Supplemental Appendix:

Intermediated Asymmetric Information, Compensation and Career Prospects

Ron Kaniel

University of Rochester, FISF Fudan and CEPR

Dmitry Orlov

University of Wisconsin-Madison

9th June, 2025

Contents

Auxiliary Results: Value Function and Stopping Set Monotonicity

Consider a stopping time τ that is measurable with respect to the filtration $\mathcal{F}_t = \sigma(\eta \le t, \mathcal{F})$, where $\mathcal{F}$ is a sufficiently rich independent filtration. In this context, consider the continuation value from optimally stopping a deterministic, measurable, and bounded cumulative cash flow process Y at a time weakly less than η:

$$v_t(p) \overset{def}{=} \sup_{\tau \ge t} \mathrm{E}\left[\int_t^{\tau\wedge\eta} e^{-r(s-t)}\, dY_s \;\middle|\; \eta > t, \tilde{p}_0 = p\right] = \sup_{\tau \ge t} \mathrm{E}\left[\int_t^{\tau} e^{-r(s-t)} \left(p + (1-p)e^{-\lambda(s-t)}\right) dY_s\right] \quad \text{(A.1)}$$

Lemma A4.1 in Davis (2018) shows that any stopping time τ measurable with respect to η takes the form $\tau = T \wedge \eta$ for some deterministic $T \ge 0$, and, given that the right-hand side of (A.1) is a deterministic problem, it follows that the supremum in (A.1) is well-defined. Define $\bar{t}(\tilde{p}_0)$ as the supremum at all times when it is optimal to stop given the prior belief $\tilde{p}_0$ and the cumulative dividend process Y:

$$\bar{t}(p) \overset{def}{=} \sup\left\{t:\ \exists\,(t_n)_{n\in\mathbb{N}}: \quad \lim_{n\to\infty} t_n = t \qquad \textit{and} \qquad v_0(p) = \lim_{n\to\infty} \mathrm{E}_{\tilde{p}_0=p}\left[\int_0^{t_n\wedge\eta} e^{-rs}\, dY_s\right]\right\}. \quad \text{(A.2)}$$

Lemma A.1 (Weak monotonicity). *Function $v_t(p)$ increases weakly in p and, consequently, $\bar{t}(p)$ is weakly increasing in $\tilde{p}_0$.*

Proof. It is without loss to consider $t = 0$. Consider $p' < p''$ and consider a random time ξ, independent of $\tilde{p}_0$ and η, distributed according to

$$\mathrm{P}(\xi \ge s) = \frac{\mathrm{P}_{p'}(\eta \ge s)}{\mathrm{P}_{p''}(\eta \ge s)} = \frac{p' + (1-p')\cdot e^{-\lambda s}}{p'' + (1-p'')\cdot e^{-\lambda s}}.$$

By construction, $\mathrm{P}(\xi \ge s)$ is strictly decreasing in s for $p'' > p'$. Then

$$\begin{aligned}
v_0(p'') &= \sup_\tau \mathrm{E}_{p''}\left[\int_0^{\tau\wedge\eta} e^{-rs} dY_s\right] = \sup_\tau \mathrm{E}_{p''}\left[\int_0^{\tau} \mathbb{1}\{t \le \eta\}\, e^{-rs} dY_s\right] \\
&= \sup_\tau \mathrm{E}\left[\int_0^{\tau} \left(p'' + (1-p'')\,e^{-\lambda s}\right) e^{-rs} dY_s\right] \ge \sup_\tau \mathrm{E}\left[\int_0^{\tau} \mathbb{1}\{s \le \xi\} \left(p'' + (1-p'')\,e^{-\lambda s}\right) e^{-rs} dY_s\right] \\
&= \sup_\tau \mathrm{E}\left[\int_0^{\tau} \mathrm{P}(s \le \xi) \left(p'' + (1-p'')\,e^{-\lambda s}\right) e^{-rs} dY_s\right] = \sup_\tau \mathrm{E}\left[\int_0^{\tau} \left(p' + (1-p')\,e^{-\lambda s}\right) e^{-rs} dY_s\right] \\
&= \sup_\tau \mathrm{E}\left[\int_0^{\tau} \mathrm{P}_{p'}(s \le \eta)\, e^{-rs} dY_s\right] = \sup_\tau \mathrm{E}_{p'}\left[\int_0^{\tau\wedge\eta} e^{-rs} dY_s\right] = \sup_\tau \mathrm{E}_{p'}\left[\int_0^{\tau} e^{-rs} dY_s\right] = v_0(p').
\end{aligned}$$

We can repeat this argument for any $t \ge 0$ and, consequently, $v_t(p)$ is weakly increasing in p. By the standard continuation value argument, if it is weakly optimal to wait until time t for type p', then it is weakly more profitable to wait until time t for type $p'' > p'$. Hence, $\bar{t}(p)$ is weakly increasing in p. □

Lemma A.2 (Strong monotonicity). *Suppose $\exists\, 0 \le t' < t'' \le \bar{t}(p)$ s.t. Y_t is weakly increasing over (t', t'')*

and $Y_{t''-} > Y_{t'+}$. Then $v_t(p) < v_t(\hat{p})$ for all $\hat{p} > p$ and $t \le t'$.

Proof. It is without loss to consider $t = 0$. Since $Y_{t''-} > Y_{t'+}$ there exists $\varepsilon > 0$ s.t. $Y_{t''-\varepsilon} > Y_{t'+\varepsilon}$. Given $\hat{p} > p$, consider a random variable ξ, independent of $\tilde{p}_0$ and η, distributed according to

$$\mathrm{P}(\xi \ge t) = \begin{cases} \dfrac{\mathrm{P}_{p'}(\eta \ge t)}{\mathrm{P}_{p''}(\eta \ge t)} = \dfrac{p' + (1-p') \cdot e^{-\lambda t}}{p'' + (1-p'') \cdot e^{-\lambda t}} & \text{if} \quad t \in [0, t'+\varepsilon) \cup (t''-\varepsilon, +\infty), \\ \dfrac{\mathrm{P}_{p'}(\eta \ge t'+\varepsilon)}{\mathrm{P}_{p''}(\eta \ge t'+\varepsilon)} = \dfrac{p' + (1-p') \cdot e^{-\lambda(t'+\varepsilon)}}{p'' + (1-p'') \cdot e^{-\lambda(t'+\varepsilon)}} & \text{if} \quad t \in (t'+\varepsilon, t''-\varepsilon). \end{cases} \tag{A.3}$$

The probability $\mathrm{P}(\xi \ge t)$ defined via (A.3) is constant for $t \in (t'+\varepsilon, t''-\varepsilon)$ and strictly decreasing for $t \in (t'+\varepsilon, t''-\varepsilon)$. It is continuous at $t = t'+\varepsilon$, and has a downward jump at $t = t''-\varepsilon$.
Consider a sequence of times $(t_n)_{n\in\mathbb{N}}$ s.t. $t_2 - \varepsilon \le t_n \to \bar{t}(p')$ and $\lim_{n\to\infty} \mathrm{E}_{\tilde{p}_0=p}\Big[\int_0^{t_n\wedge\eta} e^{-rs}\, dY_s\Big] = v_0(p)$. Then

$$\begin{aligned} &v(p'') - v(p') = \sup_\tau \mathrm{E}_{p''}\Big[\int_0^{\tau\wedge\eta} e^{-rs} dY_s\Big] - \sup_\tau \mathrm{E}_{p'}\Big[\int_0^{\tau\wedge\eta} e^{-rs} dY_s\Big] \\ \ge& \sup_\tau \mathrm{E}_{p''}\Big[\int_0^{\tau\wedge\xi\wedge\eta} e^{-rs} dY_s\Big] - \sup_\tau \mathrm{E}_{p'}\Big[\int_0^{\tau\wedge\eta} e^{-rs} dY_s\Big] \\ \ge& \lim_{n\to\infty} \mathrm{E}\left[\int_0^{t_n} e^{-rs} \cdot \Big[\mathrm{P}(\xi \ge s) \cdot \Big(p'' + (1-p'') \cdot e^{-\lambda s}\Big) - \Big(p' + (1-p') \cdot e^{-\lambda s}\Big)\Big]\, dY_s\right] \\ =&\mathrm{E}\left[\int_{t'+\varepsilon}^{t''-\varepsilon} e^{-rs} \cdot \left[\frac{p' + (1-p') \cdot e^{-\lambda(t'+\varepsilon)}}{p'' + (1-p'') \cdot e^{-\lambda(t'+\varepsilon)}} \cdot \Big(p'' + (1-p'') \cdot e^{-\lambda s}\Big) - \Big(p' + (1-p') \cdot e^{-\lambda s}\Big)\right] dY_s\right] \\ =&\mathrm{E}\left[\int_{t'+\varepsilon}^{t''-\varepsilon} e^{-rs} \cdot \left[\frac{(p''-p') \cdot \Big(e^{-\lambda(t'+\varepsilon)} - e^{-\lambda s}\Big)}{p'' + (1-p'') \cdot e^{-\lambda(t'+\varepsilon)}}\right] dY_s\right] > 0, \end{aligned} \tag{A.4}$$

where the strict inequality (A.4) follows from $Y_{t''-\varepsilon} - Y_{t'+\varepsilon} > 0$. □

Lemma A.3 (Generalized strong monotonicity)**.** *Suppose process Y is differentiable and given by $dY_t = y_t\, dt$, where y_t is deterministic. Suppose there exist p and $t' < \bar{t}(p)$ s.t. $v_t(p) > 0$. Then $v_t(\hat{p}) > 0$ for all $t \in [0, t']$ and $\hat{p} > p$.*

Proof. Since Y_t is differentiable, it follows from Ito's lemma that $v_t(p)$ is continuously differentiable in t for every p. It follows from Lemma A.1 that $v_t(\hat{p}) \ge v_t(p)$. Suppose $v_{t'}(\hat{p}) > v_{t'}(p) > 0$. By continuity there exists $\varepsilon > 0$ s.t. $v_t(\hat{p}) > v_t(p)$ for every $t \in (t'-\varepsilon, t')$. Suppose, alternatively, $v_{t'}(\hat{p}) = v_{t'}(p) > 0$. It follows from the standard martingale argument that for every p and $t < \bar{t}(p)$

$$r \cdot v_t(p) = y_t - \lambda(1 - \pi_t(p)) \cdot v_t(p) + \dot{v}_t(p). \tag{A.5}$$

Then

$$\begin{aligned}\dot{v}_t(\hat{p}) - \dot{v}_t(p) &= r(v_t(\hat{p}) - v_t(p)) + \lambda(1-\pi_t(\hat{p}))\cdot v_t(\hat{p}) - \lambda(1-\pi_t(p))\cdot v_t(p)\\ &= [r+\lambda(1-\pi_t(\hat{p}))]\cdot(v_t(\hat{p}) - v_t(p)) - \lambda\left(\pi_t(p) - \pi_t(\hat{p})\right)\cdot v_t(p).\end{aligned}$$

By continuity, $\exists\,\epsilon, h > 0$ s.t. $v_t(p) > h$ for every $t \in (t'-\epsilon, t')$. Since $v_{t'}(\hat{p}) = v_{t'}(p)$, by continuity of $v_t(p)$ and $\hat{v}_t(p)$ there exists $\varepsilon \in (0,\epsilon)$ s.t.

$$[r+\lambda(1-\pi_t(\hat{p}))]\cdot(v_t(\hat{p}) - v_t(p)) - \lambda\left(\pi_t(p) - \pi_t(\hat{p})\right)\cdot h < 0.$$

Consequently, $\dot{v}_t(\hat{p}) < \dot{v}_t(p)$ for every $t \in (t'-\varepsilon, t')$ and, hence, for every $t \in (t'-\varepsilon, t')$

$$v_t(\hat{p}) = v_{t'}(\hat{p}) - \int_t^{t'} \dot{v}_s(\hat{p})\,ds = v_{t'}(p) - \int_t^{t'} \dot{v}_s(\hat{p})\,ds < v_{t'}(p) - \int_t^{t'} \dot{v}_s(p)\,ds = v_t(p).$$

The above argument shows that there exists $\varepsilon > 0$ s.t. $v_t(\hat{p}) > v_t(p) > 0$ for every $t \in (t'-\varepsilon, t')$. In what follows, we show that this implies that $v_t(\hat{p}) > 0$ for every $t \leq t'$. Consider, without loss, $t = 0$ and $v_0(\hat{p})$. Define a random variable ξ, independent of $\tilde{p}_0$ and η, distributed according to

$$\mathrm{P}(\xi \geq s) = \begin{cases} \dfrac{\mathrm{P}_p(\eta \geq s)}{\mathrm{P}_{\hat{p}}(\eta \geq s)} = \dfrac{p + (1-p)\cdot e^{-\lambda s}}{\hat{p} + (1-\hat{p})\cdot e^{-\lambda s}} & \textit{if} \quad s \in [0, t'-\varepsilon],\\ \dfrac{\mathrm{P}_p(\eta \geq t'-\varepsilon)}{\mathrm{P}_{\hat{p}}(\eta \geq t'-\varepsilon)} = \dfrac{p + (1-p)\cdot e^{-\lambda (t'-\varepsilon)}}{\hat{p} + (1-\hat{p})\cdot e^{-\lambda (t'-\varepsilon)}} & \textit{if} \quad s \in (t'-\varepsilon, t']. \end{cases} \tag{A.6}$$

