot hold given that \( e_{T-1} < e_T \) and that \( \pi_{l,T} > \frac{b_T - \bar{u}}{1 - \delta_l} \) (this latter holds because \( e_l < e_T \)). Thus, \( b_{T-1} < \delta_l \pi_{l,T} - \frac{\delta_l}{1 - \delta_l} \bar{\pi} \). Yet, this gives rise to another contradiction. Given this, \( b_{T-1} \) can be increased by a small \( \epsilon > 0 \) and the implemented effort level can be modified in a way that \( c(e_{T-1}) = b_{T-1} + \epsilon \). This increases the payoff of both types if \( \nu_t = 1 \) for all \( t \leq T - 1 \). If there exists a \( t < T - 1 \) such that \( \nu_t < 1 \), then the incentive of the low type to default at \( t \) is distorted since \( \pi_{l,t+1} \) increases due to the increase in surplus at \( T - 1 \); in particular, the low type strictly prefers paying the bonus at \( t \) rather than defaulting. Note however that increasing \( b_t \) by the amount of the increase in \( \delta_l \pi_{l,t+1} \) (without changing the implemented effort level \( e_t \)) leaves the low type’s continuation payoff and strategy at \( t \) unaffected. The high type has a strictly higher continuation payoff due to her higher discount factor. Moreover, given that the agent’s continuation payoff at \( t \) is strictly higher with the increase in \( b_t \) (there is no change in \( e_t \)), the output requirement at \( t = 0 \) can be increased by a small amount. Thus, both types are strictly better off, a contradiction. Hence, it follows that \( b_{T-2} < b_{T-1} \) must hold. This in turn implies that \( e_{T-2} < e_{T-1} \). Suppose not so that \( e_{T-2} \geq e_{T-1} \), which in turn implies that \( \lambda_{T-1} < 1 \) and that \( b_{T-2} \geq \lambda_{T-1} b_{T-1} \). However, since \( \nu_{T-2} > 0 \), the low type’s enforcement constraint (see above) combined with the inequality \( b_{T-2} \geq \lambda_{T-1} b_{T-1} \) implies that \( \lambda_{T-1} b_{T-1} \leq b_l \), in

---

43The claim is still true if \( \nu_t = 0 \) at some \( t < T - 1 \), and thus, equilibrium beliefs are degenerate from \( t + 1 \) onwards with probability 1—the proof of Proposition 6 can be directly applied for any \( t < T \) such that \( \mu_t = 1 \) in order to show that \( b_r < b_{r+1} \) and \( e_r < e_{r+1} \) for all \( r \in \{t, ..., T - 1\} \).
contradiction with Claims 3 and 4. Thus, \( e_{T-2} < e_{T-1} \). Next, assume that \( b_t < b_{t+1} \) and \( e_t < e_{t+1} \) for all \( t \in \{ \tau, \tau + 1, \ldots, T-2 \} \) by the induction hypothesis. I will show that \( b_{\tau-1} < b_\tau \) and \( e_{\tau-1} < e_\tau \) must hold. The proof of this is very similar to the proof above for the claim that \( b_{T-2} < b_{T-1} \) and \( e_{T-2} < e_{T-1} \). Suppose towards a contradiction that \( b_{\tau-1} \geq b_\tau \). Since \( \nu_{\tau-1} > 0 \),

\[
b_\tau \leq b_{\tau-1} \leq \delta_l \pi_{l,\tau} - \frac{\delta_l}{1 - \delta_l} \bar{\pi}.
\]

I will now argue that \( b_\tau < \delta_l \pi_{l,\tau+1} - \frac{\delta_l}{1 - \delta_l} \bar{\pi} \) must hold. Otherwise, the inequality above implies that \( \pi_{l,\tau} \geq \pi_{l,\tau+1} \). But this cannot hold given that \( b_t \) and \( e_t \) are strictly increasing for all \( t \in \{ \tau, \tau + 1, \ldots, T \} \) and \( \pi_{l,T} > \frac{\pi_l - \bar{\pi}}{1 - \delta_l} \). Thus, \( b_\tau < \delta_l \pi_{l,\tau+1} - \frac{\delta_l}{1 - \delta_l} \bar{\pi} \). But this is also a contradiction, similar to what I argued above; there is a modified hybrid contract which gives a strictly higher payoff to both types. Thus, \( b_{\tau-1} < b_\tau \). To see why \( e_{\tau-1} < e_\tau \), suppose towards a contradiction that \( e_{\tau-1} \geq e_\tau \). This implies that \( \lambda_\tau < 1 \) and that \( b_{\tau-1} \geq \lambda_\tau b_\tau \). Since \( \nu_{\tau-1} > 0 \) and \( \nu_\tau < 0 \), the low type’s enforcement constraint combined with the inequality \( b_{\tau-1} \geq \lambda_\tau b_\tau \) and the fact that

\[
\pi_{l,\tau} = s(c(\lambda_\tau, b_\tau)) - \bar{u} + \lambda_\tau b_\tau + \frac{\delta_l}{1 - \delta_l} \bar{\pi},
\]

implies that \( \lambda_\tau b_\tau \leq b_t \), in contradiction with Claims 3 and 4.

Finally, I establish the following claim regarding the assumptions A1 and A1’ and complete the proof. Note that below I focus on hybrid equilibria in which the low type’s payoff is higher than \( \frac{\pi_l - \bar{\pi}}{1 - \delta_l} \), which is without loss of generality because otherwise, Claim 3 implies that the hybrid equilibrium is strictly dominated by a separating equilibrium. Note that this is true regardless of the form of the punishment strategy as Claim 3 does not make any assumption thereof. Thus, using Claim 3 enables me to obtain sharper results regarding the optimality of A1.

**Claim 6** Let \( \tilde{b} \) be such that \( \tilde{b} = c(e(\tilde{b})) \) and \( y(c(\tilde{b})) = y(e(t)) - c(e(t)) \), and consider hybrid equilibria that give the low type an equilibrium payoff greater than \( \frac{\pi_l - \bar{\pi}}{1 - \delta_l} \). If \( y'(e_h)/c'(e_h) \) is sufficiently larger than 1 (for example, higher than \( \frac{\mu_0}{b} \)), then it is optimal to impose the worst punishment after a default. The same is also true if \( \delta_l \) and \( \delta_h \) are not far from each other or if \( \mu_0 \) is sufficiently high.

**Proof.** Assume that \( \pi_{l,t}^d > \frac{\pi_l}{1 - \delta_l} \) for some \( t \geq 0 \) in the optimal hybrid contract, where \( \pi_{l,t}^d \) is (as defined before) the punishment payoff of a type-\( \theta \) principal who defaulted at \( t \). There are two types of periods to consider: the period in which the low type defaults with probability one, and the periods in which the low type strictly randomizes between returning and defaulting. I start by period \( T \), the period in which the low type defaults with
probability one (that is, \( \nu_T = 0 \) and \( \nu_t > 0 \) for all \( t < T \)). Note that at period \( T \),

\[
b_T + \delta_h \pi_{h,T+1}^d = \delta_h \pi_{h,T+1},
\]

and

\[
b_T + \delta_l \pi_{l,T+1}^d \geq \delta_l \pi_{l,T+1},
\]

