Online Appendix to "Honest equilibria in reputation games: The role of time preferences"

Melis Kartal*

May 19, 2017

A 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 not used in the optimal hybrid contract, and hence, default takes place through refusing to pay the bonus.

CLAIM 1: *In the optimal hybrid equilibrium, $u_t = \frac{\bar{u}}{1-\delta}$ 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, $u_0 = \frac{\bar{u}}{1-\delta}$ must hold, otherwise w_0 can be reduced by a small amount and make both types better off. Next, suppose towards a contradiction that $u_{t+1} > \frac{\bar{u}}{1-\delta}$ at some $t \geq 0$. Note that either (i) $b_t > \delta_l \pi_{l,t+1} - \delta_l \pi^d_{l,t+1}$; or (ii) $b_t \leq \delta_l \pi_{l,t+1} - \delta_l \pi^d_{l,t+1}$, 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 $\pi^d_{\theta,t+1}$ represents the punishment payoff of a type-θ principal who defaulted at t. If (i) holds, then $u_{t+1} > \frac{\bar{u}}{1-\delta}$ 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 ε; thus, u_{t+1} reduces by ε whereas $\pi_{h,t+1}$ and $\pi_{l,t+1}$ both increase by ε. 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 ε is small enough), and

*Department of Economics, Vienna University of Economics and Business, Welthandelsplatz 1, 1020, Vienna, Austria (e-mail: melis.kartal@wu.ac.at). 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.

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 ε, and

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

together with $\delta_h > \delta$ implies that

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

Thus, $b_t + \delta\varepsilon$ is enforceable with the high type principle. With ε sufficiently small, $b_t + \delta\varepsilon + \pi^d_{l,t+1} > \delta_l(\pi^i_{l,t+1} + \varepsilon)$ 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 $\delta_h^{t-1}(\delta_h - \delta)\varepsilon > 0$ in the modified separating contract, a contradiction.

Next, consider the case in which (ii) holds, and suppose towards a contradiction that $u_{t+1} > \frac{\bar{u}}{1-\delta}$ 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 ε, which reduces u_{t+1} by ε and increases $\pi_{h,t+1}$ and $\pi_{l,t+1}$ by ε—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

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

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 $\left[\delta_h^t(\delta_h - \delta_l)\right]\varepsilon > 0$, 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, $u_t = \frac{\bar{u}}{1-\delta}$ 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 the total payment P_t is observed and $\mu^1_{t+1} = \mu^2_t$ for every t. Therefore, I focus on $\mu_t \equiv \mu^1_t$ for $t \geq 1$, where μ_t denotes the posterior belief of the sender at the beginning of $t \geq 1$, and μ_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 $\tilde{T} < \infty$ such that the low type strictly prefers defaulting at period $\tilde{T}$. 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 $\{w_t, b_t\} = C_l$ and e_l 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 $\frac{s_l-\bar{u}}{1-\delta_h}$, and hence, a separating equilibrium is strictly better, or (ii) the high type's payoff from the hybrid contract is higher than $\frac{s_l-\bar{u}}{1-\delta_h}$, 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_l. 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 $\tilde{T} < \infty$ such that the low type strictly prefers defaulting at $\tilde{T}$. 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 λ_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 ν_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 $\frac{s_l-\bar{u}}{1-\delta_h}$; 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 $v_0 = 0$. 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 δ_l and δ_h are not far from each other or if μ_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 $v_t < 1$ (and $v_\tau > 0$ 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 $v_t < 1$ and $v_\tau > 0$ 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 $v_{t-1} > 0$ and $v_t < 1$ by hypothesis,

$$\begin{aligned} b_{t-1} &\leq \delta_l \pi_{l,t} - \frac{\delta_l}{1-\delta_l}\bar{\pi} = \delta_l \left(s(e(\lambda_t, b_t)) - \bar{u}) + \lambda_t b_t + \frac{\delta_l}{1-\delta_l}\bar{\pi} \right) - \frac{\delta_l}{1-\delta_l}\bar{\pi} \\ &< \delta_l \frac{s_l - \bar{u} - \bar{\pi}}{1-\delta_l} = b_l. \end{aligned}$$

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 λ 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.[1] The second inequality above follows because $\lambda_t b_t < b_l$, and thus,

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

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

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

[1] 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, $u_1 > \frac{\bar{u}}{1-\delta}$, and e_0, the equilibrium effort at $t = 0$, could be increased by a small amount because $u_1 > \frac{\bar{u}}{1-\delta}$. 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.