Then

$$\begin{aligned} v_0(\hat{p}) &= \sup_{\tau\geq 0} \mathrm{E}_{\hat{p}}\left[\int_t^{\tau\wedge\eta} e^{-rs}\cdot y_s\,ds\right] \geq \sup_{\tau\geq t'-\varepsilon} \mathrm{E}_{\hat{p}}\left[\int_t^{\tau\wedge\eta} e^{-rs}\cdot y_s\,ds\right] \geq \sup_{\tau\geq t'-\varepsilon} \mathrm{E}_{\hat{p}}\left[\int_t^{\tau\wedge\xi\wedge\eta} e^{-rs}\cdot y_s\,ds\right] \qquad (\text{A.7})\\ &= \mathrm{E}_{\hat{p}}\left[\int_0^{(t'-\varepsilon)\wedge\xi\wedge\eta} e^{-rs}\cdot y_s\,ds\right] + \sup_{\tau\geq t'-\varepsilon} \mathrm{E}_{\hat{p}}\left[\int_{t'-\varepsilon}^{\tau\wedge\xi\wedge\eta} e^{-rs}\cdot y_s\,ds\right]\\ &= \int_0^{t'-\varepsilon} e^{-rs}\cdot \mathrm{P}_{\tilde{p}_0=\hat{p}}(\xi\geq s, \eta\geq s)\cdot y_s\,ds + \sup_{\tau\geq t'-\varepsilon} \mathrm{E}_{\hat{p}}\left[\int_{t'-\varepsilon}^{\tau} e^{-rs}\cdot \mathrm{P}_{\tilde{p}_0=\hat{p}}(\xi\geq s, \eta\geq s)\cdot y_s\,ds\right]\\ &= \int_0^{t'-\varepsilon} e^{-rs}\cdot \mathrm{P}_{\tilde{p}_0=\hat{p}}(\xi\geq s)\cdot \mathrm{P}_{\tilde{p}_0=\hat{p}}(\eta\geq s)\cdot y_s\,ds\\ &+ \sup_{\tau\geq t'-\varepsilon} \mathrm{E}_{\hat{p}}\left[\int_{t'-\varepsilon}^{\tau} e^{-rs}\cdot \mathrm{P}_{\tilde{p}_0=\hat{p}}(\xi\geq s)\cdot \mathrm{P}_{\tilde{p}_0=\hat{p}}(\eta\geq s)\cdot y_s\,ds\right]\\ &= \int_0^{t'-\varepsilon} e^{-rs}\cdot \mathrm{P}_p(\eta\geq s)\cdot y_s\,ds + \sup_{\tau\geq t'-\varepsilon} \mathrm{E}_{\hat{p}}\left[\int_{t'-\varepsilon}^{\tau} e^{-rs}\cdot \mathrm{P}_{\tilde{p}_0=\hat{p}}\big(\xi\geq t'-\varepsilon\big)\cdot \mathrm{P}_{\tilde{p}_0=\hat{p}}(\eta\geq s)\cdot y_s\,ds\right]\\ &= \mathrm{E}_{\tilde{p}_0=p}\left[\int_0^{t'-\varepsilon} e^{-rs}\cdot y_s\,ds\right] \end{aligned}$$

$$+ \mathrm{P}_{\tilde{p}_0=\hat{p}}\big(\xi \geq t' - \varepsilon\big) \cdot \mathrm{P}_{\tilde{p}_0=\hat{p}}\big(\eta \geq t' - \varepsilon\big) \cdot \sup_{\tau \geq t'-\varepsilon} \mathrm{E}_{\hat{p}}\bigg[\int_{t'-\varepsilon}^{\tau} e^{-rs} \cdot \frac{\mathrm{P}_{\tilde{p}_0=\hat{p}}(\eta \geq s)}{\mathrm{P}_{\tilde{p}_0=\hat{p}}(\eta \geq t'-\varepsilon)} \cdot y_s \, ds\bigg]$$

$$= \mathrm{E}_{\tilde{p}_0=p}\left[\int_0^{t'-\varepsilon} e^{-rs} \cdot y_s \, ds\right] + \mathrm{P}_{\tilde{p}_0=p}\big(\eta \geq t' - \varepsilon\big) \cdot e^{-r(t'-\varepsilon)} \cdot v_{t'-\varepsilon}(\hat{p})$$

$$> \mathrm{E}_{\tilde{p}_0=p}\left[\int_0^{t'-\varepsilon} e^{-rs} \cdot y_s \, ds\right] + \mathrm{P}_{\tilde{p}_0=p}\big(\eta \geq t' - \varepsilon\big) \cdot e^{-r(t'-\varepsilon)} \cdot v_{t'-\varepsilon}(p) \tag{A.8}$$

$$= \sup_{\tau > t'-\varepsilon} \mathrm{E}_{\tilde{p}_0=p}\bigg[\int_0^{\tau\wedge\eta} e^{-rs} \cdot y_s \, ds\bigg] = v_0(p) \geq 0. \tag{A.9}$$

The first weak inequality in (A.7) follows from the imposition of a constraint that τ be greater than $t' - \varepsilon$. The second weak inequality in (A.7) introduces probabilistic termination that takes place if $s \geq \xi$, where ξ is distributed according to (A.6). Strict inequality (A.8) follows from $v_{t'-\varepsilon}(\hat{p}) > v_{t'-\varepsilon}(p)$ derived earlier. The second equality in (A.9) follows from the fact that $t' - \varepsilon < t' < \bar{t}(p)$, i.e., imposing a restriction on $\tau > t' - \varepsilon$ does not reduce expected value. Weak inequality (A.9) follows the definition of $v_0(p)$ in (A.1) – $\tau = 0$ is feasible and generates value 0. Derivations (A.7) – (A.9) imply that $v_0(\hat{p}) > v_0(p)$. We can repeat this argument for any $t \in [0, t' - \varepsilon)$ and ε. This implies $v_t(\hat{p}) > v_t(p) \geq 0$ for any $t \in [0, t')$ implying that it is strictly suboptimal for type $\hat{p}$ to stop prior to t'. □

Lemma A.4. *Suppose there exists $t' < \bar{t}(p)$ s.t. $y_t \neq 0$ on a set of positive Lebesgue measure in $[t', \bar{t}(p)]$. Then there exists $\hat{t} \in [t', \bar{t}(p))$ s.t. $v_{\hat{t}}(p) > 0$.*

Proof. Since $t < \bar{t}(p)$ it implies that $v_t(p) \geq 0$. Given p denote $\hat{t} \overset{def}{=} \sup\{t \in [t', \bar{t}(p)) : \quad v_t(p) > 0\}$. If $\hat{t} = t'$, it implies that $v_t(p) = 0$ for all $t \in \big[\hat{t}, \bar{t}(p)\big]$. Differentiating $v_t(p)$ with respect to t obtain that $y_t = 0$ almost everywhere on $\big[\hat{t}, \bar{t}(p)\big]$ which contradicts the assumption of the Lemma. □

Proof of Proposition 2

The only distinction between Propositions 1 and 2 is that the intermediary's optimal retention rule solves a real option problem. We first characterize the intermediary's churning decision assuming clients keep their beliefs constant and then show how the implied stopping sets $\mathbb{T}(\tilde{p}_0, l)$ can then be used to construct the equilibrium churning rule.

Stopping set $\mathbb{T}(\tilde{p}_0, l)$. The intermediary's expected value at time t from employing the agent of ex-ante ability $\tilde{p}_0$ while the clients believe that the average employed agent is $q_t = Q_t(l)$ and the agent who leaves the intermediary has ex-ante type l, i.e., $k_t = \pi_t(l)$ is

$$V_t(\tilde{p}_0, l) \overset{def}{=} \sup_{\tau} \mathrm{E}_{\tilde{p}_0, t<\eta}\bigg[\int_t^{\tau} e^{-r(s-t)} \left[A\left(Q_s(l) \cdot \xi_s\right) - w_R\left(\pi_s(\tilde{p}_0) \cdot \xi_s, \pi_s(l) \cdot \xi_s\right)\right] ds + e^{-r(\tau-t)} \cdot V\bigg], \tag{A.10}$$

where $\xi_t \overset{def}{=} \mathbb{1}\{\eta > t\}$ and V is the outside option of the intermediary. Since we take l as constant, it follows from (19) that $dC_t(\tilde{p}_0) = w_R(\pi_t(\tilde{p}_0) \cdot \xi_t, \pi_t(l) \cdot \xi_t)\, dt$, i.e., the lowest retention compensation is the agent's reservation wage. The intermediary optimally chooses a stopping time τ when to let the agent go. Since $Q_t(l_t) \cdot \xi_t = \pi_t(l_t) \cdot \xi_t = 0$ for $t > \eta$, the intermediary finds it unprofitable to retain the agent for $t > \eta$, implying that $\tau \leq \eta$ without loss.

Function $V_t(\tilde{p}_0, l)$ is weakly increasing in $\tilde{p}_0$ since *(i)* the reservation wage $w_R(\tilde{p}, k)$ is weakly decreasing in $\tilde{p}$ and *(ii)* the intermediary can attain the payoff corresponding to a lower ex-ante type $\tilde{p}_0' < \tilde{p}_0$ by choosing a mixed-strategy stopping rule τ that would mimic the arrival of a Poisson shock – the argument we formalized in Lemma A.1 above. Define

$$\mathbb{T}\left(\tilde{p}_0, l\right) \overset{def}{=} \{t < \eta :\ V_t(\tilde{p}_0, l) = V\} \tag{A.11}$$

as the set of times t when conditional on having reached time t, it is weakly optimal for the intermediary to let the agent go along the path of good performance. Weak monotonicity of $V_t(\tilde{p}_0, l)$ in $\tilde{p}_0$ implies that $\mathbb{T}\left(\tilde{p}_0', l\right) \subseteq \mathbb{T}\left(\tilde{p}_0, l\right)$ for all $\tilde{p}_0' \geq \tilde{p}_0$.

Lemma A.5. *Consider a weakly increasing continuous process $l = (l_t)_{t\geq 0}$ s.t. $dl_t = 0$ for every $t \notin \mathbb{T}\left(\tilde{p}_0, l_t\right)$. Then for any stopping time $\tau \in [t, \eta]$*[1]

$$V_t(\tilde{p}_0, l_t) \geq \mathrm{E}_{\tilde{p}_0,t}\left[\int_t^\tau e^{-r(s-t)}\left[A\left(Q_s(l_s)\right) - w_R(\pi_s(\tilde{p}_0), \pi_s(l_s)) - rV\right] \cdot \mathbb{1}\{s \notin \mathbb{T}\left(\tilde{p}_0, l_s\right)\}\, ds\right] + V. \tag{A.12}$$

where (A.12) *holds with equality if $V_\tau(\tilde{p}_0, l_\tau) = V$ or $\tau = \eta$.*

Proof. It is without loss to prove (A.12) for $t = 0$. Applying the standard martingale argument and Ito's lemma to (A.10) at $s \notin \mathbb{T}\left(\tilde{p}_0, l\right)$, obtain that for $s < \eta$ the continuation value $V_s(\tilde{p}_0, l)$ satisfies

$$rV_s(\tilde{p}_0, l) = A\left(Q_s(l)\right) - w_R\left(\pi_s(\tilde{p}_0), \pi_s(l)\right) + \dot{V}_s(\tilde{p}_0, l) + \lambda(1 - \pi_s(\tilde{p}_0)) \cdot \left(V - V_s(\tilde{p}_0, l)\right). \tag{A.13}$$

Define process $Z = (Z_s)_{s\geq 0}$ as

$$\begin{aligned} Z_s \overset{def}{=} \int_0^{s\wedge\eta} & e^{-ry} \cdot \left[A(Q_y(l_y)) - w_R(\pi_y(\tilde{p}_0), \pi_y(l_y)) - rV\right] \cdot \mathbb{1}\{y \notin \mathbb{T}\left(\tilde{p}_0, l_y\right)\}\, dy \\ & + e^{-rs} \cdot \left[V_y(\tilde{p}_0, l_y) - V\right] \cdot \mathbb{1}\{y < \eta\}\,. \end{aligned} \tag{A.14}$$

Applying Ito's lemma to (A.14) for $s \in \mathbb{T}\left(\tilde{p}_0, l_s\right) \cap [0, \eta)$, substituting $\dot{V}_s(\tilde{p}_0, l)$ from (A.13) and using the

[1] Note that in the definition of (A.10), process l is taken as constant, whereas the integrand in the right-hand side of (A.12) depends on the dynamics of process l.

fact that $dl_s = 0$ for $s \notin \mathbb{T}(\tilde{p}_0, l_s)$, obtain

$$\begin{aligned} e^{rs} \cdot dZ_s &= [A(Q_s(l_s)) - w_R(\pi_s(\tilde{p}_0), \pi_s(l))]\, ds - rV_s(\tilde{p}_0, l_s)\, ds + [V - V_t(\tilde{p}_0, l_s)]\, dN_s^\theta \\ &\quad + [rV_s(\tilde{p}_0, l_s) - \lambda(1-\pi_s(\tilde{p}_0)) \cdot (V - V_s(\tilde{p}_0, l_s)) + w_R(\pi_s(\tilde{p}_0), \pi_s(l_s)) - A(Q_s(l_s))]\, ds \\ &= [V - V_s(\tilde{p}_0, l_s)] \cdot \big[dN_s^\theta - \lambda(1-\pi_s(\tilde{p}_0))\, ds\big]. \end{aligned} \tag{A.15}$$

It follows from (A.15) that $\mathrm{E}_{s-}[dZ_s] = 0$ for $s \notin \mathbb{T}(\tilde{p}_0, l_s)$. For $t \in \mathbb{T}(\tilde{p}_0, l_s)$ or $s > \eta$ we have $V_s(\tilde{p}_0, l_s) = V$ which implies from (A.14) that $dZ_s = 0$. Consequently, $\mathrm{E}_{s-}[dZ_s] = 0 \quad \forall\, t$. As long as $e^{-\rho s} A(\pi_s(\overline{p})$ is bounded in s for some $\rho < r$, (A.14) implies that Z is an L^1 martingale. Optional stopping theorem $Z_0 = \mathrm{E}_{\tilde{p}_0}[Z_\tau]$ can be written for $\tau \le \eta$ as

$$\begin{aligned} V_0(\tilde{p}_0, l) - V &= \mathrm{E}\left[\int_0^\tau e^{-rs}\left[A(Q_s(l_s)) - w_R(\pi_s(\tilde{p}_0), \pi_s(l_s)) - rV\right] \mathbb{1}\{s \notin \mathbb{T}(\tilde{p}_0, l_s)\} ds + e^{-r\tau}\left[V_\tau(\tilde{p}_0, l_\tau) - V\right]\right] \\ &\le \mathrm{E}\left[\int_0^\tau e^{-rs}\left[A(Q_s(l_s)) - w_R(\pi_s(\tilde{p}_0), \pi_s(l_s)) - rV\right] \cdot \mathbb{1}\{s \notin \mathbb{T}(\tilde{p}_0, l_s)\}\, ds\right]. \end{aligned} \tag{A.16}$$

Inequality (A.16) follows from $V_\tau(\tilde{p}_0, l_\tau) \ge V$ and it is satisfied with equality if $V_\tau(\tilde{p}_0, l_\tau) = V$ or $\tau \ge \eta$. □

Lemma A.6 shows that if it is optimal to let the agent go at $t > 0$ along the path of good performance and this agent leaves the industry, then it was optimal to let this agent go before time t.