where (as before) \( \pi_{\theta,t+1} \) represents the payoff of a type-\( \theta \) principal who has not defaulted until \( t + 1 \). If it were the case that \( b_T + \delta_h \pi_{h,T+1}^d < \delta_h \pi_{h,T+1} \), then the punishment payoff \( \pi_{\theta,T+1}^d \) for defaulting in period \( T \) can be slightly increased to make the low type strictly better off with no effect on the high type, a contradiction. Assume that \( b_T \) is increased by a small amount \( \varepsilon > 0 \), whereas \( \delta_l \pi_{l,T+1}^d \) is decreased by the same amount. It follows that \( \delta_h \pi_{h,T+1}^d \) falls by more than \( \varepsilon \) due to the fact that \( \delta_h > \delta_l \), and thus the high type’s enforcement constraint is satisfied. This change increases the payoff of the optimal hybrid equilibrium at \( T \) for both types provided that \( \mu_T \frac{y'(e_T)}{c'(e_T)} > 1 \). Moreover, \( \mu_T \) (more generally, every \( \lambda_t \)) is bounded below away from zero (no matter how small \( \mu_0 \) might be) in every hybrid equilibrium such that the low type’s payoff is at least \( \frac{\varepsilon - \delta}{1 - \delta} \). This is because \( e(\mu_T, b_T) \) must be such that \( y(e(\mu_T, b_T)) = y(e(\hat{b})) > y(e_t) - c(e_t) > 0 \) holds (otherwise, the low type’s payoff is strictly lower than \( \frac{\varepsilon - \delta}{1 - \delta} \)). Thus, \( \mu_T > \frac{b}{b_0} > 0 \). Since \( \frac{y'(e_T)}{c'(e_T)} > \frac{y'(e_h)}{c'(e_h)} \), if \( \frac{y'(e_h)}{c'(e_h)} \) is sufficiently high, for example higher than \( \frac{b_0}{b} \), then the increase in \( b_T \) surely increases the payoff of the optimal hybrid equilibrium at \( T \). Note that the increase in \( \pi_{l,T} \) can affect the incentive of the low type to default before \( T \) if there exists a \( t < T \) such that \( \nu_t < 1 \). However, as I argued before, increasing \( b_t \) by the amount of the increase in \( \delta_l \pi_{l,t+1} \) (without changing the implemented effort level \( e_t \)) leaves the low type’s continuation payoff and strategy at \( t \) unaffected. The high type has a strictly higher continuation payoff due to her higher discount factor. Moreover, given that the agent’s continuation payoff at \( t \) is strictly higher with the increase in \( b_t \) (no change in \( e_t \)), the output requirement at \( t = 0 \) can be increased by a small amount. Thus, both types are strictly better off, a contradiction.

Next, consider an arbitrary period \( t < T \) such that \( \nu_t < 1 \) and \( \pi_{l,t+1}^d > \frac{\pi_{l,T+1}}{1 - \delta} \). Then, \( b_t + \delta_l \pi_{l,t+1}^d = \delta_l \pi_{l,t+1} \) and \( b_t + \delta_h \pi_{h,t+1}^d \leq \delta_h \pi_{h,t+1} \). Assume that \( b_t \) is increased by a small amount \( \varepsilon > 0 \) whereas \( \delta_l \pi_{l,t+1}^d \) is decreased by the same amount. It follows that the low type is indifferent between paying \( b_t + \varepsilon \) and defaulting. Moreover, the high type strictly prefers paying \( b_t + \varepsilon \) since the reduction in \( \delta_h \pi_{h,t+1}^d \) is larger than \( \varepsilon \) due to the fact that \( \delta_h > \delta_l \). This increase makes both types strictly better off if \( \lambda_t \frac{y'(e_t)}{c'(e_t)} > 1 \). This will be the case if for example \( \frac{y'(e_h)}{c'(e_h)} \) is higher than \( \frac{b_0}{b} \). The proof of this claim is very similar to the proof above and therefore, omitted.

Steps in the proof above already show that \( A1 \) is optimal if \( \mu_0 \) is sufficiently high. Next, I show that it is optimal if \( \delta_h \) and \( \delta_l \) are sufficiently close. To show this I will argue that if \( \delta_h \) and \( \delta_l \) are sufficiently close, then \( \lambda_t \frac{y'(e_t)}{c'(e_t)} > 1 \) at every \( t \geq 0 \). Suppose towards a contradiction
that there exists a \( t \) such that \( \lambda_t \frac{y'(e_t)}{y'(e_t)} \leq 1 \); that is, \( \lambda_t \leq \frac{y'(e_t)}{y'(e_t)} \) \( < 1 \). Thus, \( \lambda_t \) is bounded above away from 1 because \( \frac{c'(e_t)}{y'(e_t)} \) \( < \frac{c'(e_t)}{y'(e_t)} \) \( < 1 \). Note that the following must hold so that the high type is willing to honor the bonus promise \( b_t \).

\[
b_t \leq \delta_h \pi_{h,t+1} - \sum_{\tau=t+1}^{\infty} \delta_h^{\tau-t} (s_{\tau} - \bar{u}),
\]

where \( \pi_{h,t+1} < \frac{s_h - \bar{u}}{1 - \delta_h} \) if \( s_{\tau} - \bar{u} > \bar{\pi} \) for some \( \tau > t \). Moreover, it must be the case that if \( \delta_h \) and \( \delta_l \) are sufficiently close, then

\[
\lambda_t b_t > y(e(\lambda_t, b_t)) - (y(e_t) - c(e_t)) \geq \sum_{\tau=t+1}^{\infty} \delta_l^{\tau-t} (s_l - s_{\tau}).
\]

I start with the second inequality. If this inequality does not hold, then the low type’s payoff is strictly lower than \( \frac{s_l - \bar{u}}{1 - \delta_l} \) and the equilibrium is strictly dominated. The first inequality holds because \( \lambda_t \) is bounded above away from one and therefore, there exists a \( \epsilon > 0 \) such that if \( \delta_h - \delta_l < \epsilon \), then \( \lambda_t b_t < b_l \). Hence, the first inequality follows. Thus, it follows that

\[
\lambda_t \left( \delta_h \pi_{h,t+1} - \sum_{\tau=t+1}^{\infty} \delta_h^{\tau-t} (s_{\tau} - \bar{u}) \right) \geq \lambda_t b_t \geq \sum_{\tau=t+1}^{\infty} \delta_l^{\tau-t} (s_l - s_{\tau}).
\]

However, note that as \( \delta_h \) and \( \delta_l \) get closer and closer, it must be the case that

\[
\sum_{\tau=t+1}^{\infty} \delta_l^{\tau-t} (s_l - s_{\tau}) > \lambda_l \left( \delta_h \pi_{h,t+1} - \sum_{\tau=t+1}^{\infty} \delta_h^{\tau-t} (s_{\tau} - \bar{u}) \right)
\]

because \( \lambda_l \) is bounded away from 1, and \( s_h \) and \( s_l \) get closer as \( \delta_h \) and \( \delta_l \) get closer. Hence, it follows that \( \sum_{\tau=t+1}^{\infty} \delta_l^{\tau-t} (s_l - s_{\tau}) > \lambda_l b_t \), a contradiction.

51
B  Online Appendix

B.1  Proofs of Claims in Section 4

Assume that the two principal types, the high type and the low type differ in (and are privately-informed regarding) their productivity and are identical in every other respect. Let $y_\theta$ represent the production function of type-$\theta$ principal, where $\theta \in \{h, l\}$ and $y_h(e) > y_l(e)$. Suppose that there exists a separating equilibrium $\{w_t, b_t\}_{t=0}^\infty$ implementing $e_t$ in period $t \geq 0$.

As before, $C_l = \{w_l, b_l\}$ ( $C_h = \{w_h, b_h\}$ ) denotes the optimal symmetric-information contract of type-$l$ (type-$h$) principal, which implements $e_l$ ($e_h$) in every period. First, note that the optimal separating contract is such that a low type principal who imitates the high type strictly prefers defaulting at some $t \geq 0$ because otherwise the separating contract $\{w_t, b_t\}_{t=0}^\infty$ is one that can be implemented with the low type in a symmetric-information setting and thus, either $\{w_t, b_t\}_{t=0}^\infty$ is such that $\{w_t, b_t\} = C_l$ for every $t$, which is a contradiction, or the imitation payoff of the low type from $\{w_t, b_t\}_{t=0}^\infty$ is strictly lower than $\frac{y_l(e_l) - w_l - b_l}{1-\delta}$; this is also a contradiction because then $\{w_t, b_t\}_{t=0}^\infty$ can be strictly improved upon. In particular, there exists a large enough but finite $T$ such that if the high type starts offering $C_h$ from $T$ onwards, then the imitation payoff of the low type increases by a very small amount—i.e., the imitation payoff is still lower than $\frac{y_l(e_l) - w_l - b_l}{1-\delta}$, and thus, the low type strictly prefers revealing her type and offering $C_l$ (notice that the proof is very similar to the proof of Lemma 5). As a result, the high type is strictly better off, a contradiction. Thus, a low type principal who imitates the high type strictly prefers defaulting at some finite $t \geq 0$. This in turn implies that the high type starts offering $C_h$ in finite time. Thus, there exists a $T < \infty$ such that the high type offers $C_h$ from $T$ onwards in the optimal separating equilibrium. It follows that

$$\frac{y_l(e_l) - w_l - b_l}{1-\delta} \geq \sum_{t=0}^{T-1} \delta^t (y_l(e_t) - w_t - b_t) + \delta^T \left( y_l(e_h) - w_h + \delta \frac{\bar{y}}{1-\delta} \right)$$

is a necessary condition as one of the incentive compatibility constraints which ensure that the low type is deterred from imitation. Adding and subtracting $b_h$ and using the enforcement constraint of the high type (i.e., $b_h = \frac{\delta}{1-\delta} (s_h(e_h) - \bar{u} - \bar{\pi})$), it follows that

$$\frac{y_l(e_l) - w_l - b_l}{1-\delta} \geq \sum_{t=0}^{T-1} \delta^t (y_l(e_t) - w_t - b_t) + \delta^T (y_l(e_h) - w_h - b_h) + \delta^{T+1} s_h(e_h) - \bar{u} \frac{1}{1-\delta}$$

must hold. For the high type to prefer separating, the following must hold.