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 $k \in \{\tau, \tau+1, ..., t-1\}$. As a result, $b_{\tau-1} < \delta_l \frac{s_l-\bar{u}-\bar{\pi}}{1-\delta_l} = b_l$. Thus, $b_\tau < b_l$ and $s_\tau < s_l$ 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 (i) $b_t < \lambda_{t_k} b_{t_k}$ at every $t < t_k$, and (ii) $b_l \leq \lambda_{t_0} b_{t_0}$.*

PROOF: I start with part (i). If $\upsilon_0 = 0$, then $t_0 = 0$ and there is nothing to prove, so assume that $\upsilon_0 > 0$. Take an arbitrary $t_k > 0$, and suppose towards a contradiction that $b_{t_k-1} \geq \lambda_{t_k} b_{t_k}$. 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 $\upsilon_0 > 0$. Hence, $b_{t_k-1} < \lambda_{t_k} b_{t_k}$ must hold. Next, assume that $b_t < \lambda_{t_k} b_{t_k}$ holds for all $t \in \{\tau, \tau+1, \ldots, t_k-1\}$ by the induction hypothesis. I now show that $b_{\tau-1} < \lambda_{t_k} b_{t_k}$ also holds. Again, by the definition of t_k, $\nu_{\tau-1} > 0$. Thus,

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

I will now show that

$$\sum_{i=\tau}^{t_k} \delta_l^{i-\tau} (s(e(\lambda_i, b_i)) - \bar{u}) + \delta_l^{t_k-\tau} \lambda_{t_k} b_{t_k} + \frac{\delta_l}{1-\delta_l}\bar{\pi} < s(e(\lambda_{t_k}, b_{t_k})) - \bar{u} + \lambda_{t_k} b_{t_k} + \frac{\delta_l}{1-\delta_l}\bar{\pi}. \quad (1)$$

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 $\lambda_t b_t = c(e(\lambda_t, b_t))$ 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

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

in order to prove that (1) holds. Note that $\lambda_{t_k} b_{t_k} > b_l$ 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, (1) holds. From (1), it follows that $b_{\tau-1} < \lambda_{t_k} b_{t_k}$ must hold. Otherwise, the implication is that $\lambda_{t_k} b_{t_k} < b_l$, but this is a contradiction by Claims 3 and 4. Finally, I show that part (ii) holds. Suppose not. Then, $b_l > \lambda_{t_0} b_{t_0}$. 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,

$$b_l \leq \lambda_{t_0} b_{t_0} < \prod_{k=0}^{K} \lambda_{t_k} b_{t_K} \leq b_h \prod_{k=0}^{K} \lambda_{t_k} = \mu_0 b_h$$