Lemma A.6. *Suppose $t \in \mathbb{T}(\tilde{p}_0, l)$ such that $U(\pi_t(\tilde{p}_0), \pi_t(l)) \le L$. Then $[0, t] \in \mathbb{T}(\tilde{p}_0, l)$.*

Proof. Suppose $\exists\, t \in \mathbb{T}(\tilde{p}_0, l)$ s.t. $U(\pi_t(\tilde{p}_0), \pi_t(l)) \le L$. Since $U(\tilde{p}, k)$ is increasing in $\tilde{p}$ and k, it follows that $U(\pi_s(\tilde{p}_0), \pi_s(l)) \le L$ and $w_R(\pi_s(\tilde{p}_0), \pi_s(l)) = rL \;\; \forall s \le t$. Hence $\forall s \le t$

$$A(Q_s(l)) - w_R(\pi_s(\tilde{p}_0), \pi_s(l)) - rV \overset{(i)}{<} A(Q_t(l)) - rL - rV \overset{(ii)}{\le} 0. \tag{A.17}$$

Inequality (i) in (A.17) follows from $Q_t(l)$ increasing in t. Inequality (ii) in (A.17) is necessary for $t \in \mathbb{T}(\tilde{p}_0, l)$ – otherwise the intermediary would *strictly* profitable to retain the agent at t.

Suppose, by contradiction, $\exists\, t_1 < t_2 \le t$ s.t. $(t_1, t_2) \cap \mathbb{T}(\tilde{p}_0, l) = \emptyset$. It follows from (A.13) that

$$\begin{aligned} V_{t_1}(\tilde{p}_0, l) - V &= \mathrm{E}_{\tilde{p}_0, t_1 < \eta}\left[\int_{t_1}^{\eta} e^{-r(s-t_1)}\left[A(Q_s(l)) - w_R(\pi_s(\tilde{p}_0), \pi_s(l)) - rV\right] \mathbb{1}\{s \notin \mathbb{T}(\tilde{p}_0, l)\}\, ds\right] \\ &= \underbrace{\mathrm{E}_{\tilde{p}_0, t_1 < \eta}\left[\int_{t_1}^{t \wedge \eta} e^{-r(s-t_1)} \cdot \left[A(Q_s(l)) - rL - rV\right] \mathbb{1}\{s \notin \mathbb{T}(\tilde{p}_0, l)\}\, ds\right]}_{<0 \text{ due to (A.17) and } (t_1, t_2) \cap \mathbb{T}(\tilde{p}_0, l) = \emptyset} \\ &\quad + e^{-r(t-t_1)} \cdot \mathrm{P}_{\tilde{p}_0, t_1 < \eta}(\eta > t) \cdot \underbrace{\left[V_t(\tilde{p}_0, l) - V\right]}_{=0 \text{ since } t \in \mathbb{T}(\tilde{p}_0, l)} < 0. \end{aligned} \tag{A.18}$$

Inequality (A.18) contradicts the existence of (t_1, t_2) such that $(t_1, t_2) \cap \mathbb{T}(\tilde{p}_0, l) = \emptyset$. $V_t(\tilde{p}_0, l)$ defined in (A.13) is continuous and set $\mathbb{T}(\tilde{p}_0, l)$ is, hence, closed. Consequently, $[0, t] \in \mathbb{T}(\tilde{p}_0, l)$. □

Hiring threshold. As on page 28 of the Print Appendix, denote

$$p_A \overset{def}{=} \inf\{p \geq \underline{p} : \quad U(p, p) \geq L\} \tag{A.19}$$

to be the lowest skilled agent willing to open his own firm if clients perceive him accurately. Additionally, define by l_R the lowest agent who the intermediary is willing to retain for a positive amount of time:

$$l_R \overset{def}{=} \left\{l \geq \underline{p} : \quad 0 \notin \mathbb{T}(l, l)\right\}. \tag{A.20}$$

Threshold l_R is defined using stopping set notation $\mathbb{T}(l, l)$ to account for the intermediary's optimal retention rule, which may be complex given a general shape of $A(\pi_t(p))$. If $A(\pi_t(p))$ is concave in t, then definition (A.20) coincides with the definition of l_R in (23) of the Pring Appendix.

Consider an initial cutoff p_A, which is the minimum between p_A and l_R defined above:

$$l_0^* = \min\{p_A,\, l_R\}. \tag{A.21}$$

Definition (A.21) is equivalent (25) if $A(\pi_t(p))$ is weakly concave in t, in which case the optimal stopping set $\mathbb{T}(l, l) = \{t \geq t^*(l)\}$.

Churning dynamics. Given initial l_0^* in (A.21) and $\gamma_t(l)$ defined in (26), construct process $l^* = (l_t^*)_{t \geq 0}$ forward via the differential equation

$$dl_t^* = \left[\gamma_t(l_t^*)/\pi_t'(l_t^*)\right] \cdot \mathbb{1}\{t \in \mathbb{T}(l_t^*, l_t^*)\}\, dt. \tag{A.22}$$

Denote by $\bar{t} \overset{def}{=} \inf\{t : \ l_t^* = \bar{p}\}$ the time when l^* reaches the top of the support of distribution $F(\cdot)$. Definition (A.22) is a first-order differential equation pinning down a candidate process $(l_t^*)_{t \geq 0}$.

Lemma A.7. *Process l^* defined via* (A.21) *and* (A.22) *is differentiable and weakly increasing.*

Proof. Process l^* is differentiable by construction. The necessary first-order optimality condition for $t \in \mathbb{T}(l, l)$ in (A.11) is $A(Q_t(l)) - w_R(\pi_t(l), \pi_t(l)) - rV \leq 0$ – otherwise it is strictly optimal for the intermediary to retain the agent at time t. If $l_0^* = p_A$ then it follows from the monotonicity of $\pi_t(p)$ in t that $w_R(\pi_t(l), \pi_t(l)) = A(\pi_t(l))$ for each $t \geq 0$, implying that $\gamma_t(l) \geq 0$. If, alternatively, $l_0^* < p_A$, then it follows from Lemma A.6 that $U(\pi_t(l), \pi_t(l)) \geq L$ for $t \in \mathbb{T}(l, l)$ which implies that $w_R(\pi_t(l), \pi_t(l)) = A(\pi_t(l))$ for $t \in \mathbb{T}(l, l)$. Consequently, $\gamma_t(l) \geq 0$ for $t \in \mathbb{T}(l, l)$. These arguments imply that the right-hand side of (A.22) is non-negative. □

Given process l^*, define a (churning) set $\mathbb{T}^* \stackrel{def}{=} \{t: \ t \in \mathbb{T}(l_t^*, l_t^*)\}$. By construction, $\mathbb{T}^*$ can replace $\mathbb{T}(l_t^*, l_t^*)$ in (A.22).

Lemma A.8. *Consider process l^* defined by* (A.21) *and* (A.22). *Client beliefs $k_t = \pi_t(l_t^*)$ and $q_t = Q_t(l_t^*)$ and intermediary's turnover strategy τ and retention wages $w_t(\tilde{p}_0)$ defined as*

$$
\begin{aligned}
\tau &= \min\{\inf\{t: \ \tilde{p}_0 \le l_t^* \ \textit{and} \ t \in \mathbb{T}\}, \ \eta\}, \\
w_t(\tilde{p}_0) &= w_R\left(\pi_t(\tilde{p}_0), \pi_t(l_t^*)\right) - \gamma_t(l_t^*) \cdot \partial_2 U\left(\pi_t(\tilde{p}_0), \pi_t(l_t^*)\right) \cdot \mathbb{1}\{t \in \mathbb{T}^*\},
\end{aligned}
\tag{A.23}
$$

constitute a Perfect Bayesian Equilibrium.

Proof. It follows from (19) and Lemma 1 that wage (A.23) satisfies the agent's retention constraint (7) with equality at every t. By construction, the intermediary retains all agents with ability $\tilde{p}_t > l_t^*$, implying that $q_t = \mathrm{E}\left[\tilde{p}_t \,|\, \tilde{p}_t > l_t^*, X_t\right] = Q_t(l_t^*) \cdot \mathbb{1}\{t < \eta\}$. Continuity and monotonicity of process l^* implies that $\mathrm{E}\left[\tilde{p}_0 | \tau = t\right] = l_t^*$ or, equivalently, $\mathrm{E}\left[\tilde{p}_t | \tau = t\right] = k_t$. Consequently, turnover time τ in (A.23) is Bayes-consistent with client beliefs $k_t = \pi_t(l_t^*)$ and $q_t = Q_t(l_t^*)$. Applying Ito's Lemma to the identity $k_t = \pi_t(l_t^*)$ and using (A.22) obtain

$$
dk_t = d\pi_t(l_t^*) = \dot{\pi}_t(l_t^*)dt + \pi_t'(l_t^*) \cdot dl_t^* = \underbrace{\lambda k_t(1-k_t)dt - k_t dN_t^\theta}_{\textit{learning from performance}} + \underbrace{\gamma_t(l_t^*) \cdot \mathbb{1}\{t \in \mathbb{T}^*\}}_{\textit{learning from churning}}. \tag{A.24}
$$

Suppose $\tilde{p}_0 < l_t^*$ and $t \notin \mathbb{T}^* \ \Leftrightarrow \ t \notin \mathbb{T}(l_t^*, l_t^*)$. For $t < \eta$, the intermediary's flow profit is

$$
\begin{aligned}
&\left[A\left(Q_t(l_t^*)\right) - w_t(\tilde{p}_0) - rV\right] \cdot \mathbb{1}\{t \notin \mathbb{T}(l_t^*, l_t^*)\} \\
=& \left[A\left(Q_t(l_t^*)\right) - w_R(\pi_t(\tilde{p}_0), \pi_t(l_t^*)) - rV\right] \cdot \mathbb{1}\{t \notin \mathbb{T}(\tilde{p}_0, l_t^*)\} \qquad \text{(A.25)} \\
&+ \left[A\left(Q_t(l_t^*)\right) - w_R(\pi_t(\tilde{p}_0), \pi_t(l_t^*)) - rV\right] \cdot \mathbb{1}\{t \in \mathbb{T}(\tilde{p}_0, l_t^*) \setminus \mathbb{T}(l_t^*, l_t^*)\} \\
\le& \left[A\left(Q_t(l_t^*)\right) - w_R(\pi_t(\tilde{p}_0), \pi_t(l_t^*)) - rV\right] \cdot \mathbb{1}\{t \notin \mathbb{T}(\tilde{p}_0, l_t^*)\}. \qquad \text{(A.26)}
\end{aligned}
$$

Equality (A.25) follows from $\mathbb{T}(l_t^*, l_t^*) \subseteq \mathbb{T}(\tilde{p}_0, l_t^*)$ for $\tilde{p}_0 \le l_t^*$ which follows from the monotonicity of $V_t(\tilde{p}_0, l)$ in $\tilde{p}_0$ established earlier. Inequality (A.26) follows from the necessary optimality condition that for $t \in \mathbb{T}(\tilde{p}_0, l_t^*)$ it must be that $A\left(Q_t(l_t^*)\right) - w_R\left(\pi_t(\tilde{p}_0), \pi_t(l_t^*)\right) \le rV$ since, otherwise, the intermediary would strictly benefit from retaining the agent at time t.

Suppose $\tilde{p}_0 < l_t^*$ and $t \in \mathbb{T}^* \ \Leftrightarrow \ t \in \mathbb{T}(l_t^*, l_t^*)$. By definition of l_0^* in (A.21), either $U(l_0^*, l_0^*) > L$, in which case $w_R(l_0^*, l_0^*) = A(l_0^*)$, or $0 \notin \mathbb{T}(l_0^*, l_0^*)$. If $0 \notin \mathbb{T}(l_0^*, l_0^*)$, then $U(\pi_t(l_0^*), \pi_t(l_0^*)) \ge L$ for $t \in \mathbb{T}(l_0^*, l_0^*)$ – if this was not the case then Lemma A.6 proves that $[0, t] \in \mathbb{T}(l_0^*, l_0^*)$ which contradicts $0 \notin \mathbb{T}(l_0^*, l_0^*)$. Consequently, $w_R(\pi_t(l_0^*), \pi_t(l_0^*)) = A(\pi_t(l_0^*))$ for all $t \in \mathbb{T}^*$. Since l_t^* is weakly increasing, as we've shown in Lemma A.7, it follows from the definition of the agent's reservation wage that $w_R(\pi_t(\tilde{p}_0), \pi_t(l_t^*)) = A(\pi_t(l_t^*)) \ \ \forall t \in \mathbb{T}$.

The intermediary's flow profit for $t \in \mathbb{T}^* \cap [0,\eta)$ is

$$\begin{aligned}&\left(A\left(Q_t(l_t^*)\right) - w_t(\tilde{p}_0) - rV\right)\cdot \mathbb{1}\{t \in \mathbb{T}\left(l_t^*, l_t^*\right)\} = \\ &\underbrace{\left(A\left(Q_t(l_t^*)\right) - A(\pi_t(l_t^*)) - rV\right)}_{\leq 0,\ turnover\ optimality\ t\in\mathbb{T}(l_t^*,l_t^*)}\cdot \underbrace{\left(1 - \frac{\partial_2 U\left(\pi_t(\tilde{p}_0), \pi_t(l_t^*)\right)}{\partial_2 U\left(\pi_t(l_t^*), \pi_t(l_t^*)\right)}\right)}_{\geq 0\ due\ to\ \tilde{p}_0<l_t^*\ and\ single\text{-}crossing}\cdot \mathbb{1}\{t \in \mathbb{T}\left(l_t^*, l_t^*\right)\} \leq 0. \end{aligned} \tag{A.27}$$

Inequality (A.27) follows from the necessary optimality condition for $t \in \mathbb{T}\left(l_t^*, l_t^*\right)$ discussed above and the fact that $\partial_2 U(\tilde{p}, k)$ is increasing in $\tilde{p}$, as established in (11) of the main text.