$$\sum_{t=0}^{T-1} \delta^t (y_h(e_t) - w_t - b_t) + \delta^T (y_h(e_h) - w_h - b_h) + \delta^{T+1} s_h(e_h) - \bar{u} \frac{1}{1-\delta} \geq \frac{y_h(e_l) - w_l - b_l}{1-\delta}.$$

These inequalities imply that

$$\sum_{t=0}^{T-1} \delta^t (y_h(e_t) - y_l(e_t)) + \delta^T (y_h(e_h) - y_l(e_h)) \geq \frac{y_h(e_l) - y_l(e_l)}{1-\delta}.$$
First, observe that this inequality can never hold if \( y_h(e) - y_l(e) = \eta \) for every \( e \) because \( T \) is a finite number, as I explained above. Thus, there exists no separating equilibrium if \( y_h(e) - y_l(e) = \eta \) for every \( e \).

Next, consider the case where \( y_h(e) > y_l(e) \) and \( y_h(e) - y_l(e) \) is increasing in \( e \). Let \( \kappa > 0 \) be such that
\[
\frac{y_h(e_t) - y_l(e_t)}{1 - \delta} < \frac{y_h(e_h) - y_l(e_h)}{1 - \delta} \leq \frac{y_h(e_l) - y_l(e_l)}{1 - \delta} + \kappa.
\]

Suppose towards a contradiction that there exists a separating equilibrium regardless of \( \kappa > 0 \). It follows that, for every \( \kappa > 0 \),
\[
\sum_{t=0}^{T-1} \delta^t (y_h(e_t) - y_l(e_t)) + \delta^T (y_h(e_h) - y_l(e_h)) \geq \frac{y_h(e_l) - y_l(e_l)}{1 - \delta} \geq \frac{y_h(e_h) - y_l(e_h)}{1 - \delta} - \kappa.
\]

But \( y_h(e_t) - y_l(e_t) < y_h(e_h) - y_l(e_h) \) for every \( e_t \) such that \( t < T \) (because \( e_t < e_h \) and \( y_h(e) - y_l(e) \) is increasing in \( e \)). Thus, the inequality above cannot hold for every \( \kappa > 0 \), a contradiction. Notice that in the proof above I assumed that \( T \) is uniformly bounded for all \( \kappa > 0 \) (which may be partly justified since \( T \) is finite in the optimal separating equilibrium for fixed \( y_l \) and \( y_h \)). What if \( \lim_{\kappa \to 0} T(\kappa) = \infty \), where \( T(\kappa) = \min\{t \geq 0 | C_t = C_h \} \) given \( \kappa > 0 \) in the optimal separating contract? I am able to rule this out and show that \( \lim_{\kappa \to 0} T(\kappa) < \infty \) if there is a possibly large but finite number of effort levels, as I discuss in more detail below. However, showing the same with a continuum of effort levels is very difficult. Nevertheless, I am confident that it is impossible to construct a separating equilibrium even if \( \lim_{\kappa \to 0} T(\kappa) = \infty \) since both sides of the inequality below
\[
\sum_{t=0}^{T(\kappa)-1} \delta^t [(y_h(e_t) - y_l(e_t)) - (y_h(e_l) - y_l(e_l))] \geq \delta^{T(\kappa)} \left( \frac{y_h(e_l) - y_l(e_l)}{1 - \delta} - (y_h(e_h) - y_l(e_h)) \right)
\]
converge to zero as \( \kappa \to \infty \), and yet it is impossible to make sure that the right-hand side converges to zero faster than the left-hand side because not only \( (y_h(e_t) - y_l(e_t)) - (y_h(e_l) - y_l(e_l)) \) converges to zero as \( \kappa \to 0 \) but also for fixed \( t \geq 0 \), \( \lim_{\kappa \to 0} e_t^\kappa \leq e_t \) holds, which increases the convergence rate of the left-hand side (\( \lim_{\kappa \to 0} e_t^\kappa \leq e_t \) must hold for every \( t \geq 0 \) if \( \lim_{\kappa \to 0} T(\kappa) = \infty \); otherwise, I can show that \( T(\kappa) \) is uniformly bounded above by some \( \bar{T} < \infty \)). These issues do not arise if there is a large but finite number of effort levels; in that case, I can show that \( T \) is uniformly bounded above for all \( \kappa > 0 \) (assuming that a separating equilibrium exists). Thus, I can also show that separation is not generally possible even if \( y_h(e) > y_l(e) \) and \( y_h(e) - y_l(e) \) is increasing in \( e \). However, there always exists a separating equilibrium if types differ in their time preferences, and this is still true with discrete effort levels (the proof of Proposition 3 does not rely on the existence of a continuum of effort levels).
B.2 Dynamic Intuitive Criterion (DIC)

In this part, I will provide a detailed discussion of DIC, and I will explain how hybrid contracts can be eliminated using DIC. Let \( C_t = \{ w_t, b_t \} \) describe the period-\( t \) hybrid equilibrium contract that promises to pay \( b_t \) and implements \( e_t \). Fix an arbitrary hybrid equilibrium \( \{ C_t \}_{t=0}^\infty \). Assume that information revelation is complete with probability one at the end of period \( T \). To be more precise, let \( T = \min\{t \geq 0 | \mu_t = 0 \} \). Thus, if the low type principal honored all the promised payments up until period \( T \), then \( \lambda_T = \mu_T < 1 \) and \( \mu_{T+1} \in \{0,1\} \).

Assume that \( T > 0 \) for now. For simplicity, I focus on equilibria where the high type offers \( C_h \) from \( T + 1 \) onwards; i.e., \( \{ C_t \}_{t=T+1}^\infty = \{ C_h, C_h, \ldots \} \). Let \( C_{T-1} = \{ w_{T-1}, b_{T-1} \} \), where \( e_{T-1} \) is the equilibrium effort level. Consider the deviation \( \{ D_t \}_{t \geq T-1} \) such that (i)-(iii) hold:

(i) \( D_{T-1} = \{ w'_t, b_{T-1} \} \), where \( b_{T-1} \) is contingent on \( e_{T-1} \) (thus, \( b_{T-1} \) and \( e_{T-1} \) in \( D_{T-1} \) are the same as in contract \( C_{T-1} \)),

\[
w'_t = w_{T-1} + \frac{\delta_t + \delta_h}{2},
\]

and the term \( \Delta \) is derived as follows. Let \( \Delta = y(e'_T) - y(e_T) > 0 \), where \( e_T \) is the effort level implemented by \( C_T = \{ w_t, b_t \} \), and \( e'_T \) is described below in part (ii).