since $b_{t_K} \leq b_h$ and $\prod_{k=0}^{K} \lambda_{t_k} = \mu_0$. As a result, $\mu_0 > \frac{b_l}{b_h}$ 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 $\frac{s_l-\bar{u}}{1-\delta_l}$, it can be checked that the high type obtains a payoff strictly lower than $\frac{s_l-\bar{u}}{1-\delta_h}$, 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 μ_0, as I will now show (the payoff of the optimal separating equilibrium does not depend on μ_0). Take an arbitrary μ_0 and an arbitrary sequence $\{\mu_0^n\}$ such that $\lim_{n\to\infty}\mu_0^n = \mu_0$. Let π and π^n denote the payoff of the optimal hybrid contract with μ_0 and μ_0^n, respectively. I will show that π^n converges to π as μ_0^n converges to μ_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 μ_0^n 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 μ_0. Since $\nu_{t_0} < 1$, it follows that $\lambda_{t_0} \in (0,1)$. The hybrid contract that I will construct given μ_0^n is exactly the same as the optimal hybrid contract with μ_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 $\nu_{t_0}^n = \nu_{t_0}\frac{\mu_0^n}{1-\mu_0^n}\frac{1-\mu_0}{\mu_0}$ and $\lambda_{t_0}^n = \mu_0^n + (1-\mu_0^n)\nu_{t_0}^n$ in the hybrid contract with μ_0^n. 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 μ_0 at period t_0. Note that $\nu^n_{t_0} < 1$ with all sufficiently large n. Moreover, as μ^n_0 goes to μ_0, $\nu^n_{t_0}$ goes to ν_{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 μ_0 and μ^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 π as μ^n_0 converges to μ_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 μ_0 given the optimal hybrid contract with μ^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 μ^n_0, and b_{t_0} is the bonus in the hybrid contract with μ^n_0. Similar to the construction above, the posterior at $t_0 + 1$ is identical in the two contracts with μ_0 and μ^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 π^n by only $(\gamma\,\delta_h^{t_0} + (1-\gamma)\delta_l^{t_0})\left(s(e^n_{t_0}) - s(e_{t_0})\right)$, which is strictly smaller than ε if n is large enough because y and s are continuous, and $\lambda^n_{t_0}$ and λ_{t_0} are arbitrarily close by construction with large enough n. Thus, $\lim_{n\to\infty} \pi^n > \pi$ 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 $\mu_0 \leq \frac{b_l}{b_h} + \varepsilon$. For the following claim, consider μ_0 such that $\mu_0 > 1 - \epsilon$ for small $\epsilon > 0$. I construct a hybrid equilibrium such that at $t = 0$ the fixed wage is w_h, and the bonus reward 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 $\epsilon > 0$, the payoff of the high type and low type approximate $\frac{s_h - \bar{u}}{1-\delta_h}$ and $s_h - \bar{u} + b_h + \delta_l \frac{\bar{\pi}}{1-\delta_l}$, respectively. But the separating equilibrium payoff for the high type is bounded above away from $\frac{s_h - \bar{u}}{1-\delta_h}$ due to a fixed cost of signaling which is independent of μ_0, while the low type's payoff is only $\frac{s_l - \bar{u}}{1-\delta_l}$. Thus, if μ_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 μ_γ such that $\mu_\gamma > \frac{b_l}{b_h}$ and the optimal contract is separating if $\mu_0 \leq \mu_\gamma$ and hybrid otherwise. First, I show why a single cutoff exists. This is because, while the prior belief μ_0 does not affect the payoff of the optimal separating contract for fixed γ, the payoff of the optimal hybrid contract strictly increases in μ_0. To see why, let $\mu'_0 > \mu_0$. I will now modify the optimal contract with μ_0 and generate a hybrid contract with μ'_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 μ_0; that is, $\nu_{t_k} < 1$ and $\mu_{t_k} \neq \mu_{t_k+1}$. Since $\mu'_0 > \mu_0$, there exists a t_K such that $\mu_{t_K} < \mu'_0 \leq \mu_{t_K+1}$ (note that $\mu_{t_0} = \mu_0$ by the definition of t_k). Then, the default rate of the low type in the modified contract with μ'_0 is zero at every $t < t_K$, and the implemented effort levels are compatible with this, that is, $\nu'_t = 1$ and $c(e'_t) = b_t$ at every $t < t_K$, where b_t represents the bonus at period t in the original hybrid contract with μ_0. Everything else is the same as in the original contract until t_K. In period t_K, ν'_{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 μ'_0 generates a strictly higher payoff than the optimal hybrid contract with μ_0 for both types since (i) the implemented effort in the contract with μ'_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 μ_γ is increasing in γ to complete the proof of part (i). To see why, suppose towards a contradiction that there exists a $\tilde{\gamma}$ such that $\mu_{\gamma'} < \mu_{\tilde{\gamma}}$ for some $\gamma' > \tilde{\gamma}$. At the prior $\mu_{\tilde{\gamma}}$,

$$\tilde{\gamma} U^h_{sep} + (1 - \tilde{\gamma}) U^l_{sep} = \tilde{\gamma} U^h_{hyb}(\mu_{\tilde{\gamma}}) + (1 - \tilde{\gamma}) U^l_{hyb}(\mu_{\tilde{\gamma}})$$

where U^θ_{sep} (U^θ_{hyb}) represents the optimal separating (hybrid) equilibrium payoff of type θ. Note that U^θ_{hyb} 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 $\mu_{\gamma'}$,

$$\begin{aligned} \gamma' U^h_{sep} + (1 - \gamma') U^l_{sep} &= \gamma' U^h_{hyb}(\mu_{\gamma'}) + (1 - \gamma') U^l_{hyb}(\mu_{\gamma'}) \\ &< \gamma' U^h_{hyb}(\mu_{\tilde{\gamma}}) + (1 - \gamma') U^l_{hyb}(\mu_{\tilde{\gamma}}) \end{aligned}$$

where the strict inequality follows because $\mu_{\gamma'} < \mu_{\tilde{\gamma}}$. This implies that