The intermediary's maximum expected continuation value from employing the agent of ability $\tilde{p}_0 \leq l_t^*$ from time $t < \eta$ onward is bounded by

$$V_t^*(\tilde{p}_0) \overset{def}{=} \sup_{\tau \leq \eta}\left\{\mathrm{E}_{\tilde{p}_0}\left[\int_t^\tau e^{-r(s-t)}\left[A\left(Q_s(l_s^*)\right) - w_s(\tilde{p}_0) - rV\right]ds\right] + V\right\} \leq \tag{A.28}$$

$$\sup_{\tau\leq\eta}\mathrm{E}_{\tilde{p}_0}\left[\int_t^\tau e^{-r(s-t)}\left[A(Q_s(l_s^*)) - w_R(\pi_s(\tilde{p}_0), \pi_s(l_s^*) - rV\right]\cdot \mathbb{1}\{s \notin \mathbb{T}\left(\tilde{p}_0, l_s^*\right)\}\,ds\right] + V \leq V_t(\tilde{p}_0, l_t^*), \tag{A.29}$$

where $w_s(\tilde{p}_0)$ is given by (A.23). Inequality (A.28) follows from (A.26) and (A.27). Inequality (A.29) follows from Lemma A.5 for $\tilde{p}_0 \leq l_t^*$ since $dl_s^* = 0$ for every $s \in [t, +\infty) \setminus \mathbb{T}\left(l_s^*, l_s^*\right) \subseteq [t, +\infty) \setminus \mathbb{T}\left(\tilde{p}_0, l_s^*\right)$. It follows that $V_t^*(\tilde{p}_0) \leq V_t(\tilde{p}_0, l_t^*) = V$ for $t \in \mathbb{T}\left(\tilde{p}_0, l_t^*\right)$ and $\tilde{p}_0 \leq l_t^*$, implying that it is weakly optimal for the intermediary to let go of the agent $\tilde{p}_0 \leq l_t^*$ and $t \in \mathbb{T}\left(\tilde{p}_0, l_t^*\right)$.

For $\tilde{p}_0 \geq l_t^*$ we reverse the above argument to obtain that the intermediary's expected continuation value $V_t^*(\tilde{p}_0) \geq V_t(\tilde{p}_0, l_t^*)$, implying that it is weakly optimal for the intermediary to retain the agent of ability $\tilde{p}_0 \geq l_t^*$ whenever $t \notin \mathbb{T}\left(\tilde{p}_0, l_t^*\right) \subseteq \mathbb{T}\left(l_t^*, l_t^*\right)$. This argument implies that at $\tilde{p}_0 = l_t^*$ we have $V_t^*(l_t^*) \geq V_t(l_t^*, l_t^*)$, whereas (A.28) and (A.29) imply that $V_t^* \leq V_t(l_t^*, l_t^*)$. Consequently, it follows that $V_t^*(l_t^*) = V_t(l_t^*, l_t^*)$. The latter equality is useful in generalizing the proof of equilibrium uniqueness (Observation 2) below.

Since for $\tilde{p}_0 < l_t^*$ it follows that $V_t^*(\tilde{p}_0) \leq V_t(l_t^*, l_t^*)$ it is weakly optimal to let go of type $\tilde{p}_0$ agent if $V_t(l_t^*, l_t^*) = V$, whereas if $\tilde{p}_0 > l_t^*$, then $V_t^*(\tilde{p}_0) \geq V_t(\tilde{p}_0, l_t^*)$ and it is weakly optimal to retain this agent. Consequently, turnover time τ, specified by (A.23), is incentive compatible for the intermediary given the necessary and sufficient retention wage $w_t(\tilde{p}_0)$ in (A.23). In particular, the above implies that it is optimal to let go of agents $\tilde{p}_0 < l_0^*$ at $t = 0$, i.e., not hire them to begin with. □

Promotion discontinuity. By continuity of the value function $V_t(l_t^*, l_t^*)$, it follows that $\mathbb{T}$ is closed. Without loss, we focus on the first churning period and the subsequent quiet period. Denote t^* to be the start of the first churning period:

$$t^* \overset{def}{=} \inf\{t:\ \exists\, \varepsilon > 0\ s.t.\ (t, t+\varepsilon) \in \mathbb{T}\}$$

Denote $\hat{t}$ to be the start of the quiet period that follows the first churning period:

$$\hat{t} \stackrel{def}{=} \inf\{t > t^* : \exists \varepsilon > 0 \ s.t. \ (t, t+\varepsilon) \notin \mathbb{T}\}.$$

The optimality of churning for $t \in (t^*, \hat{t})$ implies that $\gamma_{\hat{t}}(l_{\hat{t}}) \geq 0$ defined in (26), as we've shown in Lemma A.7. If $\gamma_{\hat{t}}(l_{\hat{t}}) > 0$, it follows from (A.22) and (A.23) that there is a discrete reduction in turnover at the transition from the churning to a quiet period occurring at time $\hat{t}$. We have shown as part of the proof of Lemma A.6 that $U(\pi_t(l_0^*), \pi_t(l_0^*)) \geq L$ for $t \in \mathbb{T} \cap (0, +\infty)$, implying that $w_R(\pi_t(\tilde{p}_0), \pi_t(l_{t^*})) = A(\pi_t(l_{t^*}))$ for $\tilde{p}_0 \geq l_{t^*}$ and $t \geq t^*$. For $t \in (\hat{t} - \varepsilon, \hat{t})$ the agent's wage is given by (A.23). The increase in the agent's compensation at $\hat{t}$ is, thus, equal to $\gamma_{\hat{t}}(l_{\hat{t}}) \cdot \partial_2 U(\pi_{\hat{t}}(\tilde{p}_0), \pi_{\hat{t}}(l_{\hat{t}}))$.

Lemma A.9 below shows that $\gamma_{\hat{t}}(l_{\hat{t}}) > 0$ for all, but a finite number of parameters λ if the revenue function $A(\cdot)$ is analytic. The intuition is that $\gamma_{\hat{t}}(l_{\hat{t}}) = 0$ requires the pair $(\hat{t}, l_{\hat{t}})$ to satisfy an over-determined system comprised of the intermediary's indifference over retaining the agent of type $l_{\hat{t}}$ both locally at time $\hat{t}$ and for the duration of the subsequent quiet period. Satisfying such an overdetermined system is infeasible for generic parameters. This is true even before we were to impose an additional independent condition that the state $(\hat{t}, l_{\hat{t}})$ needs to be reached from initial conditions via the dynamics specified by (A.21) and (A.22).

Lemma A.9 (Promotion discontinuity)**.** *For any analytical revenue function $A(\cdot)$ such that $e^{-\rho t} \cdot A(\pi_t(\bar{p}))$ is bounded for some $\rho < r$ there exists at most a finite set of values λ for which there exists $(\hat{t}, l_{\hat{t}})$ such that $\gamma_{\hat{t}}(l_{\hat{t}}) = 0$ and $(\hat{t} - \varepsilon, \hat{t}) \subseteq \mathbb{T}$ and $(\hat{t}, \hat{t} + \varepsilon) \cap \mathbb{T} = \emptyset$ for some $\varepsilon > 0$.*

Proof. In what follows, we show that $(\hat{t}, l_{\hat{t}})$ must be a solution to an overdetermined system that is degenerate for all but a finite number of learning intensities λ. We refer to $\gamma_{\hat{t}}(l_{\hat{t}}) = 0$ as a continuous promotion at time $\hat{t}$.

By definition of $\pi_t(l)$ and $Q_t(l)$, the log-likelihood ratios of these beliefs is given by

$$\begin{aligned} \ln\left(\frac{\pi_t(l)}{1-\pi_t(l)}\right) &= \ln\left(\frac{l}{e^{-\lambda t} \cdot (1-l)}\right) = \lambda t + \ln\left(\frac{l}{1-l}\right), \\ \ln\left(\frac{Q_t(l)}{1-Q_t(l)}\right) &= \ln\left(\frac{Q_0(l)}{e^{-\lambda t} \cdot (1-Q_0(l))}\right) = \lambda t + \ln\left(\frac{Q_0(l)}{1-Q_0(l)}\right). \end{aligned} \tag{A.30}$$

Given a belief p define function $a(\cdot)$ which maps belief p's log-likelihood ratio $\ln\left(\frac{p}{1-p}\right)$ to revenue $A(p)$:

$$a\left(\ln\left(\frac{p}{1-p}\right)\right) = A(p) \qquad \Leftrightarrow \qquad a(y) \stackrel{def}{=} A\left(\frac{e^y}{1+e^y}\right). \tag{A.31}$$

If function $A(\cdot)$ is analytic, then function $a(\cdot)$ is a superposition of two analytic functions and is, hence, also analytic. It follows from (A.30) and (A.31) that

$$A(\pi_t(l)) = a\left(\lambda t + \ln\left(\frac{l}{1-l}\right)\right), \qquad A(Q_t(l)) = a\left(\lambda t + \ln\left(\frac{Q_0(l)}{1-Q_0(l)}\right)\right). \tag{A.32}$$

By definition of $\gamma_t(l)$ in (26), the pair $\left(\hat{t}, l_{\hat{t}}\right)$ must satisfy $A\left(Q_{\hat{t}}\left(l_{\hat{t}}\right)\right) - A\left(\pi_{\hat{t}}\left(l_{\hat{t}}\right)\right) = rV$, which can be expressed as

$$a\left(\lambda\hat{t} + \ln\left(\frac{Q_0(l_{\hat{t}})}{1-Q_0(l_{\hat{t}})}\right)\right) - a\left(\lambda\hat{t} + \ln\left(\frac{l_{\hat{t}}}{1-l_{\hat{t}}}\right)\right) = rV. \tag{A.33}$$

Next, we show that is must be the case that

$$a'\left(\lambda\cdot\hat{t} + \ln\left(\frac{Q_0(l_{\hat{t}})}{1-Q_0(l_{\hat{t}})}\right)\right) - a'\left(\lambda\cdot\hat{t} + \ln\left(\frac{l_{\hat{t}}}{1-l_{\hat{t}}}\right)\right) = 0. \tag{A.34}$$

(i) Suppose, by contradiction,

$$a'\left(\lambda\cdot\hat{t} + \ln\left(\frac{Q_0(l_{\hat{t}})}{1-Q_0(l_{\hat{t}})}\right)\right) - a'\left(\lambda\cdot\hat{t} + \ln\left(\frac{l_{\hat{t}}}{1-l_{\hat{t}}}\right)\right) < 0,$$

which implies that $\exists\,\varepsilon_1 > 0$ such that for every $s \in (\hat{t}-\varepsilon_1, \hat{t})$

$$a\left(\lambda\cdot s + \ln\left(\frac{Q_0(l_{\hat{t}})}{1-Q_0(l_{\hat{t}})}\right)\right) - a\left(\lambda\cdot s + \ln\left(\frac{l_{\hat{t}}}{1-l_{\hat{t}}}\right)\right) > rV$$

In this case $\left(\hat{t}-\varepsilon_1, \hat{t}\right) \notin \mathbb{T}\left(l_{\hat{t}}, l_{\hat{t}}\right)$ implying that $\exists\,\varepsilon_2 > 0$ such that churning agents $(l_{\hat{t}}-\varepsilon_2, l_{\hat{t}})$ is strictly suboptimal prior to time $\hat{t}$.

(ii) Suppose, by contradiction

$$a'\left(\lambda\cdot\hat{t} + \ln\left(\frac{Q_0(l_{\hat{t}})}{1-Q_0(l_{\hat{t}})}\right)\right) - a'\left(\lambda\cdot\hat{t} + \ln\left(\frac{l_{\hat{t}}}{1-l_{\hat{t}}}\right)\right) > 0,$$

which implies that $\exists\,\varepsilon_1 > 0$ such that for every $s \in (\hat{t}, \hat{t}+\varepsilon_1)$

$$a\left(\lambda\cdot s + \ln\left(\frac{Q_0(l_{\hat{t}})}{1-Q_0(l_{\hat{t}})}\right)\right) - a\left(\lambda\cdot s + \ln\left(\frac{l_{\hat{t}}}{1-l_{\hat{t}}}\right)\right) > rV.$$

In this case, the intermediary's profit from retaining the agent at time $\hat{t}$ is

$$\mathrm{E}_{\tilde{p}_0 = l_{\hat{t}}, \eta > t}\left[\int_t^{\min\{t+\varepsilon_1, \eta\}} e^{-r(s-t)}\cdot\left[A\left(Q_s\left(l_{\hat{t}}\right)\right) - A\left(\pi_s\left(l_{\hat{t}}\right)\right) - rV\right]\,ds + V\right] > V,$$

which implies that $\exists\,\varepsilon_2 > 0$ such that churning agents $(l_{\hat{t}}-\varepsilon_2, l_{\hat{t}})$ is strictly suboptimal before $\hat{t}$.

The above two cases imply that t(A.34) must be satisfied at $(\hat{t}, l_{\hat{t}})$. Consider a supplementary (simplified) system of conditions:

$$\begin{cases} a(s+z) - a(s) = rV, \\ a'(s+z) - a'(s) = 0. \end{cases} \tag{A.35}$$

Lemma A.10. *If $a(\cdot)$ is analytic and bounded, then there exists at most a finite number of solutions (s, z) to* (A.35).

Proof. Suppose there is an infinite number of solutions to (A.35). Suppose, by contradiction, there exists a sequence of solutions $\{(s_n, z_n)\}$ to (A.35) that admit a limit (s_∞, z_∞). Since $a(\cdot)$ is analytic, then (s_∞, z_∞) satisfies the first two conditions of (A.35), but may fail the second order condition in the limit, even if all solutions (s_n, z_n) satisfy it. In what follows, we show that it must be the case that $a^{(j)}(s_\infty + z_\infty) = a^{(j)}(s_\infty)$ for every $j \in \mathbb{N}$.