(ii) If the offer \( D_{T-1} \) is accepted, and the agent exerts effort \( e_{T-1} \), then \( D_T = \{ w_t, b_t \} \) is offered. Thus, \( w_T \) and \( b_T \) are the same as in the equilibrium hybrid contract \( C_T \), but, unlike in \( C_T \), the bonus payment \( b_T \) rewards \( y(e'_T) \), where \( e'_T \) is the highest possible effort level that satisfies both \( c(e) \leq b_T \) and \( w_T + b_T - c(e) \geq \bar{u} \). Given this, \( e'_T > e_T \) must hold (assuming for the moment that the agent believes that only the high type principal deviates to \( \{ D_t \}_{t \geq T-1} \) because \( \mu_T < 1 \) and \( \{ C_t \}_{t=T+1}^\infty = \{ C_h, C_h, \ldots \} \), whereas if only the high type principal deviates to \( \{ D_t \}_{t \geq T-1} \) then \( D_T \) and \( b_T \) will be honored with probability one.\(^{45}\)

(iii) \( \{ D_t \}_{t \geq T+1} = \{ C_h, C_h, \ldots \} \). That is, if the agent accepts the offer \( D_T \), and exerts effort \( e'_T \), then the principal offers \( C_h \) from \( t = T + 1 \) onwards just as in the original hybrid contract \( \{ C_t \}_{t=0}^\infty \). Note that if the agent believes that the deviation \( \{ D_t \}_{t \geq T-1} \) comes from a high type, then \( e_{T-1} \) and \( e'_T \) are incentive compatible.

It is easy to show that the deviation \( \{ D_t \}_{t \geq T-1} \) is equilibrium-dominated for the low type even if the agent chooses \( e_{T-1} \) and \( e'_T \). To see why, note that at \( T - 1 \),

\[
\pi_{t,T-1} = y_{T-1} - w_{T-1} - b_{T-1} + \max\{b_{T-1} + \delta_t \pi^d_{t,T}, \delta_t \pi_{t,T} \}
\]

given \( \{ C_t \}_{t=0}^\infty \), where \( \pi^d_{t,T} \) represents the equilibrium punishment payoff for the low type after defaulting at period \( T - 1 \). However, the deviation to \( \{ D_t \}_{t \geq T-1} \) gives the low type a

\(^{44}\)In a hybrid equilibrium in which either \( T = \infty \) or \( T < \infty \) but behavior distortion of the high type continues after beliefs have become degenerate, one can still find a deviation path that would make the high type strictly better off and the low type worse off using arguments similar to those presented below.

\(^{45}\)I assume that \( b_T > 0 \) without loss of generality. Otherwise, I set \( b_T = \epsilon \), and \( c(e'_T) = \epsilon \) for small \( \epsilon > 0 \), and \( \Delta = y(e'_T) \) in the deviation contract.
maximum possible payoff of

$$y_{t-1} - w'_{t-1} - b_{t-1} + \max\{b_{t-1} + \delta_l \pi_{l,T}^d, \delta_l (\pi_{l,T}^i + \Delta)\},$$

which is strictly lower than $\pi_{l,T-1}$ ($y_{t-1}$, $b_{t-1}$ and $v_{t-1}$ are the same across $C_{T-1}$ and $D_{T-1}$ whereas $w'_{t-1} > w_{t-1} + \delta_l \Delta$ by construction). Moreover, $\{D_t\}_{t \geq T-1}$ is enforceable for the high type at every $t \geq T - 1$, and since

$$y_{t-1} - w'_{t-1} - b_{t-1} + \delta_h(y_{T} + \Delta - w_{T} - b_{T}) > y_{t-1} - w_{t-1} - b_{t-1} + \delta_h(y_{T} - w_{T} - b_{T})$$

holds, the high type is strictly better off—assuming that the agent believes that the bonus will be honored with probability one at every $t \geq T - 1$. Thus, it is always possible to find a deviation path such that a hybrid equilibrium $\{C_t\}_{t=0}^\infty$ is not robust to DIC provided that $T > 0$ in $\{C_t\}_{t=0}^\infty$. What happens if $T = 0$ so that $\mu_1 \in \{0, 1\}$? In that case, it is possible to rule out every hybrid equilibrium as being unreasonable unless $b_0$ in $C_0$ is such that

$$b_0 = \delta_l \left(s_h - \bar{u} + b_h + \frac{\delta_l}{1 - \delta_l} \bar{\pi}\right)$$

holds.$^{46}$ Note that such a hybrid equilibrium is undominated by a separating equilibrium only if $\mu_0 > \frac{b_l}{b_0}$, which is a more stringent condition than $\mu_0 > \frac{b_l}{b_0}$ as $b_0 < b_h$. Thus, such an equilibrium is not robust to DIC if $\mu_0 \leq \frac{b_l}{b_0}$. If $\mu_0 > \frac{b_l}{b_0}$, then whether or not this is a reasonable equilibrium depends on the precise parameters of the game as well as the cost and the production functions.

Now, I provide a discussion of the Dynamic Intuitive Criterion. Unlike the case with the standard Intuitive Criterion, which is typically applied to one-shot games, one concern is that when a deviation contract is observed in some period, the complete path of deviation $\{D_t\}_{t \geq k}$ is not yet fully observed (although $\{D_t\}_{t \geq k}$ is announced by the principal, this is only cheap talk with the exception of the fixed wage in $D_k$). So, what should the agent infer from a single deviation $D_k$ when $\{D_t\}_{t \geq k}$ is not yet fully observable? Note that the type of the principal does not enter the payoff function of the agent directly. What matters for the agent is only $\lambda_l$, the probability with which a payment promise is fulfilled; the type of the principal matters only indirectly and due to its implication regarding the probability, $\lambda_l$. I explore the inference on this probability given a deviation, which I denote by $\lambda_l$.

Assume that the principal deviates for the first time at an arbitrary period $k \geq 0$. For simplicity of the argument below, I focus on deviation contracts $\{D_t\}_{t \geq k}$ such that $c(e_l') \leq b_l'$ and $w'_{t} + b_{l}' - c(e_l') \geq \bar{u}$ hold at every $t \geq k$.\(^{47}\) Otherwise, I assume that the agent rejects $D_t$.

\(^{46}\)Note that since $T = 0$,

$$b_0 \geq \delta_l \left(s_h - \bar{u} + b_h + \frac{\delta_l}{1 - \delta_l} \bar{\pi}\right)$$

must hold. If this holds with strict inequality, then the deviation contract $D_0 = \{w'_0, b'_0\}$ and $\{D_t\}_{t \geq 1} = \{C_h, C_h, \ldots\}$ where $b'_0 = b_0 \mu_0$ and $w'_0 = w_0 + (1 - b_0 \mu_0) - \epsilon$ does the job provided that $\epsilon > 0$ is small enough.

\(^{47}\)Thus, it can be checked easily whether the agent’s participation and incentive compatibility constraints are satisfied in the deviation contract.
Second, a prerequisite for \( \{D_t\}_{t \geq k} \) to come from a high type principal is that every contract \( D_t \) in \( \{D_t\}_{t \geq k} \) is enforceable for the high type.

The inference of the agent following a deviation at \( k \) is determined based on the equilibrium dominance concept as follows: Does a low type principal benefit (relative to her equilibrium payoff from \( \{C_t\}_{t=0}^{\infty} \) from offering \( D_k = \{w'_k, b'_k\} \), paying \( w'_k \) to the agent, obtaining output \( y(e'_k) \) and then defaulting on \( b'_k \)? If the answer to this question \( Q^k \) is no, then the agent infers that \( \hat{\lambda}_k = 1 \), accepts the offer, and chooses his effort level in accordance with \( D_k \). If \( b'_k \) is honored and \( D_{k+1} \) is offered at \( k+1 \), then the agent asks the question \( Q^k_{k+1} \): Does the low type benefit from offering \( D_{k+1} \) at \( k+1 \), paying \( w'_{k+1} \) to the agent, obtaining output \( y(e'_{k+1}) \) and then refusing to pay \( b'_{k+1} \), having offered and honored \( D_k \)? If the answer is again no, then the agent infers that \( \hat{\lambda}_{k+1} = 1 \), accepts the offer, and chooses his effort level accordingly. Inductively, let \( Q^k \) stand for the following question: Does the low type benefit from offering \( D_t \) at \( t > k \), paying \( w'_t \), obtaining output \( y(e'_t) \) and then defaulting on \( b'_t \) having offered and honored \( D_k \), \( D_{k+1}, \ldots, D_{t-1} \) ? If the answer is again no, then the agent accepts the offer, infers that \( \hat{\lambda}_k = 1 \) and chooses his effort level accordingly.