$$(\gamma' - \tilde{\gamma})(U^h_{sep} - U^l_{sep}) < (\gamma' - \tilde{\gamma})(U^h_{hyb}(\mu_{\tilde{\gamma}}) - U^l_{hyb}(\mu_{\tilde{\gamma}})).$$

and $U^h_{sep} < U^h_{hyb}(\mu_{\tilde{\gamma}})$ since $U^l_{sep} = \frac{s_l - \bar{u}}{1 - \delta_l}$ and $U^l_{hyb}(\mu_{\tilde{\gamma}}) \geq \frac{s_l - \bar{u}}{1 - \delta_l}$ (the latter holds since if $U^l_{hyb}(\mu_{\tilde{\gamma}}) < \frac{s_l - \bar{u}}{1 - \delta_l}$, then the hybrid equilibrium would be strictly dominated by a separating equilibrium due to Claim 3). But if $U^h_{sep} < U^h_{hyb}(\mu_{\tilde{\gamma}})$, then

$$\tilde{\gamma} U^h_{sep} + (1 - \tilde{\gamma}) U^l_{sep} < \tilde{\gamma} U^h_{hyb}(\mu_{\tilde{\gamma}}) + (1 - \tilde{\gamma}) U^l_{hyb}(\mu_{\tilde{\gamma}}),$$

a contradiction. Note that the steps above shows that if $\tilde{\gamma}$ is such that $U^l_{hyb}(\mu_{\tilde{\gamma}}) > \frac{s_l - \bar{u}}{1 - \delta_l}$, then it must be the case that $\mu_{\gamma'} > \mu_{\tilde{\gamma}}$ for all $\gamma' > \tilde{\gamma}$. Hence, μ_γ is strictly increasing at sufficiently low levels of γ. This is because if γ is low enough, then $U^l_{hyb}(\mu_\gamma) > \frac{s_l - \bar{u}}{1 - \delta_l}$ must hold in the optimal hybrid contract. To see why, suppose towards a contradiction that $U^l_{hyb}(\mu_\gamma) = \frac{s_l - \bar{u}}{1 - \delta_l}$ for all $\gamma \in (0, 1)$. By Claim 3, $U^l_{hyb}(\mu_\gamma) = \frac{s_l - \bar{u}}{1 - \delta_l}$ 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 $b_0 \mu_\gamma = b_l$ so that $U^l_{hyb}(\mu_\gamma) = \frac{s_l - \bar{u}}{1 - \delta_l}$ can hold. Moreover, $b_0 < b_h$ since $b_0 \mu_\gamma = b_l$ and $\mu_\gamma > \frac{b_l}{b_h}$. Thus, b_0 can be increased slightly and make the low type better off. If γ is low enough (for example, lower than $\mu_\gamma \frac{y'(e_h)}{c'(e_h)}$), then $U^l_{hyb}(\mu_\gamma) = \frac{s_l - \bar{u}}{1 - \delta_l}$ 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 μ_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 $t < T$.[2] 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 $\pi_{l,T} > \frac{s_T - \bar{u}}{1-\delta_l}$ (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 $t < T - 1$. 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

[2]The 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 $b_\tau < b_{\tau+1}$ and $e_\tau < e_{\tau+1}$ for all $\tau \in \{t, ..., T - 1\}$.

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 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 $t \in \{\tau, \tau+1, ..., T-2\}$ 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 $t \in \{\tau, \tau+1, ..., T\}$ and $\pi_{l,T} > \frac{s_T - \bar{u}}{1-\delta_l}$. 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

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

implies that $\lambda_\tau b_\tau \leq b_l$, 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 $\frac{s_l - \bar{u}}{1-\delta_l}$, 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 $y(e(\tilde{b})) = y(e_l) - c(e_l)$, and consider hybrid equilibria that give the low type an equilibrium payoff greater than $\frac{s_l - \bar{u}}{1-\delta_l}$. If $y'(e_h)/c'(e_h)$ is sufficiently larger than 1 (for example, higher than $\frac{b_h}{\tilde{b}}$), then it is optimal to impose the worst punishment after a default. The same is also true if δ_l and δ_h are not far from each other or if μ_0 is sufficiently high.*