Consider (s_n, z_n) and (s_{n+1}, z_{n+1}) satisfying (A.35). Suppose, without loss, that n is large enough so that $s_n + z_n > s_{n+1}$ and $s_{n+1} + z_{n+1} > s_n$. Since $a'(s_n) = a'(s_n + z_n)$ and $a'(s_{n+1}) = a'(s_{n+1} + z_{n+1})$, by the intermediary function theorem, there exist continuous mappings, e.g., time changes

$$y_1(\cdot) : [0, 1] \to [s_n), s_{n+1}], \qquad y_2(\cdot) : [0, 1] \to [s_n + z_n, s_{n+1} + z_{n+1}]. \tag{A.36}$$

such that $y_1(0) = s_n$, $y_1(1) = s_{n+1}$, $y_2(0) = s_n + z_n$, $y_2(1) = s_{n+1} + z_{n+1}$, and for every $x \in [0, 1]$

$$a'(y_2(x)) - a'(y_1(x)) = 0. \tag{A.37}$$

Since the above holds for all $x \in [0, 1]$, differentiating it with respect to s obtain

$$a''(y_2(x)) \cdot y_2'(x) - a''(y_1(x)) \cdot y_1'(x) = 0. \tag{A.38}$$

Additionally, by construction, $a(y_2(x)) - a(y_1(x)) = rV$ at $x = 0$ and $x = 1$. Since the difference $a(y_2(x)) - a(y_1(x))$ is equal to the same value at $x = 0$ and $x = 1$ it must contain a local maximum or minimum over $[0, 1]$, implying that there exists $\hat{x} \in [0, 1]$ such that $a'(y_2(\hat{x}) \cdot y_2'(\hat{x}) - a'(y_1(\hat{x}) \cdot y_1'(\hat{x}) = 0$. Note that by construction $a'(y_2(x)) = a'(y_1(x))$ for $x \in [0, 1]$, implying that the former condition implies that $y_1'(\hat{x}) = y_2'(\hat{x})$. Substituting it into (A.38) obtain that $a''(y_2(\hat{x})) = a''(y_1(\hat{x}))$. This argument, whilst winded, implies that for a given (s_n, z_n) and (s_{n+1}, z_{n+1}) there exists $\hat{s}_n \overset{def}{=} y_1(\hat{x}) \in [s_n, s_{n+1}]$ and $\hat{z}$ such that $\hat{s} + \hat{z} \overset{def}{=} y_2(\hat{x}) \in [s_n + z_n, s_{n+1} + z_{n+1}]$ such that

$$a'(\hat{s}_n) = a'(\hat{s}_n + \hat{z}_n) \qquad \textit{and} \qquad a''(\hat{s}_n) = a''(\hat{s}_n + \hat{z}_n). \tag{A.39}$$

Moreover, since $(s_n, z_n) \to (s_\infty, z_\infty)$ then $(\hat{s}_n, \hat{z}_n) \to (s_\infty, z_\infty)$. Repeating the above construction for higher order derivatives, obtain that $a^{(j)}(s_\infty) = a^{(j)}(s_\infty + z_\infty)$ for all $j \in \mathbb{N}$. Since function $a(\cdot)$ is analytic, it implies that $a(s+z_\infty)-a(s)$ is constant and, from the first condition of (A.35), is equal to rV. This, however, contradicts the fact that $A(\cdot)$ and, consequently, $a(\cdot)$, is bounded since $a(N\cdot z)=N\cdot a(z)\to\infty$ as $N\to\infty$. Consequently, if there exists an infinite set of solutions to (A.35), it cannot have a converging subsequence.

By the above argument, if there exists an infinite sequence of solutions to (A.35), then there must be a sequence of solutions $\{s_n^*, z_n^*\}$ such that either $s_n^*\to+\infty$ and/or $z_n^*\to+\infty$. If $A(\cdot)$ is bounded, which corresponds to $a(+\infty)$ being bounded, then the first equation in (A.35) requires that $a(s)\leq a(1)-rV$, requiring that any solution is uniformly bounded by $a^{-1}(a(1)-rV)$. Consequently, it must be the case that $z_n^*\to+\infty$ while $s_n^*\to s_\infty^*$. The second and third conditions (A.35) would then require that $a'(s_\infty^*)=0$ and $a''(s_\infty^*)\leq 0$, implying a contradiction with monotonicity of $a(\cdot)$. Consequently, there can be, at most, a finite number of solutions $(\hat{s},\hat{z})$ to (A.35). □

Any solution $\left(\hat{t}, l_{\hat{t}}\right)$ to a system of equations (A.33) and (A.34) corresponds to one of the finite number of solutions $(\hat{s},\hat{z})$ to (A.35) via

$$\begin{cases} \lambda\hat{t}+\ln\left(\dfrac{l_{\hat{t}}}{1-l_{\hat{t}}}\right)=\hat{s}, \\ \lambda\hat{t}+\ln\left(\dfrac{Q_0(l_{\hat{t}})}{1-Q_0(l_{\hat{t}})}\right)=\hat{s}+\hat{z}. \end{cases} \tag{A.40}$$

Conditions (A.40) imply

$$\ln\left(\frac{Q_0(l_{\hat{t}})}{1-Q_0(l_{\hat{t}})}\right)-\ln\left(\frac{l_{\hat{t}}}{1-l_{\hat{t}}}\right)=\hat{z}. \tag{A.41}$$

Denote the interval $\left[\hat{l}_1,\hat{l}_2\right]$ as a convex set of solutions to (A.41) given $\hat{t}$. Fix any $\hat{l}\in\left[\hat{l}_1,\hat{l}_2\right]$. The corresponding solution to (A.41) is $\hat{t}=\frac{1}{\lambda}\cdot\left[\hat{s}-\ln\left(\frac{\hat{l}}{1-\hat{l}}\right)\right]$ and $l_{\hat{t}}=\hat{l}$.

By definition of $\hat{t}$, it is the start of a quiet period. It follows from Lemma A.26 that, if $a(\cdot)$ is bounded, then this quiet period is finite – denote by $\hat{T}$ the start of the next churning period. The local optimality condition (A.33) must then be satisfied at $\hat{T}$, which pins down one of the possible values of $\hat{T}$. Consider the set of solutions to

$$\mathbb{S}(\hat{s})\overset{def}{=}\left\{S>\hat{S}:\quad a(S+\hat{z})-a(S)=rV\right\}, \tag{A.42}$$

where $\hat{z}$ was defined earlier. By continuity of $a(\cdot)$, set $\mathbb{S}(\hat{s})$ is closed. Moreover, since $(\hat{s},\hat{z})$ is a locally unique solution to (A.35), it follows that $\inf\mathbb{S}(\hat{s})>\hat{s}$. Consider a solution $\hat{S}\in\mathbb{S}(\hat{s})$ such that $\nexists\,\varepsilon>0$ such that $\left(\hat{S}-\varepsilon,\hat{S}\right)\in\mathbb{S}(\hat{s})$. Under such selection criterion, $\hat{S}$ is a locally unique solution to (A.42).[2] If $\hat{S}$ is a solution to (A.42), then $\hat{T}=\frac{1}{\lambda}\cdot\left[\hat{S}-\ln\left(\frac{\hat{l}}{1-\hat{l}}\right)\right]$ is a solution to (A.33) where $\hat{l}\in\left[\hat{l}_1,\hat{l}_2\right]$ was chosen earlier. For $\left(\hat{t},\hat{T}\right)$ to be a quiet period, there must be a solution $\hat{S}$ satisfying the above restrictions.

The quiet period $\left(\hat{t},\hat{T}\right)$ is incentive compatible for the intermediary if she finds it weakly optimal to retain

[2]Such selection is without loss since, in what follows, the intermediary would be indifferent across such solutions and, consequently, they generate identical expected continuation value.

the agent of ability $l_{\hat{t}}$ for every $t \in \left[\hat{t}, \hat{T}\right]$, which can be written as

$$\int_t^{\hat{T}} e^{-ry} \cdot \left[\hat{l} + \left(1-\hat{l}\right) e^{-\lambda y}\right] \cdot \left[a\left(\lambda y + \ln\left(\frac{Q_0\left(\hat{l}\right)}{1-Q_0\left(\hat{l}\right)}\right)\right) - a\left(\lambda y + \ln\left(\frac{\hat{l}}{1-\hat{l}}\right)\right) - rV\right] dy \geq 0 \quad \text{(A.43)}$$

which must be satisfied with equality at $t = \hat{t}$. Rewrite (A.43) by substituting

$$t = \frac{1}{\lambda} \cdot \left[s - \ln\left(\frac{\hat{l}}{1-\hat{l}}\right)\right] \qquad \Leftrightarrow \qquad s = \lambda t + \ln\left(\frac{\hat{l}}{1-\hat{l}}\right)$$

to obtain, using the definition of $\hat{z}$,

$$\int_s^{\hat{S}} e^{-\frac{r}{\lambda} y} \cdot \left(1 + e^{-y}\right) \cdot \left(a\left(y + \hat{z}\right) - a(y) - rV\right) \, dy \geq 0 \quad \text{(A.44)}$$

with equality at $s = \hat{s}$. The solution $\left(\hat{s}, \hat{z}, \hat{S}\right)$ is locally unique and, importantly, independent of λ. In what follows, we prove that there can be, at most, a finite number of solutions in λ to

$$G\left(1/\lambda\right) \overset{def}{=} \int_{\hat{s}}^{\hat{S}} e^{-\frac{r}{\lambda} y} \cdot \left(1 + e^{-y}\right) \cdot \left(a\left(y + \hat{z}\right) - a(y) - rV\right) \, dy = 0. \quad \text{(A.45)}$$

Function $G(1/\lambda)$ is a Laplace transform of a bounded continuous function $(1 + e^{-y}) \cdot (a(y + \hat{z}) - a(y) - rV) \cdot \mathbb{1}\left\{y \in \left[\hat{s}, \hat{S}\right]\right\}$. Consequently, $G(1/\lambda)$ is analytic for $\lambda \geq 0$. Suppose there exists a converging sequence $\{\lambda_n\}_{n\geq 1}$ such that $\lim_{n\to\infty} \lambda_n = \lambda_\infty$ and $G(1/\lambda_n) = 0$. By continuity, it implies that $G(1/\lambda_\infty) = 0$. Since $G(1/\lambda)$ is differentiable for $\lambda \geq 0$, the intermediate value theorem it follows that there exists $\lambda_n^1 \in (\lambda_n, \lambda_{n+1})$ such that $G'(1/\lambda_n^1) = 0$. By construction, $\lim_{n\to\infty} \lambda_n^1 = \lambda_\infty$, implying that $G'(1/\lambda_\infty) = 0$. We continue this argument by induction, implying that $G^{(k)}(\lambda_\infty) = 0$ for every $k \in \mathbb{N}$. Since $G(\cdot)$ is analytic, it implies that $G(\lambda) \equiv 0$, which would imply that $a(y + \hat{z}) - a(y) = rV$ for all $y \in \left[\hat{s}, \hat{S}\right]$. This argument implies that $\exists \, \varepsilon > 0$ such that any two solutions $G(\lambda) = G(\lambda') = 0$ satisfy $|\lambda' - \lambda| > \varepsilon$, which implies at most a countable number of solutions.

The maximum arrival intensity λ that satisfies (A.44) for all $s \in \left[\hat{s}, \hat{S}\right]$ is bounded from above. Consider a value $\hat{y} \in [\hat{s}, \hat{S}]$ such that $a\left(\hat{y} + \hat{z}\right) - a(\hat{y}) - rV < 0$. This value must exist for (A.44) to be satisfied with equality at $s = \hat{s}$. Then

$$\lim_{\lambda\to\infty} \left[\int_{\hat{y}}^{\hat{S}} \frac{1}{\lambda} \cdot e^{-\frac{r}{\lambda} y} \cdot \left(1 + e^{-y}\right) \cdot \left(a\left(y + \hat{z}\right) - a(y) - rV\right)\right] dy = a\left(\hat{y} + \hat{z}\right) - a(\hat{y}) - rV < 0.$$

Consequently, there exists a $\bar{\lambda}$ such that there does not exist a $\lambda > \bar{\lambda}$ for which $\left(\hat{s}, \hat{z}, \hat{S}\right)$ is consistent with a continuous promotion. Combined with the earlier derivation that any two solutions to $G(1/\lambda) = 0$ must be at least an ε away from each other implies that there is at most a finite number of values of λ for which $(\hat{s}, \hat{z}, \hat{S})$ is consistent with a continuous promotion at $\hat{t}$. Combined with the fact that there is, at most, a finite number of possible solutions $(\hat{s}, \hat{z})$, then there is, at most, a finite set of values λ which a continuous promotion may take place. □

In addition to the general parameters for which a promotion requires a discontinuity, there is an additional, independent requirement that we have not utilized due to its complexity. Specifically, if there is a continuous promotion at $\hat{t}$ then the pair $(\hat{t}, l_{\hat{t}})$ must be reached via (27) from the initial condition $t = 0$ and $l_0 = k_0$. Consequently, even if we were to take one of the finite learning intensity parameters λ for which $\left(\hat{t}, \hat{l}, \hat{T}\right)$ is consistent with a continuous promotion, it would require a simultaneous condition that such a history is reached along the equilibrium path. Consequently, for general revenue functions $A(\cdot)$ and distribution functions $F(\cdot)$, a transition from a churning to a quiet period features a promotion discontinuity.

Proof of Observation 1 (hiring threshold comparative statics)

We have shown the ranking of the intermediary's hiring threshold l_0^* and the agent's autarky entry threshold l_A. In what follows we complete the proof of Observation 1 by establishing the comparative statics of l_0^*.

Consider the equilibrium characterized by Proposition 2, with Proposition 1 being a special case for $A(\pi_t(\tilde{p}))$ being concave in t. The ex-ante type $l_0^* = \min\{p_A,\, l_R\}$ defined by (A.21) is the lowest skilled agent hired by the intermediary, where p_A was defined in (A.19) and l_R in (A.20).

Lemma A.11 (p_A properties)**.** *Threshold p_A is weakly increasing in the agent's outside option L, weakly decreasing in the informativeness λ, and is independent of the intermediary's outside option V.*

Proof. It follows directly from the definition of $p_A = \inf\{\tilde{p} \geq \underline{p} : U(p,p) \geq L\}$ that p_A is weakly increasing in L and is independent of V. Suppose $p_A > \underline{p}$, implying that $U(p_A, p_A) = L$.