Let’s fix an arbitrary equilibrium \( \{C_t\}_{t=0}^{\infty} \). If there exists no \( k \geq 0 \) and \( \{D_t\}_{t \geq k} \) such that:

(i) the answer to question \( Q^k_t \) results in inferring \( \hat{\lambda}_t = 1 \) for every \( t \in \{k, k+1, \ldots\} \) (which ensures that the low type cannot benefit from the deviation path at any \( t \geq k \)), and

(ii) the high type is strictly better off with \( \{D_t = \{w'_t, b'_t, 1, e'_t\}\}_{t \geq k} \) where effort \( e'_t \) is incentive compatible and \( b'_t \) is enforceable at every \( t \in \{k, k+1, \ldots\} \),

then equilibrium \( \{E_t\}_{t=0}^{\infty} \) is “reasonable”. Otherwise, it is not reasonable. But these conditions are equivalent to the definition of DIC given in the main text.

### B.3 Proofs of Claims in Section 7.1

I start the analysis with the case in which type-\( g \) is more able and has a lower cost of effort than type-\( b \) in a way that \( c^b_H - c^g_H \leq c^b_M - c^g_M \), where \( c^e_i \) denotes the cost of effort \( e \in \{M, H\} \) for type-\( i \) seller, \( i \in \{g, b\} \). As stated in the main text, I focus on the nontrivial case in which type-\( g \) has an incentive to separate himself from type-\( b \) (equivalently, type-\( b \) has an incentive to imitate type-\( g \)). This would be true if for example type-\( g \) can be induced to exert high effort in a symmetric-information setting with relational incentives whereas type-\( b \) can only be induced to exert medium effort (thus, type-\( g \) is similar to the high type principal and type-\( b \) is similar to the low type principal in the main model). From now on, I will focus on such parameter values. With a type-\( g \) seller, this translates to the following enforcement constraints in a symmetric-information setting:

\[
p_H + \frac{\delta}{1 - \delta} \bar{\pi} \leq \Phi p_H + (1 - \Phi) p_M - c^g_H,
\]

and

\[
p_H - c_H + \frac{\delta}{1 - \delta} \bar{\pi} \leq p_M - c_H + \delta \left( \Phi p_H + (1 - \Phi) p_M - c^g_H \right).
\]
where $p_M$ and $p_H$ denote the equilibrium product price set given $q = q_M$ and $q = q_H$, respectively. I follow the literature in assuming that prices will be bid up to the respective buyer valuations given their beliefs. It follows that $p_M = u_M$ and $p_H = u_H$ given that a seller is “truthful” in his quality announcement. I also follow the literature in assuming that if the seller trades with short-lived buyers, then past quality realizations of a seller are perfectly observable (although they are not verifiable). The term $\pi$ denotes the per period punishment payoff of the seller if the seller is dishonest in his message regarding quality. Given these, the first constraint implies that the type-$g$ seller does not benefit from exerting low effort, falsely claiming that $q = q_H$ and selling a low-quality product at a rip-off price, $p_H$. The second constraint implies that the seller does not benefit from exerting high effort and falsely claiming that $q = q_H$ when in fact $q = q_M$. A reasonable assumption regarding the punishment payoff is that $\pi = u_L$. Buyers who observe the dishonesty of a seller believe that the dishonest seller will always exert low effort from then onwards and hence they will not pay more than $u_L$ for his product. For simplicity, and without loss of generality, I assume that $\pi = u_L = 0$ from now on. Assuming that $\pi = u_L = 0$ and that type-$b$ seller can be induced to exert medium effort, the following must hold:

$$p_M \leq \delta \frac{\Phi p_M - c_M^b}{1 - \delta}.$$  

This constraint implies that type-$b$ seller will not benefit from (i) exerting low effort and falsely claiming that $q = q_M$, and (ii) exerting medium effort and falsely claiming that $q = q_M$ when in fact $q = q_L$. Since type-$b$ seller cannot be motivated to exert high effort, either

$$p_H > \frac{\Phi p_H + (1 - \Phi)p_M - c_H^b}{1 - \delta},$$

or

$$p_H - c_H > p_M - c_H + \delta \frac{\Phi p_H + (1 - \Phi)p_M - c_H^b}{1 - \delta},$$

or both will hold. I assume that $c_H^b$ is high enough so that the first inequality holds; hence,

$$p_H > \frac{\Phi p_H + (1 - \Phi)p_M - c_H^b}{1 - \delta}. \quad (13)$$

If these conditions above about type-$b$ and type-$g$ sellers are satisfied, then type-$b$ seller can be motivated to exert medium effort—but not high effort—and type-$g$ seller can be motivated to exert both high effort and medium effort in a symmetric-information setting.\footnote{This may be because buyers engage in Bertrand competition or the seller runs an auction. See Tadelis (1999), Mailath and Samuelson (2001) and Jullien and Park (2014).}

Now, consider the private information setting. I first show that there is no separating equilibrium if $c_H^b - c_H^g \leq c_M^b - c_M^g$. Intuitively, comparing the benefit of separation for the high type to the benefit of imitation for the low type is enough to see that there can be no separation. The benefit of separation for the high type is equal to $\Phi(p_H - p_M) + (1 - \Phi)p_M - (c_H^g - c_M^g)$.

\footnote{It is straightforward to verify that there exist parameter values $c_M^b$, $c_H^b$, $p_H$, $p_M$ and $\Phi$ such that all of the conditions are satisfied.}
whereas the benefit of imitation for the low type is

\[
\max \left\{ p_H, \frac{p_H - c_H^b}{1 - \Phi\delta} \right\} - \frac{\Phi p_M - c_M^b}{1 - \delta},
\]

which is strictly greater than \( \frac{\Phi(p_H - p_M) + (1 - \Phi)p_M - (c_H^g - c_M^g)}{1 - \delta} \) due to (13). The cost of advertising (or of another type of money burning) is however the same for both types. Notice that this argument can still hold even if \( c_H^b - c_H^g > c_M^b - c_M^g \). Thus, the condition \( c_H^b - c_H^g > c_M^b - c_M^g \) is also not sufficient for separation.

Formally, let \( a_t \) denote the cost of advertising in period \( t \geq 0 \) — I assume that \( a_t \) is undertaken at the beginning of period \( t \geq 0 \) before trade. I will first show that it is never possible to find a sequence \( \{a_t\}_{t=0}^{\infty} \) that separates the two types if \( c_H^b - c_H^g \leq c_M^b - c_M^g \). In a separating contract, the high type is willing to separate himself (rather than pooling with type-\( b \)) as long as

\[
1 \sum_{t=0}^{\infty} \delta^t (\mathbb{E}(p_t|e_t) - c_{e_t}^b - a_t) \geq \frac{\Phi p_M - c_M^g}{1 - \delta}
\]

where \( \mathbb{E}(p_t|e_t) \) and \( c_{e_t}^b \) denote the “truthful” expected equilibrium price and cost given equilibrium effort \( e_t \) in period \( t \), respectively. Note that this implies that

\[
1 \sum_{t=0}^{\infty} \delta^t (\mathbb{E}(p_t|e_t) - c_{e_t}^b - a_t) \geq \frac{\Phi p_M - c_M^b}{1 - \delta}
\]

since \( c_H^b - c_H^g \leq c_M^b - c_M^g \). I will now show that this inequality must be strict, which will imply that the low type cannot be deterred from imitation and thus separation is impossible. Given that seller types are “truthful” in their quality announcement, prices will be bid up to the respective buyer valuations given their beliefs, and thus, \( p_t \in \{p_L, p_M, p_H\} \) in every \( t \geq 0 \). Thus, in the separating contract \( p_t = p_L \) if the seller announces \( q = q_L \), \( p_t = p_M \) if the seller announces \( q = q_M \), and \( p_t = p_H \) if the seller announces \( q = q_H \). To see why the inequality above is strict, assume that \( a_t > 0 \) for at least one \( t \geq 0 \); otherwise, there is no costly signaling and type-\( g \) seller cannot separate. Given this, there must exist a period \( T < \infty \) such that \( e_T = H \) is prescribed in the separating contract (otherwise, costly advertising is wasteful and type-\( g \) would prefer imitating a type-\( b \) seller). In period \( T \), type-\( b \) seller who imitates a type-\( g \) seller can cheat by, for example, exerting low effort and announcing that \( q = q_H \). Thus, the payoff of a type-\( b \) seller who imitates a type-\( g \) seller is (weakly) greater than

\[
1 \sum_{t=0}^{T-1} \delta^t (\mathbb{E}(p_t|e_t) - c_{e_t}^b - a_t) + p_H - a_T.
\]

\(^{50}\) Of course, the separating contract must be enforceable for the high type; i.e., the quality announcement must be truthful. However, these constraints are not needed for the result.
But $p_H - a_T$ is strictly greater than

$$\sum_{t=T}^{\infty} \delta^{t-T} \left( \mathbb{E}(p_t | e_t) - c_{e_t}^b - a_t \right)$$

because $\mathbb{E}(p_t | e_t)$ is, by definition, the truthful expected equilibrium price and thus, $\sum_{t=T}^{\infty} \delta^{t-T} \times (\mathbb{E}(p_t | e_t) - c_{e_t}^b - a_t)$ is weakly smaller than $\frac{\Phi p_H + (1 - \Phi) p_M - c_H}{1 - \delta} - \sum_{t=T}^{\infty} a_t$, which in turn is strictly smaller than $p_H - a_T$ due to (13). Hence, separation is not possible. Next, I show that $c_H - c_H^g > c_M^b - c_M^b$ is not a sufficient condition for the existence of a separating equilibrium.