PROOF: Assume that $\pi^d_{\theta,t+1} > \frac{\bar{\pi}}{1-\delta_\theta}$ for some $t \geq 0$ in the optimal hybrid contract, where $\pi^d_{\theta,t+1}$ is (as defined before) the punishment payoff of a type-θ 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,

$$b_T + \delta_h \pi^d_{h,T+1} = \delta_h \pi_{h,T+1},$$

and

$$b_T + \delta_l \pi^d_{l,T+1} \geq \delta_l \pi_{l,T+1},$$

where (as before) $\pi_{\theta,t+1}$ represents the payoff of a type-θ principal who has not defaulted until $t+1$. If it were the case that $b_T + \delta_h \pi^d_{h,T+1} < \delta_h \pi_{h,T+1}$, then the punishment payoff $\pi^d_{\theta,T+1}$ 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 $\delta_l \pi^d_{l,T+1}$ is decreased by the same amount. It follows that $\delta_h \pi^d_{h,T+1}$ falls by more than ε 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, μ_T (more generally, every λ_t) is bounded below away from zero (no matter how small μ_0 might be) in every hybrid equilibrium such that the low type's payoff is at least $\frac{s_l - \bar{u}}{1-\delta_l}$. This is because $e(\mu_T, b_T)$ must be such that $y(e(\mu_T, b_T)) = y(e(\tilde{b})) > y(e_l) - c(e_l) > 0$ holds (otherwise, the low type's payoff is strictly lower than $\frac{s_l - \bar{u}}{1-\delta_l}$). Thus, $\mu_T > \frac{\tilde{b}}{b_h} > 0$. 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 $\frac{b_h}{\tilde{b}}$, 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 $\pi^d_{\theta,t+1} > \frac{\bar{\pi}}{1-\delta_\theta}$. Then, $b_t + \delta_l \pi^d_{l,t+1} = \delta_l \pi_{l,t+1}$ and $b_t + \delta_h \pi^d_{h,t+1} \leq \delta_h \pi_{h,t+1}$. Assume that b_t is increased by a small amount $\varepsilon > 0$ whereas $\delta_l \pi^d_{l,t+1}$ 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 $\delta_h \pi^d_{h,t+1}$ is larger than ε 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 $\frac{b_h}{\tilde{b}}$. 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 μ_0 is sufficiently high. Next, I show that it is optimal if δ_h and δ_l are sufficiently close. To show this I will argue that if δ_h and δ_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 $\lambda_t \frac{y'(e_t)}{c'(e_t)} \leq 1$; that is, $\lambda_t \leq \frac{c'(e_t)}{y'(e_t)} < 1$. Thus, λ_t is bounded above away from 1 because $\frac{c'(e_t)}{y'(e_t)} < \frac{c'(e_h)}{y'(e_h)} < 1$. 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 δ_h and δ_l are sufficiently close, then

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

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 λ_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

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

However, note that as δ_h and δ_l get closer and closer, it must be the case that

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

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

B Proofs of Claims in Section III

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_θ represent the production function of type-θ 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

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

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

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

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

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

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

$$\frac{y_h(e_l) - y_l(e_l)}{1-\delta} < \frac{y_h(e_h) - y_l(e_h)}{1-\delta} \leq \frac{y_h(e_l) - y_l(e_l)}{1-\delta} + \kappa.$$

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 $T(\kappa) = \min\{t \geq 0 | C_t = C_h \text{ given } \kappa > 0\}$ 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

$$\sum_{t=0}^{T(\kappa)-1} \delta^t \left[(y_h(e_t) - y_l(e_t)) - (y_h(e_l) - y_l(e_l))\right] \geq \delta^{T(\kappa)} \left(\frac{y_h(e_l) - y_l(e_l)}{1-\delta} - (y_h(e_h) - y_l(e_h)) \right)$$

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 $t \geq 0$, $\limsup_{\kappa\to 0} e_t^\kappa \leq e_l$ holds, which increases the convergence rate of the left-hand side ($\limsup_{\kappa\to 0} e_t^\kappa \leq e_l$ 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).

C 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 $T = \min\{t \geq 0 | \nu_t = 0\}$. 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., $\{C_t\}_{t=T+1}^\infty = \{C_h, C_h, ...\}$.[3] 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) $D_{T-1} = \{w'_{T-1}, b_{T-1}\}$, 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}),

$$w'_{T-1} = w_{T-1} + \Delta \frac{\delta_l + \delta_h}{2},$$

and the term Δ 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).

[3] 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.

(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 $\{C_t\}_{t=T+1}^{\infty} = \{C_h, C_h, ...\}$, 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.[4].