Consider a higher arrival rate $\hat{\lambda} \geq \lambda$. Bayes rule implies that the Bayesian posterior from observing good performance given the higher informativeness $\hat{\lambda}$ is

$$\hat{\pi}_t(p) \overset{def}{=} \mathrm{P}\Big(\theta = 1 \,|\, \tilde{p}_0 = p, \eta > t, \hat{\lambda}\Big) = \frac{p}{p + (1-p)\cdot e^{-\hat{\lambda} t}} = \frac{p}{p + (1-p)\cdot e^{-\lambda \cdot t \cdot \hat{\lambda}/\lambda}} = \pi_{t\cdot\hat{\lambda}/\lambda}(p). \tag{A.46}$$

Denote by $\hat{U}(p,k)$ to be the agent's expected payoff from opening his firm and operating it until the arrival $\hat{\eta} \sim \mathrm{Exp}\left(\hat{\lambda}(1-\theta)\right)$:

$$\hat{U}(\tilde{p},k) \overset{def}{=} \mathrm{E}_{\tilde{p}}\bigg[\int_0^{\hat{\eta}} e^{-rt}\cdot\left[A\left(\hat{\pi}_t(k)\right) - rL\right]\, dt\bigg] + L = \mathrm{E}_{\tilde{p}}\bigg[\int_0^{\hat{\eta}} e^{-rt}\cdot\Big[A\left(\pi_{t\cdot\hat{\lambda}/\lambda}(k)\right) - rL\Big]\, dt\bigg] + L$$

The expected value to opening own firm given the higher arrival rate $\hat{\lambda}$ under the original belief p_A defined in (A.19) is

$$\begin{aligned}
\hat{U}(p_A, p_A) - L =& \mathrm{E}\left[\int_0^\infty e^{-rt} \cdot \left(p_A + (1-p_A)\, e^{-\hat{\lambda} t}\right) \cdot \left[A\left(\pi_{t \cdot \hat{\lambda}/\lambda}(k)\right) - rL\right] dt\right] \\
\overset{s=t\cdot\frac{\hat{\lambda}}{\lambda}}{=}& \frac{\lambda}{\hat{\lambda}} \cdot \mathrm{E}\left[\int_0^\infty e^{-r\cdot\frac{\lambda}{\hat{\lambda}}\cdot s} \cdot \left(p_A + (1-p_A)\cdot e^{-\lambda s}\right) \cdot \left[A\left(\pi_s(k)\right) - rL\right] ds\right] \\
\geq& \frac{\lambda}{\hat{\lambda}} \cdot \mathrm{E}\left[\int_0^\infty e^{-rs} \cdot \left(p_A + (1-p_A)\cdot e^{-\lambda s}\right) \cdot \left[A\left(\pi_s(k)\right) - rL\right] ds\right] \qquad (A.47) \\
=& \frac{\lambda}{\hat{\lambda}} \cdot \left[U(p_A, p_A) - L\right] \geq 0. \qquad (A.48)
\end{aligned}$$

Inequality (A.47) follows from the fact that the integrand in (A.47) is increasing, and a reduction in discounting increases this expected value. Inequality (A.48) follows from the definition of p_A. The ranking of $\hat{U}(p_A, p_A) \geq L$ whenever $U(p_A, p_A) \geq L$ implies that $\hat{p}_A \overset{def}{=} \inf\left\{p \geq \underline{p} :\ \hat{U}(p,p) \geq L\right\} \leq p_A$. □

Comparative statics with respect to L and V. If $l_0^* = l_R < p_A$ then it follows from the proof of Lemma A.12 (as part of the proof of equilibrium uniqueness in Observation 2 below) that $V(l_0, l_0)$ is increasing in l_0 at $l_0 = l_R < p_A$. As can be seen from (A.53), $V(l_0, l_0) - V$ is weakly decreasing in the intermediary's outside option V and the agent's outside option L. Consequently, differentiating the indifference condition $V(l_0, l_0) - V = 0$ obtain that l_0 is weakly increasing in V and L whenever $l_R < p_A$. If $l_0^* = p_A < l_R$, then it follows from Lemma A.11 that l_0^* is weakly increasing in L and independent of V. Combining the two cases obtain that $l_0^* = \min\{l_R, p_A\}$ is weakly increasing in the intermediary's outside option V and the agent's outside option L.

Comparative static with respect to λ. If $l_R > p_A$ then $l_0^* = p_A$ and Lemma A.11 shows that p_A it is decreasing in λ. Consider the remaining case when $l_R < p_A$. Bayes rule implies that

$$\hat{Q}_t(l) \overset{def}{=} \mathrm{E}\left[\tilde{p}_0 \,|\, \tilde{p}_0 \geq l,\ \hat{\lambda}, \eta > t\right] = \hat{\pi}_t(Q_0(l)) = \pi_{t\cdot\hat{\lambda}/\lambda}(Q_0(l)) = Q_{t\cdot\hat{\lambda}/\lambda}(l). \qquad (A.49)$$

The expected payoff of hiring the lowest skilled agent of ability $l_0 = l_R$ under the arrival rate $\hat{\lambda}$ is

$$\begin{aligned}
& \sup_\tau \mathrm{E}\left[\int_0^{\tau\wedge\eta} e^{-rt} \cdot \left[A\left(\hat{Q}_t(l_0)\right) - \hat{w}_R\left(\hat{\pi}_t(l_0), \hat{\pi}_t(l_0)\right) - rV\right] dt\right] \\
=& \sup_\tau \mathrm{E}\left[\int_0^{\tau\wedge\eta} e^{-rt} \cdot \left[A\left(Q_{t\cdot\hat{\lambda}/\lambda}(l_0)\right) - \hat{w}_R\left(\pi_{t\cdot\hat{\lambda}/\lambda}(l_0), \pi_{t\cdot\hat{\lambda}/\lambda}(l_0)\right) - rV\right] dt\right] \\
\geq& \sup_\tau \mathrm{E}\left[\int_0^{\tau} e^{-rt} \cdot \mathrm{P}(\tau > \eta) \cdot \left[A\left(Q_{t\cdot\hat{\lambda}/\lambda}(l_0)\right) - w_R\left(\pi_{t\cdot\hat{\lambda}/\lambda}(l_0), \pi_{t\cdot\hat{\lambda}/\lambda}(l_0)\right) - rV\right] dt\right] \qquad (A.50) \\
=& \sup_\tau \mathrm{E}\left[\int_0^{\tau} e^{-rt} \cdot \left(l_0 + (1-l_0)\cdot e^{-\hat{\lambda} t}\right) \cdot \left[A\left(Q_{t\cdot\hat{\lambda}/\lambda}(l_0)\right) - w_R\left(\pi_{t\cdot\hat{\lambda}/\lambda}(l_0), \pi_{t\cdot\hat{\lambda}/\lambda}(l_0)\right) - rV\right] dt\right] \\
\overset{s=t\cdot\frac{\hat{\lambda}}{\lambda}}{=}& \frac{\lambda}{\hat{\lambda}} \cdot \sup_\tau \mathrm{E}\left[\int_0^{\tau} e^{-r\cdot\frac{\lambda}{\hat{\lambda}}\cdot s} \cdot \left(l_0 + (1-l_0)\cdot e^{-\lambda s}\right) \cdot \left[A\left(Q_s(l_0)\right) - w_R\left(\pi_s(l_0), \pi_s(l_0)\right) - rV\right] ds\right]
\end{aligned}$$

$$= \frac{\lambda}{\hat{\lambda}} \cdot \sup_{\tau} \mathrm{E}\left[\int_0^{\tau\wedge\eta} e^{-r\cdot\frac{\lambda}{\hat{\lambda}}\cdot s} \cdot \left[A\left(Q_s(l_0)\right) - w_R\left(\pi_s(l_0), \pi_s(l_0)\right) - rV\right] ds\right] \tag{A.51}$$

$$\geq \frac{\lambda}{\hat{\lambda}} \cdot \sup_{\tau} \mathrm{E}\left[\int_0^{\tau\wedge\eta} e^{-rs} \cdot \left[A\left(Q_s(l_0)\right) - w_R\left(\pi_s(l_0), \pi_s(l_0)\right) - rV\right] ds\right]. \tag{A.52}$$

Inequality (A.50) follows from Lemma A.11 which implies that the agent's reservation wage $\hat{w}_R(\tilde{p}, k)$ under arrival rate $\hat{\lambda}$ is weakly lower than his reservation wage $w_R(\tilde{p}, k)$ under arrival rate λ. Inequality (A.52) follows from the argument, very similar to Lemma A.1 that the intermediary can use mixed-strategy stopping rules to replicate discount rate r when faced with the discount rate $r \cdot \frac{\lambda}{\hat{\lambda}} < r$. The above argument highlights that if it is strictly optimal for the intermediary to hire agent l_0 under arrival intensity λ, then it is strictly optimal to hire this agent under arrival intensity $\hat{\lambda}$. It follows that the lowest skilled agent $\hat{l}_R$ the intermediary is willing to retain for a positive amount of time under informativeness $\hat{\lambda}$ is lower than l_R.

Proof of Observation 2 (generalization to non-concave $A(\pi_t(p))$)

The following generalizes the proof of Observation 2 from the print appendix to general revenue functions $A(\cdot)$. This generalization relies on the proof and notation of Proposition 2. Lemmas 3 and 4, established in the Print Appendix, do not rely on concavity of $A(\pi_t(p))$ in t and, consequently, hold in the context of Proposition 2. We thus, only need to establish uniqueness for separating equilibria satisfying (37).

Lemma A.12 generalizes Lemma 5 for when $A(\pi_t(p))$ is not concave in t.

Lemma A.12 (Threshold p^*)**.** *The hiring threshold p^* is necessarily equal to l_0^* defined by* (A.21).

Proof. Denote by $V_0(\tilde{p}_0; l_0)$ to be the intermediary's expected value from hiring the agent of ex-ante type $\tilde{p}_0$ given the lowest hired agent is of type l_0:

$$V(\tilde{p}_0; l_0) = \sup_{\tau\leq\eta} \mathrm{E}_{\tilde{p}_0}\left[\int_0^{\tau} e^{-rt}(A(q_t) - rV)\, dt + e^{-r\tau} U(\pi_\tau(\tilde{p}_0), \pi_\tau(l_\tau)) - U^*(\tilde{p}_0, l_0) + V\right]. \tag{A.53}$$

If $V(l_0; l_0) > V$ and $l_0 > \underline{p}$ then, by continuity, $\exists\, \varepsilon > 0$ s.t. $V_0(l_0 - \varepsilon; l_0) > V$, implying that the intermediary would find it strictly optimal to hire the agent of ability $l_0 - \varepsilon$, which would contradict an interior hiring threshold l_0. Consequently, $V(l_0; l_0) = V$ if $l_0 > \underline{p}$.

Suppose, by contradiction, $l_0 > p_A$, where p_A was defined in (A.19). Bayes consistency (3) requires that clients assign belief $\mathrm{E}\left[\tilde{p}_0 | \tilde{p}_0 < l_0\right] < l_0$ to agents not hired by the intermediary. There exists $\varepsilon > 0$ s.t. $U\left(l_0 - \varepsilon, l_0\right) > U^*(l_0 - \varepsilon, \mathrm{E}\left[\tilde{p}_0 | \tilde{p}_0 < l_0\right])$. It follows from (A.53) that $\exists\, t_\varepsilon > 0$ s.t. the intermediary would find it strictly beneficial to hire and employ the agent of ex-ante ability $l_0 - \varepsilon$ until time t_ε along the path of good performance. This contradicts the fact that $U^*(l_0, l_0) > L$, implying that $l_0 \leq p_A$.

Suppose $l_0 < p_A$. Lemma A.1, combined with the minimal retention compensation $dC_t(\tilde{p}_0)$ decreasing in $\tilde{p}_0$, implies that $V(\tilde{p}_0; l_0)$ is weakly increasing in $\tilde{p}_0$ due to the possibility of using mixed termination strategies in (A.53). Since $U(l_0, l_0) < L$, it follows from Lemma 3 in the print Appendix that $\bar{t}(l_0) > 0$, defined in the Print Appendix in the paragraph that follows equation (36). Given the separating belief condition (37) derived in Lemma 4 of the Print Appendix, the local optimality condition requires $A(Q_t(l_0)) - A(\pi_t(l_0)) = rV$ at $t = \bar{t}(l_0)$. Differentiating (A.53) with respect to l_0 at $\tilde{p}_0 = l_0$ obtain

$$\partial_2 V(l_0; l_0) = \mathrm{E}_{p^*}\left[\int_0^{\bar{t}(l_0)\wedge\eta} e^{-rt} A'(Q_t(l_0)) Q_t'(l_0)\, dt + \int_{\bar{t}(l_0)\wedge\eta}^{\infty} e^{-rt} A'(\pi_t(l_0))\pi_t'(l_0)\, dt\right] > 0.$$

The above arguments imply that if $U(l_0, l_0) < L$ then $V(\tilde{p}_0; l_0)$ is locally increasing in $\tilde{p}_0$ and l_0, implying that $\frac{d}{dl_0} V(l_0; l_0) > 0$ whenever $U(l_0, l_0) \le L$. Consequently, there exists at most one solution to $V(l_0; l_0) = V$ for $U(l_0, l_0) < L$. If such a solution does not exist, it implies that $V(\underline{p}, \underline{p}) \ge V$, which coincides with (A.21) whenever $0 \notin \mathbb{T}\left(\underline{p}, \underline{p}\right)$. If a solution $V(l_0; l_0) = V$ does exist, then it follows from $V(\tilde{p}_0; \tilde{p}_0) > V$ for $\tilde{p}_0 > l_0$ that $l_0 = \inf\{\tilde{p}_0 \ge \underline{p} : 0 \notin \mathbb{T}(\tilde{p}_0, \tilde{p}_0)\}$. Consequently, threshold l_0 satisfies (A.21) and is, thus, uniquely determined. □

Lemma A.13 proves that the equilibrium process $(l_t)_{t\ge 0}$ must be absolutely continuous and almost everywhere continuously differentiable – a fact we use in the final step of the proof of Observation 2.

Lemma A.13. *An equilibrium belief process $(l_t)_{t\ge 0}$ must be absolutely continuous and almost everywhere differentiable.*

Proof. It follows from the proof of Lemma 4 that the equilibrium process $(l_t)_{t\ge 0}$ is continuous. The minimal compensation process $C(\tilde{p}_0)$ satisfies (20). The intermediary's expected value at time t, conditional on $t < \eta$, from optimally choosing when to let go of the agent is given by

$$V_t(\tilde{p}_0) \overset{def}{=} \sup_{\tau\le\eta} \mathrm{E}_{\tilde{p}_0, t<\eta}\left[\int_t^{\tau} e^{-r(s-t)}(A(Q_s(l_s)) - rV)ds + V + e^{-r(\tau-t)} U^*(\pi_\tau(\tilde{p}_0), \pi_\tau(l_\tau)) - U^*(\pi_t(\tilde{p}_0), \pi_t(l_t))\right],$$

It follows from Lemma and 3 and A.12 that $U(\pi_t(l_0^*), \pi_t(l_0^*)) \ge L$ for any $t \ge \underline{t}(l_0^*)$. Consequently, $U^*(\pi_t(\tilde{p}_0), \pi_t(l_t)) = U(\pi_t(\tilde{p}_0), \pi_t(l_t))$ for $t \ge \underline{t}(l_0^*)$.