To see why, assume that $p_H$ is either equal or close to $\frac{\Phi p_H + (1 - \Phi) p_M - c_H^g}{1 - \delta}$. Steps similar to those above show that in that case a separating equilibrium does not exist (simply note that delaying high-effort/high-quality production is not less costly to the type-$g$ seller, thus delay is not an effective signaling tool). Hence, the condition $c_H - c_H^g > c_M^b - c_M^b$ is not sufficient. A similar analysis shows that separation is not generally possible in the case where the two types differ in $\Phi$ such that $\Phi^g > \Phi^b$. The formal proof of this claim follows very similar steps to those above in the case with differential effort costs and is available upon request.

If the two types differ in their discount factors, then there always exists a separating equilibrium. The construction in the proof of Proposition 3 can be directly applied to show that there exists a separating equilibrium. Let $\delta_i$ represent the discount factor of a type-$i$ seller, where $i \in \{ g, b \}$ and $\delta_g > \delta_b$. Let $T \geq 0$ be the smallest integer $t \geq 0$ such that

$$\delta_t \left( \Phi\left(\frac{p_H - p_M}{1 - \delta_g}\right) + \frac{(1 - \Phi) p_M - (c_H - c_M)}{1 - \delta_g} \right) > \delta_t^b \left( \max \left\{ p_H, \frac{p_H - c_H}{1 - \delta_b} \right\} - \frac{\Phi p_M - c_M}{1 - \delta_b} \right)$$

holds. Since $\delta_g > \delta_b$, it follows that $T < \infty$. Set $a_0$ (advertising at $t = 0$) equal to

$$a_0 = \delta_t^b \left( \max\left\{ p_H, \frac{p_H - c_H}{1 - \delta_b} \right\} - \frac{\Phi p_M - c_M}{1 - \delta_b} \right),$$

and $a_t = 0$ for all $t > 0$. Then, the separating equilibrium is as follows. Type-$g$ seller chooses advertising at $t = 0$ equal to $a_0$ and exerts medium effort ($e_t = M$) at every $t < T$. From period $T$ onwards, the type-$g$ seller exerts high effort—i.e., $e_t = H$ at every $t \geq T$. Type-$b$ seller chooses zero advertising and exerts medium effort ($e_t = M$) at every $t \geq 0$. Finally, quality announcement is truthful in every period with both types of sellers. This simple construction separates the two types.

The main result also extends to a market setting where sellers enter and exit the economy stochastically, and names can be traded, as modeled in Tadelis (1999) and Mailath and Samuelson (2001). In addition to the trading of names, I allow for name changes; for example, an existing type-$b$ firm with a bad reputation can try to erase the public memory about his type by choosing a new name. I also maintain the assumption in Tadelis (1999) and Mailath and Samuelson (2001) that changes in names’ ownership are unobservable.

Under these assumptions a separating equilibrium still exists, provided that type-$g$ sellers
have a sufficiently high discount factor.\footnote{In particular, I show that there exists a separating equilibrium such that good names are bought only by good sellers because they are too expensive for bad sellers.} To show why this is the case, I first spell out the assumptions of this model. As in Tadelis (1999), I assume that there is a continuum of sellers and buyers. Sellers enter and exit the economy in a way that the size of the seller population and the distribution of seller types are constant over time. In each period, a seller exits the market with probability $1 - \phi$ and the measure of sellers that exit the market is replaced by an identical measure of new sellers that enter the market. I assume that as a firm with a good name exits the market, the firm sells its name to another firm. The name of a firm that has cheated once (by a deceptive message regarding product quality) is worthless as it signals a bad type. I also assume that a small proportion $\zeta > 0$ of entrants cannot buy a name (this is a reduced-form assumption which illustrates the case where some new firms are credit-constrained and therefore unable to purchase an established name).

Good names will be scarce in my construction since good names will be owned only by type-$g$ firms but everyone wants a good name. As a result, the price of a good name will be bid up to a point where the highest bidder is indifferent between buying the name and not buying it. The construction will be such that the highest bidders are new type-$g$ firms, and hence good names are bought only by good types. Note that a new type-$g$ firm who cannot buy a name (because $\zeta > 0$) may build its name by advertising. I assume that an already-established, good name has an infinitesimal advantage over building up a new name. For example, new advertising campaigns might have the risk of being unpopular or unable to reach prospective buyers (with an infinitesimal probability) whereas an existing name is already established. As a result, a new type-$g$ firm prefers obtaining an already existing good name over building up a new name. If the new firm buys a name, then it will spend an amount $p_s$ on this name, which is discussed in more detail below. As before, the benefit of imitation for a type-$b$ seller is $\max \left\{ p_H, \frac{p_H - c_H}{1 - \phi \delta h} \right\} - \Phi \frac{p_M - c_M}{1 - \phi \delta h}$, whereas the benefit of separation for a type-$g$ seller is $\Phi \frac{(p_H - p_M) + (1 - \Phi) p_M - (c_H - c_M)}{1 - \phi \delta h}$. Given $p_s$, the sale price of a good name, the “net” benefit of separation for a type-$g$ firm is

$$\frac{\Phi (p_H - p_M) + (1 - \Phi) p_M - (c_H - c_M)}{1 - \phi \delta h} - p_s + \frac{\delta_h (1 - \phi)}{1 - \phi \delta h} p_s,$$

where the last term follows because the firm can resell its name to a new firm upon exiting the market. This in turn implies that the “net” benefit of separation for a type-$g$ firm is positive provided that the following holds:

$$p_s \leq \frac{\Phi (p_H - p_M) + (1 - \Phi) p_M - (c_H - c_M)}{1 - \delta_h}.$$

Next, note that $p_s$ must be such that $p_s \geq \max \left\{ p_H, \frac{p_H - c_H}{1 - \phi \delta h} \right\}$ so that a type-$b$ firm is deterred from buying a good name. Thus, as long as $\delta_h$ is sufficiently high, a separating equilibrium exists because there exists a $p_s$ that satisfies both constraints.\footnote{The cost of advertising for new type-$g$ firms that are unable to buy a name is equal to $p_s$.} This in turn implies that a good name will not go bad because it will be too expensive for type-$b$ sellers.
B.4 Proof of Proposition 9