(iii) $\{D_t\}_{t\geq T+1} = \{C_h, C_h, ...\}$. 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 $T-1$,

$$\pi_{l,T-1} = y_{T-1} - w_{T-1} - b_{T-1} + \max\{b_{T-1} + \delta_l \pi^d_{l,T}, \delta_l \pi_{l,T}\}$$

given $\{C_t\}_{t=0}^{\infty}$, where $\pi^d_{l,T}$ 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 maximum possible payoff of

$$y_{T-1} - w'_{T-1} - b_{T-1} + \max\{b_{T-1} + \delta_l \pi^d_{l,T}, \delta_l(\pi^i_{l,T} + \Delta)\},$$

which is strictly lower than $\pi_{l,T-1}$ (y_{T-1}, b_{T-1} and e_{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.[5] 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 $\mu_0 > \frac{b_l}{b_h}$ as $b_0 < b_h$. Thus, such an

[4]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.

[5]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 $\{D_t\}_{t\geq 1} = \{C_h, C_h, ...\}$ 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.

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* λ_t, 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, λ_t. I explore the inference on this probability given a deviation, which I denote by $\hat{\lambda}_t$.

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 $c(e'_t) \leq b'_t$ and $w'_t + b'_t - c(e'_t) \geq \bar{u}$ hold at every $t \geq k$.[6] Otherwise, I assume that the agent rejects D_t. 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^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+1}^k: 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_t^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},..., D_{t-1}? If the answer is again no, then the agent accepts the offer, infers that $\hat{\lambda}_t = 1$ 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_t^k results in inferring $\hat{\lambda}_t = 1$ for every $t \in \{k, k+1, ...\}$ (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 $t \in \{k, k+1, ...\}$;
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.

[6]Thus, it can easily be checked whether the agent's participation and incentive compatibility constraints are satisfied in the deviation contract.

D Proofs of Claims in Section VI.A

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 $c_H^b - c_H^g \leq c_M^b - c_M^g$, 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:

$$p_H + \frac{\delta}{1-\delta}\bar{\pi} \leq \frac{\Phi p_H + (1-\Phi)p_M - c_H^g}{1-\delta},$$

and

$$p_H - c_H + \frac{\delta}{1-\delta}\bar{\pi} \leq p_M - c_H + \delta\frac{\Phi p_H + (1-\Phi)p_M - c_H^g}{1-\delta},$$

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.[7] 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 $\bar{\pi}$ 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 $\bar{\pi} = u_L$. 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 $\bar{\pi} = u_L = 0$ from now on. Assuming that $\bar{\pi} = u_L = 0$ 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

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

[7] 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).

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,

$$p_H > \frac{\Phi p_H + (1-\Phi) p_M - c_H^b}{1-\delta}. \tag{2}$$

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.[8]

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 $\frac{\Phi(p_H - p_M) + (1-\Phi) p_M - (c_H^g - c_M^g)}{1-\delta}$, 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 (2). 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

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

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

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

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

[8] It is straightforward to verify that there exist parameter values c_M^b, c_H^b, p_H, p_M and Ψ such that all of the conditions are satisfied.

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 $q = q_H$.[9] 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

$$\sum_{t=0}^{T-1} \delta^t \left(\mathbb{E}(p_t|e_t) - c^b_{e_t} - a_t \right) + p_H - a_T.$$

But $p_H - a_T$ is strictly greater than

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

because $\mathbb{E}(p_t|e_t)$ is, by definition, the truthful expected equilibrium price and thus, $\sum_{t=T}^{\infty} \delta^{t-T} \times \left(\mathbb{E}(p_t|e_t) - c^b_{e_t} - a_t \right)$ is weakly smaller than $\frac{\Phi p_H + (1-\Phi) p_M - c^b_H}{1-\delta} - \sum_{t=T}^{\infty} a_t$, which in turn is strictly smaller than $p_H - a_T$ due to (2). Hence, separation is not possible. Next, I show that $c^b_H - c^g_H > c^b_M - c^g_M$ 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 $\frac{\Phi p_H + (1-\Phi) p_M - c^g_H}{1-\delta}$. 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 $c^b_H - c^g_H > c^b_M - c^g_M$ is not sufficient. A similar analysis shows that separation is not generally possible in the case where the two types differ in Φ 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 δ_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

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

[9] 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.

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

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

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.[10] 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 $\max\left\{p_H, \frac{p_H - c_H}{1 - \Phi\phi\delta_l}\right\} -$

[10] In particular, there exists a separating equilibrium such that good names are bought only by good sellers because they are too expensive for bad sellers.