It follows from the separating belief condition (37) that belief l_t satisfies $t \in \left[\underline{t}(l_t), \bar{t}(l_t)\right]$ so long as $t \ge \underline{t}(l_0)$. If $t < \bar{t}(l_t)$ then $\exists\, \varepsilon > 0$ s.t. $\underline{t}(l_t) + \varepsilon < \bar{t}(l_t)$. Belief condition (37) then requires $l_s = l_t$, i.e., $dl_s = 0$, for all $s \in [t, t+\varepsilon)$. Consider the remaining case $t = \bar{t}(l_t)$. For $V_t(l_t) = V$ it must be the case that it is weakly suboptimal to retain the agent until time $t' > t$:

$$0 \ge \mathrm{E}_{\tilde{p}_0 = l_t, t<\eta}\left[\int_t^{t'\wedge\eta} e^{-r(s-t)}(A(Q_s(l_s)) - rV)ds + e^{-r(t'\wedge\eta - t)} \cdot U^*(\pi_{t'}(\tilde{p}_0), \pi_{t'}(l_{t'})) - U(\pi_t(\tilde{p}_0), \pi_t(l_t))\right]$$

$$= \mathrm{E}_{l_t, t<\eta}\left[\int_t^{t'\wedge\eta} e^{-r(s-t)}[A(Q_s(l_s)) - rV - A(\pi_s(l_t))]\, ds + e^{-r(t'\wedge\eta - t)}\left[U^*(\pi_{t'}(l_t), \pi_{t'}(l_{t'})) - U(\pi_{t'}(l_t), \pi_{t'}(l_t))\right]\right]$$

Given the inequality above, the upper derivative as $t' \downarrow t$ of the right-hand side expression is bounded above by

$$0 \geq A(Q_t(l_t)) - rV - A(\pi_t(l_t)) + \pi_t'(l_t) \cdot \partial_2 U(\pi_t(l_t), \pi_t(l_t)) \cdot \overline{\lim}_{t'\downarrow t}\left[\frac{l_{t'} - l_t}{t' - t}\right]$$

$$\Rightarrow \qquad \overline{\lim}_{t'\downarrow t}\left[\frac{l_{t'} - l_t}{t' - t}\right] \leq \frac{rV + A(\pi_t(l_t)) - A(Q_t(l_t))}{\pi_t'(l_t) \cdot \partial_2 U(\pi_t(l_t), \pi_t(l_t))} = \frac{\gamma_t(l_t)}{\pi_t'(l_t)}, \tag{A.54}$$

where the last inequality in (A.54) follows from the definition of $\gamma_t(l)$ in (26). Suppose the upper bound inequality in (A.54) is strict. A similar argument applied to the optimality of retaining the agent of ability $\tilde{p}_0 = l_t$ from time t' until time $t \in \mathbb{T}^I(l_t)$ is

$$\overline{\lim}_{t'\uparrow t}\left[\frac{l_t - l_{t'}}{t - t'}\right] \geq \frac{rV + A(\pi_t(l_t)) - A(Q_t(l_t))}{\pi_t'(l_t) \cdot \partial_2 U(\pi_t(l_t), \pi_t(l_t))} = \frac{\gamma_t(l_t)}{\pi_t'(l_t)}. \tag{A.55}$$

Inequality (A.54) at $t = \bar{t}(l_t)$ and the fact that $dl_t = 0$ for $t < \bar{t}(l_t)$, coupled with the assumption that l is weakly increasing, imply that process l is absolutely continuous at t. Since l is weakly increasing, it is differentiable almost everywhere, and since it is absolutely continuous, it can be expressed as the integral of its derivative $l_t = \int_0^t \dot{l}_t\, dt$. It then follows from (A.54) and (A.55) that $\dot{l}_t = \frac{\gamma_t(l_t)}{\pi_t'(l_t)}$ at $t = \bar{t}(l_t)$. □

Lemma A.14. *The equilibrium constructed in Proposition 2 is unique: belief $l = (l_t)_{t\geq 0}$ satisfies* (A.21) *and* (A.22)*.*

Proof. Lemma A.12 proves that l_0 is uniquely pinned down by (A.21), i.e., $l_0 = l_0^*$. Lemma A.13 shows that the equilibrium process $l = (l_t)_{t\geq 0}$ is absolutely continuous and differentiable almost everywhere. If l is differentiable then, for $t < \eta$, (19) can be written as

$$dC_t(\tilde{p}_0) = w_R(\pi_t(\tilde{p}_0), \pi_t(l_t))\, dt - \pi_t'(l_t) \cdot \partial_2 U^*(\pi_t(\tilde{p}_0), \pi_t(l_t))\, dl_t. \tag{A.56}$$

Substitute (20) in (36) using (43) and take the first-order condition with respect to t to obtain that $\forall\, t \in \mathbb{T}^I(\tilde{p}_0)$, where $\mathbb{T}^I(\tilde{p}_0)$ was defined in (36) of the Print Appendix,

$$A(Q_t(l_t)) - rV - A(\pi_t(l_t)) + \dot{l}_t \cdot \pi_t'(l_t) \cdot \partial_2 U^*(\pi_t(\tilde{p}_0), \pi_t(l_t)) = 0. \tag{A.57}$$

Denote $V_t(\tilde{p}_0)$ to be the intermediary's continuation value at time t given client beliefs $(l_t)_{t\geq 0}$:

$$V_t(\tilde{p}_0) \overset{def}{=} \sup_{\tau\leq\eta} \mathrm{E}_{\tilde{p}_0, t<\eta}\left[\int_t^{\tau} e^{-r(s-t)}\left(A(Q_s(l_s))\, ds - dC_s(\tilde{p}_0)\right)\, ds + e^{-r\tau} \cdot V\right].$$

We argue in the paragraph following (36) of the Print Appendix and the proof of Lemma 4, i.e., inequalities (38) – (41), continuation value $V_t(\tilde{p}_0)$ is weakly increasing in $\tilde{p}_0$. Global optimality of letting go of the agent of type $\tilde{p}_0$ at time t requires that, in addition to the first-order condition (44), the intermediary's continuation value at time t is simply equal to V, i.e., $\mathbb{T}^I(\tilde{p}_0) \subseteq \{t : V_t(\tilde{p}_0) = V\} \subseteq \{t : V_t(\tilde{p}_0, l_t) = V\}$, where the latter set inclusion follows from $V_t(\tilde{p}_0) \geq V_t(\tilde{p}_0, l_t)$, where the latter was defined in (A.10), since l_t is weakly increasing.

Suppose $l_t > l_0$ and $V_t(l_t) > V$. It follows from the separating belief condition (37) that $V_s(\tilde{p}_0) \geq V_s(l_t) > V$ for all $s \in [0, t] \subset [0, \underline{t}(l_t+))$ for $\tilde{p}_0 > l_t$. Moreover, the separating belief condition (37) implies that $\underline{t}(l_t) < t$. Consequently, $\exists\, \varepsilon > 0$ s.t. $\underline{t}(l_t+) \geq t + \varepsilon > t - \varepsilon > \underline{t}(l_t)$. Condition (37) then implies that $\dot{l}_t = 0$ for all $t \in (t - \varepsilon, t + \varepsilon)$ for ε sufficiently low.

Suppose $l_t > l_0$ and $V_t(l_t) = V$. The separating property of l_t in (37) implies that $\underline{t}(l_t) \leq t \leq \bar{t}(l_t)$. This implies that, from an ex-ante perspective, the intermediary is indifferent between letting the agent of type $\tilde{p}_0 = l_t$ go at time t and $\bar{t}(l_t)$, implying that $t \in \mathbb{T}^I(l_t)$. Consequently, the local optimality condition (44) must then be satisfied at t.

The two cases above imply that

$$\dot{l}_t = \frac{\gamma_t(l_t)}{\pi'_t(l_t)} \cdot \mathbb{1}\{V_t(l_t) = V\} \leq \frac{\gamma_t(l_t)}{\pi'_t(l_t)} \cdot \mathbb{1}\{V_t(l_t, l_t) = V\}, \tag{A.58}$$

where the inequality follows from $\mathbb{1}\{V_t(l_t) = V\} \leq \mathbb{1}\{V_t(l_t, l_t) = V\}$ and the fact that $\gamma_t(l_t) \geq 0$ is a necessary condition for $V_t(l_t, l_t) = V$.

For a given t and l_t consider process $\hat{l}_t$ s.t. $\hat{l}_s = l_s$ for every $s \leq t$ and for every $s > t$ it is given by (A.24), rewritten here as

$$\dot{\hat{l}}_s = \left[\gamma_s\left(\hat{l}_s\right) / \pi'_s\left(\hat{l}_s\right)\right] \cdot \mathbb{1}\left\{V_s\left(\hat{l}_s, \hat{l}_s\right) = V\right\}.$$

It follows from (A.58) that $\dot{\hat{l}}_s \geq \dot{l}_s$ whenever $\hat{l}_s = l_s$, it follows that $\hat{l}_s \geq l_s$ for $s \geq t$. Note by $\hat{V}_t(\tilde{p}_0)$ as the continuation value of the intermediary given process $\hat{l}$ – it is defined identically to $V_t(\tilde{p}_0)$ defined for process l. Since $\hat{l}_t = l_t$ and $\hat{l}_s \geq l_s$ it follows that $\hat{V}_t(\tilde{p}_0) \geq V_t(\tilde{p}_0)$. At the same time, since $\hat{l}$ satisfies (A.58), it follows from the proof of Lemma A.8, i.e., equations (A.28), (A.29) and two subsequent paragraphs, that $\hat{V}_t(l_t) = V_t(l_t, l_t)$. Consequently, $V_t(l_t, l_t) \leq V_t(l_t) \leq \hat{V}_t(l_t) = V_t(l_t, l_t)$, implying that $V_t(l_t) = V_t(l_t, l_t)$. This implies that the inequality in (A.58) is satisfied with equality and, consequently, process l satisfies (A.58) subject to the initial condition (A.21) and must, thus, coincide with l^*. □

Proof of Corollary 2 (quiet period comparative statics)

Consider an arrival intensity $\hat{\lambda} > \lambda$. The posterior belief about ex-ante type $\tilde{p}_0 = p$ given good performance up to time t under the higher informativeness $\hat{\lambda}$ is given by $\hat{\pi}_t(p)$ defined in (A.46). Similarly, the posterior average about types $\tilde{p}_0 \geq p$ given higher informativeness $\hat{\lambda}$ is given by $\hat{Q}_t(p)$ defined in (A.49). If the agent's outside option $L = 0$ then, since $A(\cdot) \geq 0$, it follows that $p_A = \underline{p}$ and, consequently, $l_0^* = \underline{p}$, for both arrival rates $\hat{\lambda}$ and λ.

Define by $\hat{t}^*$ the first time such that the intermediary breaks even when employing the agent $\tilde{p}_0 = l_0^* = \underline{p}$, given arrival intensity $\hat{\lambda}$:

$$\begin{aligned}\hat{t}^* &\overset{def}{=} \inf\left\{t \geq 0 : \quad A\left(\hat{Q}_t(l_0^*)\right) - A\left(\hat{\pi}_t(l_0^*)\right) \leq rV\right\} \\ &= \inf\left\{t \geq 0 : \quad A\left(Q_{t\cdot\hat{\lambda}/\lambda}(l_0^*)\right) - A\left(\pi_{t\cdot\hat{\lambda}/\lambda}(l_0^*)\right) \leq rV\right\} \\ &= \frac{\lambda}{\hat{\lambda}} \cdot \inf\left\{s \geq 0 : \quad A\left(Q_s(l_0^*)\right) - A\left(\pi_s(l_0^*)\right) \leq rV\right\} = \frac{\lambda}{\hat{\lambda}} \cdot t^* < t^*.\end{aligned}$$

Consequently, the length of the quiet period $\hat{t}^*$ under arrival intensity $\hat{\lambda} > \lambda$, is equal to $\frac{\lambda}{\hat{\lambda}} \cdot t^* < t^*$.

Consider distribution $\hat{F}(\cdot)$ dominates distribution $F(\cdot)$ in the Blackwell informativeness sense, i.e., that private prior $\hat{\tilde{p}}_0$ under distribution $\hat{F}(\cdot)$ is a mean-preserving spread of the private prior $\tilde{p}_0$ under distribution $F(\cdot)$. Mean preservation implies that $\mathrm{E}\left[\hat{\tilde{p}}_0\right] = \mathrm{E}\left[\tilde{p}_0\right]$, while the mean-preserving spread implies that the values of $\hat{\tilde{p}}_0$ are more spread out compared to the values of $\tilde{p}_0$. In particular, the lower bound of the support of distribution $\hat{F}(\cdot)$ must be lower than that of distribution $F(\cdot)$, formally captured as

$$\underline{\hat{p}} \overset{def}{=} \inf\left\{p \geq 0 : \ \hat{F}(p) > 0\right\} \leq \inf\left\{p \geq 0 : \ F(p) > 0\right\} \overset{def}{=} \underline{p}.$$

The following inequality then shows that the quiet period length $\hat{t}^*$ is longer under distribution $\hat{F}(\cdot)$ compared to the quiet period length t^* under the original distribution $F(\cdot)$:

$$\hat{t}^* = \left\{t \geq 0 : \quad A\left(\pi_t\left(\mathrm{E}\left[\tilde{p}\right]\right)\right) - A\left(\pi_t\left(\underline{\hat{p}}\right)\right) \leq rV\right\} \geq \left\{t \geq 0 : \quad A\left(\pi_t\left(\mathrm{E}\left[\tilde{p}\right]\right)\right) - A\left(\pi_t\left(\underline{p}\right)\right) \leq rV\right\} = t^*.$$

Proof of Corollary 3 (profitable exits)

At the time when the agent of posterior ability $\tilde{p}_t = k_t$ is let go, the agent's wage, while working for the intermediary, is given by (10). The compensation increase for the departing agent $p_t = k_t$ is equal to

$$\gamma_t \cdot \partial_2 U(k_t, k_t) = rV - A(Q_t(l_t^*)) + A(\pi_t(l_t^*)). \tag{A.59}$$

Differentiating (A.59) with respect to t obtain

$$\begin{aligned}\frac{d}{dt}\left[\gamma_t \cdot \partial_2 U(k_t, k_t)\right] &= A'(\pi_t(l_t^*)) \cdot \left[\pi_t'(l_t^*) \cdot \dot{l}_t^* + \dot{\pi}_t(l_t^*)\right] - A'(Q_t(l_t^*)) \cdot \left[Q_t'(l_t^*) \cdot \dot{l}_t^* + \dot{Q}_t(l_t^*)\right] \\ &= \left[A'(\pi_t(l_t^*)) \cdot \pi_t'(l_t^*) - A'(Q_t(l_t^*)) \cdot Q_t'(l_t^*)\right] \cdot \dot{l}_t^* + A'(\pi_t(l_t^*)) \cdot \dot{\pi}_t(l_t^*) - A'(Q_t(l_t^*)) \cdot \dot{Q}_t(l_t^*)\end{aligned}$$

$$\geq \left[A'(\pi_t(l_t^*)) \cdot \pi_t'(l_t^*) - A'(Q_t(l_t^*)) \cdot \pi_t'(Q_0(l_t^*)) \cdot Q_0'(l_t^*)\right] \cdot \dot{l}_t^*. \tag{A.60}$$