First, note that Lemma 5 also holds in the setting with stochastic output and unobservable effort. In fact, behavior distortion must be over at a finite $T$ in the “constrained” optimal contract, as well. Suppose not. Then, there are two possibilities. Either $b_t^*$ is bounded above away from $b_h$, or $b_t^*$ (or a subsequence) converges to $b_h$, where $b_t^*$ denotes the bonus at $t$ in the optimal contract. If $b_t^*$ is bounded above away from $b_h$, then the proof of Lemma 5 applies directly to show that this is a contradiction, and that costly signaling must end at a finite time. Next, consider the case in which $b_t^*$ (or a subsequence, which I also denote by $b_t^*$) converges to $b_h$. Note that for all $t$ sufficiently large,

$$b_t^* = \delta_h \pi_{h,t+1}^* - \frac{\delta_h}{1 - \delta_h} \bar{\pi}$$

must hold. If $t$ is large, and

$$b_t^* < \delta_h \pi_{h,t+1}^* - \frac{\delta_h}{1 - \delta_h} \bar{\pi}$$

then the high type can make a gainful increase in $b_t$, which increases her payoff by more than the imitation payoff of the low type. To see why, first note that

$$b_t^* > \delta_l \pi_{l,t+1}^i - \frac{\delta_l}{1 - \delta_l} \bar{\pi},$$

for all sufficiently large $t$. This is because $b_t^* \to b_h$, and

$$b_h > \delta_l V_l - \frac{\delta_l}{1 - \delta_l} \bar{\pi}.$$ 

As a result, at a sufficiently large $t$, the low type will have defaulted with a very high probability, which reduces drastically the impact of an increase in $b_t$ on the imitation payoff of the low type. Thus,

$$b_t^* = \delta_h \pi_{h,t+1}^* - \frac{\delta_h}{1 - \delta_h} \bar{\pi}$$

for all sufficiently large $t$. Since $\pi_{h,t+1} = (s_{t+1} - u) + b_{t+1}^* + \frac{\delta_h}{1 - \delta_h} \bar{\pi}$, $b_t^* < b_{t+1}^*$ must hold for every sufficiently large $t$. Otherwise, $b_t^* \geq b_{t+1}^*$ implies that $\pi_{h,t+1} \geq \pi_{h,t+2}$ and that $b_{t+1}^* \geq b_{t+2}^*$ for every sufficiently large $t$. But this is a contradiction. Since $b_t^* < b_{t+1}^*$, it follows that

$$\pi_{l,t+1}^i > \frac{(s_t - \bar{u}) + e_t \left(b_t^* + \frac{\delta_l}{1 - \delta_l} \bar{\pi}\right)}{1 - \delta_l (1 - e_t)}.$$

As a result,

$$V_l - \pi_{l,t+1}^i < \frac{(s_h - \bar{u}) + e_h \left(b_h^* + \frac{\delta_l}{1 - \delta_l} \bar{\pi}\right)}{1 - \delta_l (1 - e_h)} - \frac{(s_t - \bar{u}) + e_t \left(b_t^* + \frac{\delta_l}{1 - \delta_l} \bar{\pi}\right)}{1 - \delta_l (1 - e_t)}.$$
follows that

(i) First, assume that \( b \) the high type principal defaults (recall that \( b \) holds for all \( t \geq 1 \) in the initial period since \( u_t = \frac{\delta}{1-\delta} \) must hold for all \( t \geq 1 \) in the constrained optimal contract). I start with the final period of costly signaling. Let \( T \) denote the last period such that \( b^*_t \neq b_h \). Thus, \( b^*_t = b_h \) for all \( t \geq T \). First, note that \( b^*_{T-1} < b_T = b_h \) by the definition of \( T \). Otherwise, \( b^*_{T-1} > b_T \) in which case the high type principal defaults (recall that \( b^*_{T-1} \neq b_h \) by hypothesis). Next, I show that \( b^*_{T-2} < b^*_{T-1} \). Suppose towards a contradiction that \( b^*_{T-2} \geq b^*_{T-1} \). Since \( b^*_{T-1} < b_T = b_h \) it follows that \( \pi^i_{t,T-1} \leq \pi^i_{t,T} = V_t \).

There are two cases to consider:

(i) First, assume that

\[
b^*_{T-2} \leq \delta_l \pi^i_{t,T-1} - \frac{\delta_l}{1-\delta_l} \tilde{\pi}.
\]

Since \( \pi^i_{t,T-1} < \pi^i_{t,T} \) it follows that

\[
b^*_{T-1} \leq b^*_{T-2} \leq \delta_l \pi^i_{t,T} - \frac{\delta_l}{1-\delta_l} \tilde{\pi}.
\]

However,

\[
b^*_{T-1} = \delta_l \pi^i_{t,T} - \frac{\delta_l}{1-\delta_l} \tilde{\pi}
\]

cannot hold in the optimal contract, just as I argued in the proof of Proposition 6. Thus, \( b^*_{T-1} > b^*_{T-2} \), a contradiction.

(ii) Next, assume that

\[
b^*_{T-2} > \delta_l \pi^i_{t,T-1} \leq \delta_l \pi^i_{t,T} - \frac{\delta_l}{1-\delta_l} \tilde{\pi}.
\]

Assume towards a contradiction that \( b^*_{T-2} \geq b^*_{T-1} \). Consider the modified contract: \( b^*_{T-2} \) is

\[
< \left( \frac{s_h + e_h \left( b^*_h + \frac{\delta_t}{1-\delta_t} \bar{\pi} \right)}{1-\delta_t} \right) - \left( s_t + e_t \left( b^*_t + \frac{\delta_t}{1-\delta_t} \bar{\pi} \right) \right),
\]

where the last inequality uses the fact that \( e_h > e_t \). Moreover, \( \frac{s_h}{1-b_h} - \pi_{h,t+1} = \frac{b_h-b^*_h}{b_h} \). Thus,

\[
\lim_{t \to \infty} \frac{(V_t - \pi^i_{t,t+1})}{b_h - b^*_h} \leq \lim_{b_h - b^*_h} \frac{\delta_h \left[ \left( s_h + e_h \left( b^*_h + \frac{\delta_t}{1-\delta_t} \bar{\pi} \right) \right) - \left( s_t + e_t \left( b^*_t + \frac{\delta_t}{1-\delta_t} \bar{\pi} \right) \right) \right]}{(1-\delta_t) (b_h - b^*_t)} \in (0, \infty)
\]

by L’Hopital’s rule because \( s_t \) and \( e_t \) can be written as a function of \( b_t \). Using this finding, the fact that \( \delta_t < \delta_h \) and the fact that the low type will have already defaulted with a very high probability at large \( t \), one can see that the benefit of high type from proposing \( C^*_h \) at sufficiently large \( t \) must be higher than the increase in the imitation of the low type.\(^{53}\)

Thus, there is a modified separating contract that makes the high type strictly better off, a contradiction.

Next, I prove the statement about the gradually increasing pattern of \( b^*_t \). After the initial period the only costly signaling device is the offer of a sufficiently low bonus (costly signaling in the form of a high fixed wage can be used only in the initial period since \( u_t = \frac{\delta}{1-\delta} \) must hold for all \( t \geq 1 \) in the constrained optimal contract). I start with the final period of costly signaling. Let \( T - 1 \) denote the last period such that \( b^*_t \neq b_h \). Thus, \( b^*_t = b_h \) for all \( t \geq T \). First, note that \( b^*_{T-1} < b_T = b_h \) by the definition of \( T \). Otherwise, \( b^*_{T-1} > b_T \) in which case the high type principal defaults (recall that \( b^*_{T-1} \neq b_h \) by hypothesis). Next, I show that \( b^*_{T-2} < b^*_{T-1} \). Suppose towards a contradiction that \( b^*_{T-2} \geq b^*_{T-1} \). Since \( b^*_{T-1} < b_T = b_h \) it follows that \( \pi^i_{t,T-1} < \pi^i_{t,T} = V_t \).