$\frac{\Phi p_M - c_M}{1-\phi\delta_l}$, whereas the benefit of separation for a type-g seller is $\frac{\Phi(p_H - p_M)+(1-\Phi)p_M-(c_H-c_M)}{1-\phi\delta_h}$. 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 $p_s \geq \max\left\{p_H, \frac{p_H - c_H}{1-\Phi\phi\delta_l}\right\}$ so that a type-b firm is deterred from buying a good name. Thus, as long as δ_h is sufficiently high, a separating equilibrium exists because there exists a p_s that satisfies both constraints.[11] This in turn implies that a good name will not go bad because it will be too expensive for type-b sellers.

E 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

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

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

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

[11] The cost of advertising for new type-g firms that are unable to buy a name is equal to p_s.

As a result, at a sufficiently large t, the low type will have defaulted with a very high probability (one high output realization suffices), 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 given the initial hypothesis that b_t^* converges to b_h. 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,

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

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

$$\lim_{t\to\infty} \frac{(V_l - \pi_{l,t+1}^i)}{\frac{b_h - b_t^*}{\delta_h}} \leq \lim_{b_h \to b_t^*} \frac{\delta_h\left[\left(s_h + e_h\left(b_h^* + \frac{\delta_l}{1-\delta_l}\bar{\pi}\right)\right) - \left(s_t + e_t\left(b_t^* + \frac{\delta_l}{1-\delta_l}\bar{\pi}\right)\right)\right]}{(1-\delta_l)\left(b_h - b_t^*\right)} \in (0, \infty)$$

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 $\delta_l < \delta_h$ 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.[12] 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 $u_t = \frac{\bar{u}}{1-\delta}$ 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 $\pi_{l,T-1}^i < \pi_{l,T}^i = V_l$.

[12] The agent's incentives are not affected since Lemma 4 holds in this setting.

There are two cases to consider:

(i) First, assume that

$$b^*_{T-2} \leq \delta_l \pi^i_{l,T-1} - \frac{\delta_l}{1-\delta_l}\bar{\pi}.$$

Since $\pi^i_{l,T-1} < \pi^i_{l,T}$ it follows that

$$b^*_{T-1} \leq b^*_{T-2} < \delta_l \pi^i_{l,T} - \frac{\delta_l}{1-\delta_l}\bar{\pi}.$$

However,

$$b^*_{T-1} < \delta_l \pi^i_{l,T} - \frac{\delta_l}{1-\delta_l}\bar{\pi}$$

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

$$b^*_{T-2} > \delta_l \pi^i_{l,T-1} - \frac{\delta_l}{1-\delta_l}\bar{\pi}.$$

Assume towards a contradiction that $b^*_{T-2} \geq b^*_{T-1}$. Consider the modified contract: b^*_{T-2} is decreased slightly whereas b^*_{T-1} is increased in a way that

$$\frac{\partial \pi^i_{l,0}}{\partial b_{T-2}}|_{b_{T-2}=b^*_{T-2}} + \frac{\partial \pi^i_{l,0}}{\partial b_{T-1}}|_{b_{T-1}=b^*_{T-1}} \frac{\partial b_{T-1}}{\partial b_{T-2}} = 0 \tag{3}$$

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

$$\frac{\partial \pi_h}{\partial b_{T-2}}|_{b_{T-2}=b^*_{T-2}} + \frac{\partial \pi_h}{\partial b_{T-1}}|_{b_{T-1}=b^*_{T-1}} \frac{\partial b_{T-1}}{\partial b_{T-2}} < 0. \tag{4}$$