\(^{53}\)The agent’s incentives are not affected since Lemma 4 holds in this setting.
decreased slightly whereas \( b_{T-1}^* \) is increased in a way that

\[
\frac{\partial \pi_{l,0}^i}{\partial b_{T-2}} \bigg|_{b_{T-2}=b_{T-2}^*} + \frac{\partial \pi_{l,0}^i}{\partial b_{T-1}} \bigg|_{b_{T-1}=b_{T-1}^*} \frac{\partial b_{T-1}}{\partial b_{T-2}} = 0
\]

(14)

holds. Note that a small increase in \( b_{T-1}^* \) is enforceable for the high type since the surplus is always \( s_h \) after \( T - 1 \), and \( b_{T-1}^* < b_h \). Next, I show that

\[
\frac{\partial \pi_h}{\partial b_{T-2}} \bigg|_{b_{T-2}=b_{T-2}^*} + \frac{\partial \pi_h}{\partial b_{T-1}} \bigg|_{b_{T-1}=b_{T-1}^*} \frac{\partial b_{T-1}}{\partial b_{T-2}} < 0.
\]

(15)

This will establish a contradiction as it implies that the high type principal can decrease \( b_{T-2}^* \) and increase \( b_{T-1}^* \) slightly, increasing her payoff and keeping the imitation payoff of the low type same. From (14) and the fact that

\[
b_{T-1}^* \geq \delta_t \pi_{l,T}^i - \frac{\delta_t}{1 - \delta_t} \bar{\pi},
\]

it follows that

\[
\frac{\partial b_{T-1}}{\partial b_{T-2}} = -\frac{\partial \pi_{l,0}^i}{\partial b_{T-2}} \bigg|_{b_{T-2}=b_{T-2}^*} \frac{\partial \pi_{l,0}^i}{\partial b_{T-1}} \bigg|_{b_{T-1}=b_{T-1}^*} \frac{\partial b_{T-1}}{\partial b_{T-2}}
\]

\[
= -\frac{\delta_t}{1 - \delta_t} \frac{\partial \pi_{l,0}^i}{\partial b_{T-2}} \bigg|_{b_{T-2}=b_{T-2}^*} + \frac{\delta_t}{1 - \delta_t} \frac{\partial \pi_{l,0}^i}{\partial b_{T-1}} \bigg|_{b_{T-1}=b_{T-1}^*} + \frac{\delta_t}{1 - \delta_t} \frac{\partial \pi_{l,0}^i}{\partial b_{T-2}} \bigg|_{b_{T-2}=b_{T-2}^*} + \frac{\delta_t}{1 - \delta_t} \frac{\partial \pi_{l,0}^i}{\partial b_{T-1}} \bigg|_{b_{T-1}=b_{T-1}^*}
\]

\[
+ \frac{\delta_t}{1 - \delta_t} \frac{\partial \pi_{l,0}^i}{\partial b_{T-2}} \bigg|_{b_{T-2}=b_{T-2}^*} + \frac{\delta_t}{1 - \delta_t} \frac{\partial \pi_{l,0}^i}{\partial b_{T-1}} \bigg|_{b_{T-1}=b_{T-1}^*} + \frac{\delta_t}{1 - \delta_t} \frac{\partial \pi_{l,0}^i}{\partial b_{T-2}} \bigg|_{b_{T-2}=b_{T-2}^*} + \frac{\delta_t}{1 - \delta_t} \frac{\partial \pi_{l,0}^i}{\partial b_{T-1}} \bigg|_{b_{T-1}=b_{T-1}^*}
\]

\[
= -\frac{\delta_t}{1 - \delta_t} \frac{\partial \pi_{l,0}^i}{\partial b_{T-2}} \bigg|_{b_{T-2}=b_{T-2}^*} + \frac{\delta_t}{1 - \delta_t} \frac{\partial \pi_{l,0}^i}{\partial b_{T-1}} \bigg|_{b_{T-1}=b_{T-1}^*} + \frac{\delta_t}{1 - \delta_t} \frac{\partial \pi_{l,0}^i}{\partial b_{T-2}} \bigg|_{b_{T-2}=b_{T-2}^*} + \frac{\delta_t}{1 - \delta_t} \frac{\partial \pi_{l,0}^i}{\partial b_{T-1}} \bigg|_{b_{T-1}=b_{T-1}^*}
\]

Moreover,

\[
-\frac{\partial \pi_h}{\partial b_{T-2}} \bigg|_{b_{T-2}=b_{T-2}^*} = -\left( \frac{\partial \pi_h}{\partial b_{T-2}} \right) \frac{\partial \pi_h}{\partial b_{T-1}} \bigg|_{b_{T-1}=b_{T-1}^*},
\]

By the assumption that \( e''(e) \) is weakly increasing, (15) must hold. Thus, \( b_{T-2}^* \geq b_{T-1}^* \) cannot hold in the optimal contract. Hence, \( b_{T-2}^* < b_{T-1}^* \).

Next, I show that for \( t \in \{1, \ldots, T - 2\} \), \( b_{T-1}^* < b_{T-1}^* \) must hold provided that \( b_T \) is strictly increasing for all \( T \geq \tau \geq t \). The proof of this is very similar to the proof above for the claim that \( b_{T-2}^* < b_{T-1}^* \). First, one needs to verify that \( \pi_{l,T}^i < \pi_{l,T+1}^i + \pi_{h,T} < \pi_{h,T+1} \) for all \( T - 1 \geq \tau \geq t \).[54] But, this is true due to the hypothesis that \( b_T \) is monotone increasing for \( \tau \geq t \) and due to the fact that continuation payoffs are not used to motivate the agent. First, consider the case in which

\[
b_{T-1}^* \leq \delta_t \pi_{l,t}^i - \frac{\delta_t}{1 - \delta_t} \bar{\pi}.
\]

[54] This is true for \( t = T - 1 \), as I already argued. Then, the result follows by induction.
Assume towards a contradiction that \( b^*_t \leq b^*_{t-1} \). From

\[
\begin{align*}
\frac{\partial \pi^i_{l,t}}{\partial b_t} |_{b_t = b^*_t} &+ \frac{\partial \pi^i_{l,0}}{\partial b_t} |_{b_t = b^*_t} \frac{\partial b_t}{\partial b_{t-1}} = 0 \\
\end{align*}
\]

it follows that

\[
\begin{align*}
b^*_t < \delta_t \pi^i_{l,t+1} - \frac{\delta_t}{1 - \delta_t} \bar{\pi},
\end{align*}
\]

But this implies that

\[
\begin{align*}
b^*_t = \delta_h \pi_{h,t+1} - \frac{\delta_h}{1 - \delta_h} \bar{\pi}.
\end{align*}
\]

Otherwise, the high type could increase \( b^*_t \) (and the initial fixed wage) slightly and make a positive gain. But then,

\[
\begin{align*}
b^*_t = \delta_h \pi_{h,t+1} - \frac{\delta_h}{1 - \delta_h} \bar{\pi} \geq b^*_{t-1}
\end{align*}
\]

implies that \( b^*_{t-1} < b^*_t \), a contradiction. Next, consider the case where

\[
\begin{align*}
b^*_t > \delta_l \pi^i_{l,t} - \frac{\delta_l}{1 - \delta_l} \bar{\pi}.
\end{align*}
\]

Assume towards a contradiction that \( b^*_{t-1} \geq b^*_t \). This implies that

\[
\begin{align*}
b^*_t \leq b^*_{t-1} \leq \delta_l \pi^i_{l,t+1} - \frac{\delta_l}{1 - \delta_l} \bar{\pi} < \delta_l \pi^i_{l,t+1} - \frac{\delta_l}{1 - \delta_l} \bar{\pi}.
\end{align*}
\]

The last inequality implies that a small increase in \( b^*_t \) is enforceable for the high type. Also, it implies that

\[
\begin{align*}
b^*_t \geq \delta_l \pi^i_{l,t+1} - \frac{\delta_l}{1 - \delta_l} \bar{\pi},
\end{align*}
\]

otherwise the high type can make a gainful increase in \( b^*_t \). Now, let the contract change as follows: \( b^*_{t-1} \) is decreased and \( b^*_t \) is increased slightly such that

\[
\begin{align*}
\frac{\partial \pi^i_{l,0}}{\partial b_t} |_{b_t = b^*_{t-1}} + \frac{\partial \pi^i_{l,0}}{\partial b_t} |_{b_t = b^*_t} \frac{\partial b_t}{\partial b_{t-1}} = 0 \\
\end{align*}
\]

holds. But this implies that

\[
\begin{align*}
\frac{\partial \pi^i_{l,0}}{\partial b_t} |_{b_t = b^*_{t-1}} + \frac{\partial \pi^i_{l,0}}{\partial b_t} |_{b_t = b^*_t} \frac{\partial b_t}{\partial b_{t-1}} < 0,
\end{align*}
\]

resulting in a contradiction. The proof for showing this follows steps that are very similar to those I used to prove the claim that \( b^*_{T-2} < b^*_{T-1} \). Therefore, it is omitted.

In the (constrained) optimal contract, \( u_t = \frac{a}{1 - \delta} \) for \( t \geq 1 \) as I already discussed. Moreover, \( u_0 \geq \frac{a}{1 - \delta} \). As a result, \( w_t \) is strictly decreasing as long as \( b_t \) is strictly increasing.