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 (3) and the fact that

$$b^*_{T-1} \geq \delta_l \pi^i_{l,T} - \frac{\delta_l}{1-\delta_l}\bar{\pi},$$

it follows that

$$\begin{aligned}
\frac{\partial b_{T-1}}{\partial b_{T-2}} &= -\frac{\frac{\partial \pi^i_{l,0}}{\partial b_{T-2}}|_{b_{T-2}=b^*_{T-2}}}{\frac{\partial \pi^i_{l,0}}{\partial b_{T-1}}|_{b_{T-1}=b^*_{T-1}}} = -\frac{\frac{\partial s_{T-2}}{\partial b_{T-2}} + e^*_{T-2} + \left(b_{T-2} + \frac{\delta_l}{1-\delta_l}\bar{\pi}\right)\frac{\partial e}{\partial b_{T-2}} - \delta_l \pi_{l,T-1}\frac{\partial e}{\partial b_{T-2}}}{\delta_l(1-e^*_{T-2})\left[\frac{\partial s_{T-1}}{\partial b_{T-1}} + e^*_{T-1} + \left(b_{T-1} + \frac{\delta_l}{1-\delta_l}\bar{\pi}\right)\frac{\partial e}{\partial b_{T-1}} - \delta_l \pi_{l,T}\frac{\partial e}{\partial b_{T-1}}\right]} \\
&= -\frac{c''(e^*_{T-1})\left[(H-L-c'(e^*_{T-2})) + e^*_{T-2}c''(e^*_{T-2}) + b_{T-2} + \frac{\delta_l}{1-\delta_l}\bar{\pi} - \delta_l \pi_{l,T-1}\right]}{\delta_l c''(e^*_{T-2})(1-e^*_{T-2})\left[(H-L-c'(e^*_{T-1})) + e^*_{T-1}c''(e^*_{T-1}) + b_{T-1} + \frac{\delta_l}{1-\delta_l}\bar{\pi} - \delta_l \pi_{l,T}\right]}
\end{aligned}$$

Moreover,

$$-\frac{\frac{\partial \pi_h}{\partial b_{T-2}}|_{b_{T-2}=b^*_{T-2}}}{\frac{\partial \pi_h}{\partial b_{T-1}}|_{b_{T-1}=b^*_{T-1}}} = -\frac{(H-L-c'(e^*_{T-2}))c''(e^*_{T-1})}{\delta_h(H-L-c'(e^*_{T-1}))c''(e^*_{T-2})}.$$

By the assumption that $ec''(e)$ is weakly increasing, (4) 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 $t \in \{1, \ldots, T-2\}$, $b^*_{t-1} < b^*_t$ must hold provided that b_τ 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 $\pi^i_{l,\tau} < \pi^i_{l,\tau+1}$ and $\pi_{h,\tau} < \pi_{h,\tau+1}$ for all $T-1 \geq \tau \geq t$.[13] But, this is true due to the hypothesis that b_τ 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

$$b^*_{t-1} \leq \delta_l \pi^i_{l,t} - \frac{\delta_l}{1-\delta_l}\bar{\pi}.$$

Assume towards a contradiction that $b^*_{t-1} \geq b^*_t$. From

$$b^*_t \leq b^*_{t-1} \leq \delta_l \pi^i_{l,t} - \frac{\delta_l}{1-\delta_l}\bar{\pi} < \delta_l \pi^i_{l,t+1} - \frac{\delta_l}{1-\delta_l}\bar{\pi},$$

it follows that

$$b^*_t < \delta_l \pi^i_{l,t+1} - \frac{\delta_l}{1-\delta_l}\bar{\pi}.$$

But this implies that

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

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

$$b^*_t = \delta_h \pi_{h,t+1} - \frac{\delta_h}{1-\delta_h}\bar{\pi} > \delta_h \pi_{h,t} - \frac{\delta_h}{1-\delta_h}\bar{\pi} \geq b^*_{t-1}$$

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

$$b^*_{t-1} > \delta_l \pi^i_{l,t} - \frac{\delta_l}{1-\delta_l}\bar{\pi}.$$

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

$$b^*_t \leq b^*_{t-1} \leq \delta_l \pi_{h,t} - \frac{\delta_h}{1-\delta_h}\bar{\pi} < \delta_h \pi_{h,t+1} - \frac{\delta_h}{1-\delta_h}\bar{\pi}.$$

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

$$b^*_t \geq \delta_l \pi^i_{l,t+1} - \frac{\delta_l}{1-\delta_l}\bar{\pi},$$

[13] This is true for $t = T-1$, as I already argued. Then, the result follows by induction.

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

$$\frac{\partial \pi_{l,0}^i}{\partial b_{t-1}}|_{b_{t-1}=b_{t-1}^*} + \frac{\partial \pi_{l,0}^i}{\partial b_t}|_{b_t=b_{t-1}^*} \frac{\partial b_t}{\partial b_{t-1}} = 0$$

holds. But this implies that

$$\frac{\partial \pi_h}{\partial b_{t-1}}|_{b_{t-1}=b_{t-1}^*} + \frac{\partial \pi_h}{\partial b_t}|_{b_t=b_t^*} \frac{\partial b_t}{\partial b_{t-1}} < 0,$$

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, $u_t = \frac{\bar{u}}{1-\delta}$ for $t \geq 1$ as I already discussed. Moreover, $u_0 \geq \frac{\bar{u}}{1-\delta}$. As a result, w_t is strictly decreasing as long as b_t is strictly increasing.