Inequality (A.60) follows from weak concavity of $A(\pi_t(p))$ in t for $p \geq \underline{p}$, which manifests as

$$A'(\pi_t(l_t^*)) \cdot \dot{\pi}_t(l_t^*) \geq A'(Q_t(l_t^*)) \cdot \dot{Q}_t(l_t^*). \tag{A.61}$$

Differentiating $\pi_t(p)$ with respect to p obtain

$$\pi_t'(p) = \frac{1-e^{-\lambda t}}{(p+(1-p)\cdot e^{-\lambda t})^2} = \frac{1-e^{-\lambda t}}{p(1-p)\cdot e^{-\lambda t}} \cdot \dot{\pi}_t(p) \Rightarrow \qquad \dot{\pi}_t(p) = \frac{p(1-p)\cdot e^{-\lambda t}}{1-e^{-\lambda t}} \cdot \pi_t'(p).$$

Note that $Q_t(l) = \pi_t(Q_0(l))$ and, consequently, $\dot{Q}_t(l) = \dot{\pi}_t(Q_0(l))$. Substituting $\dot{\pi}_t(p)$ from (A.62) into the concavity condition (A.61) obtain the bound on $\pi_t'(\cdot)$:

$$A'(\pi_t(l_t^*)) \cdot \frac{l_t^*(1-l_t^*)\cdot e^{-\lambda t}}{1-e^{-\lambda t}} \cdot \pi_t'(l_t^*) \geq A'(Q_t(l_t^*)) \cdot \frac{Q_0(l_t^*)(1-Q_0(l_t^*))\cdot e^{-\lambda t}}{1-e^{-\lambda t}} \cdot \pi_t'(Q_0(l_t^*))$$

$$A'(Q_t(l_t^*)) \leq A'(\pi_t(l_t^*)) \cdot \frac{l_t^*(1-l_t^*)}{Q_0(l_t^*)(1-Q_0(l_t^*))} \cdot \frac{\pi_t'(l_t^*)}{\pi_t'(Q_0(l_t^*))}. \tag{A.62}$$

If $Q_0(l_t^*) < 1$ then the right hand side (A.62) is bounded. Substituting (A.62) into (A.60) obtain

$$\frac{d}{dt}\left[\gamma_t \cdot \partial_2 U(k_t, k_t)\right] \geq A'(\pi_t(l_t^*)) \cdot \dot{\pi}_t(l_t^*) \cdot \frac{e^{-\lambda t}}{1-e^{-\lambda t}} \cdot \dot{l}_t^* \cdot \left[1 - \frac{l_t^*(1-l_t^*)}{Q_0(l_t^*)(1-Q_0(l_t^*))} \cdot Q'(l_t^*)\right]. \tag{A.63}$$

The necessary condition for the right hand side of (A.63) to be positive is

$$Q_0'(l) \leq \frac{Q_0(l)(1-Q_0(l))}{l(1-l)} \qquad \forall l \in [\underline{p}, \overline{p}]. \tag{A.64}$$

A sufficient condition for (A.64) to be satisfied is for $\ln\left[\frac{Q_0(l)}{1-Q_0(l)}\right] - \ln\left[\frac{l}{1-l}\right]$ to be decreasing in l.

Suppose that $\tilde{p}_0 \sim U\left[\underline{p}, \overline{p}\right]$, i.e., $F(p) = \frac{p-\underline{p}}{\overline{p}-\underline{p}}$. Then $Q_0(l) = \frac{1}{2}\cdot(l+\overline{p})$ and (A.64) is always satisfied, as can be seen from[3]

$$\frac{1}{2} < \frac{(l+\overline{p})\cdot(2-l-\overline{p})}{4l\cdot(1-l)} \qquad \forall l \in \left[\underline{p}, \overline{p}\right]. \tag{A.65}$$

As the inequality (A.65) is strict, it implies that it is satisfied for a density function $f(p)$ that is sufficiently close to $\frac{1}{\overline{p}-\underline{p}}$, which can be captured by a sufficiently low Lipschitz constant of $f(\cdot)$.

[3]The right-hand side of the inequality is a downward facing parabola in $\overline{p}$. The inequality is satisfied at $\overline{p} = l$ and $\overline{p} = 1$, implying that it is satisfied for intermediate values as well.

Intermediary's Endogenous Outside Option V: Uniqueness and Properties

The intermediary's equilibrium expected value satisfies

$$V = \max\left\{\mathrm{E}\left[\int_0^\tau e^{-rt}\cdot[A(q_t) - w_t(\tilde{p}_t) - rV]\ dt + V\right] - I,\ 0\right\} = \max\{G(V) - I + V,\ 0\},\qquad \text{(A.66)}$$

where $G(V)$ is defined by (14). If $V > 0$, this necessarily requires that $G(V) = I$. However, if the replacement cost I is so large that hiring a new agent is unprofitable, then $V = 0$, which is captured by the max operator in (A.66).

Lemma A.15 (Unique fixed point)**.** *Suppose $I > 0$. The solution to* (A.66) *is unique if λ is sufficiently low.*

Proof. Suppose $\lambda = 0$. If $A(q_0) - \max\{A(k_0),\ rL\} - rV \geq 0$, then the intermediary employs this agent in perpetuity, generating the expected value $G(V)$ in (14) equal to

$$rG(V) = A(q_0) - \max\{rL,\ A(k_0)\} - rV \qquad \textit{if} \qquad A(q_0) - \max\{rL,\ A(k_0)\} - rV \geq 0. \qquad \text{(A.67)}$$

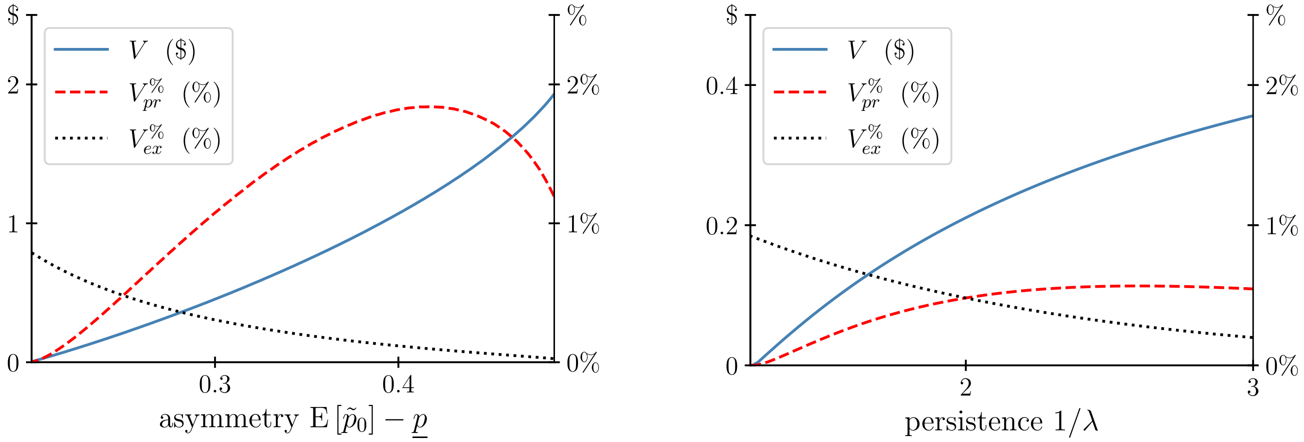
The expected value of the agent is given by $U(p,k) = A(k)/r$ and is independent of his private type $\tilde{p}$. The marginal value of reputation is, consequently, constant in private type $\tilde{p}$ and equal to $\partial_2 U(\tilde{p},k) = A'(k)/r$. If $A(q_0) - \max\{rL,\ A(k_0)\} - rV < 0$, then, once churning begins, the intermediary is indifferent between retaining all agents and letting them go. Consequently, the intermediary's net profit from employing every agent is 0. Consequently, if $\lambda = 0$, then the expected gain $G(V)$ from employing an agent is given in closed form as

$$G_{\lambda=0}(V) = \frac{1}{r}\cdot\left[\max\{A(q_0) - \max\{rL,\ A(k_0)\} - rV,\ 0\right]. \qquad \text{(A.68)}$$

Function $G_{\lambda=0}(V)$ in (A.68) is strictly decreasing for $rV < A(q_0) - \max\{rL,\ A(k_0)\}$ and equal to 0 for $rV \geq A(q_0) - \max\{rL,\ A(k_0)\}$. Consequently, if $I > 0$, then if a solution to $G_{\lambda=0}(V) = I$ exists, then it must be unique. The equilibrium we construct in Proposition 2 satisfies (A.21) and (A.22) and is continuous in λ. Consequently, there exists a $\bar{\lambda}(I)$ such that there exists a unique solution $G(V) = I$ for any $\lambda < \bar{\lambda}(I)$. □

Figure A.1 illustrates that the intermediary's equilibrium expected value V (solid line) is increasing in both ex-ante informational asymmetry $\mathrm{E}[\tilde{p}] - \underline{p}$ and its persistence $1/\lambda$. Define $V_{pr}^{\%}$ to be the fraction of the intermediary's expected profits arising from the agents paying to build a reputation

$$V_{pr}^{\%} \stackrel{def}{=} \underbrace{\mathrm{E}\left[\int_0^\tau e^{-rt}\cdot[A(k_t) - w_t(\tilde{p}_t)]\ dt\right]}_{\textit{intermediary profit from pay for reputation}} \Big/ \underbrace{\mathrm{E}\left[\int_0^\tau e^{-rt}\cdot[A(q_t) - w_t(\tilde{p}_t)]\ dt\right]}_{\textit{intermediary total profit}} \qquad \text{(A.69)}$$



(a) Equilibrium value V and percentage of revenues obtained from pay-for-reputation. $F(\cdot) \sim U[\underline{p}, 1-\underline{p}]$

(b) Equilibrium value V and percentage of revenues obtained from pay-for-reputation.

Figure A.1: Comparative statics with respect to information asymmetry $\mathrm{E}[\tilde{p}] - \underline{p}$. For the parameters considered, the intermediary hires all initial types, i.e., $\underline{p} = k_0$, and the churning set is given by $\mathbb{T} = [t^*, \bar{t}]$. Parameters: $A(p) \equiv p$, $F(\cdot) \sim U[0.25, 0.75]$, $r = 0.25$, $\lambda = 0.5$, $L = 0$, $I = 0.4$.

We see in Figure A.1a that $V_{pr}^{\%}$ (dashed line) exhibits a hump-shape pattern. There is very little churning for extreme levels of asymmetry, and thus, $V_{pr}^{\%}$ is close to 0. Consequently, the fraction of the intermediary's profits from the agents paying for reputation is maximized for intermediate levels of informational asymmetry and persistence. Considering the equilibrium value of V, obtained in (14), is critical for these insights. If one were to take V as exogenous, then the fraction of profits obtained from paying for reputation would be decreasing in information asymmetry and persistence, as can be seen from the behavior of $V_{ex}^{\%}$ in Figure A.1, which takes V as exogenous in calculating (A.69).

If $L = 0$, then, for numerical calculations it is convenient to express $V_{pr}^{\%}$ as

$$
\begin{aligned}
V_{pr}^{\%} &= \frac{\mathrm{E}\left[\int_0^\tau e^{-rt} \cdot \gamma_t \cdot \partial_2 U(\tilde{p}_t, k_t)\, dt\right]}{\mathrm{E}\left[\int_0^\tau e^{-rt} \cdot \left[A(q_t) - A(k_t) + \gamma_t \cdot \partial_2 U(\tilde{p}_t, k_t)\right]\, dt\right]} \\
&= \frac{\int_0^\infty e^{-rt} \cdot (1 - F(l_t)) \cdot \left(q_t + (1-q_t) \cdot e^{-\lambda t}\right) \cdot \gamma_t \cdot \partial_2 U(q_t, k_t)\, dt}{\int_0^\infty e^{-rt} \cdot (1 - F(l_t)) \cdot (q_t + (1-q_t) \cdot e^{-\lambda t}) \cdot \left[A(q_t) - A(k_t) + \gamma_t \cdot \partial_2 U(q_t, k_t)\right]\, dt}.
\end{aligned}
\tag{A.70}
$$

Proof of Observation 3 (independent signaling)

To simplify the derivations below, we assume, without loss, that $L = 0$. If $L > 0$, then agents of ability $\tilde{p}_0 < p_A$, defined in (A.19), will prefer leaving the industry rather than attempting to participate in the signaling game. If $p_A = \underline{p}$, we can include all agents in the argument below. Alternatively, if $p_A > \underline{p}$, then we can effectively relabel the lowest initial type to be equal to p_A in the signaling subgame below and consider an alternative revenue function $\hat{A}(p) = A(p) - rL$, which factors out the agent's opportunity cost.

For this reasons, it is without loss to conduct the proof of Observation 3 under the assumption that $L = 0$.

Definition 1 (Perfect Bayesian Equilibrium of the Signaling Game)**.** *The equilibrium is a collection of contracts* $\alpha = \{\alpha_t\}_{t\geq 0} \in [0,1]^\infty$ *and client beliefs* $\Pi : [0,1]^\infty \to [\underline{p}, 1]$ *such that*

(i) Contract optimality: the agent of type $\tilde{p}$ *chooses among contracts* $\alpha^* \in \Delta(\tilde{p})$ *that maximize his expected payoff given beliefs* $\Pi(\cdot)$*:*

$$\Delta(\tilde{p}) \stackrel{def}{=} \arg\max_{\alpha} \left\{ \int_0^\infty e^{-rt} \cdot \left(\tilde{p} + (1-\tilde{p}) \cdot e^{-\lambda t}\right) \cdot \alpha_t \cdot A\left(\pi_t(\Pi(\alpha))\right) dt \right\}.$$

(ii) Belief consistency: for each contract α *chosen with positive probability by any agent, the client belief* $\Pi(\alpha)$ *is Bayes consistent and given by* $\Pi(\alpha) \stackrel{def}{=} \mathrm{E}\left[\tilde{p} \,|\, \alpha(\tilde{p}) = \alpha\right]$*.*

We allow the agent to commit to long-term contracts with each client in order to set up possible subgame signaling as a static game. This allows us to discipline beliefs for contracts not offered in equilibrium using the D1 refinement, e.g., Cho and Sobel (1990).

Definition 2 (Subgame Signaling D1 Belief Refinement)**.** *Consider the expected value* $H(\tilde{p})$ *obtained by agent type* $\tilde{p}$ *in equilibrium. For a contract* α *not offered in equilibrium compute* $\Pi^I(\alpha, \tilde{p})$ *such that the agent of skill* $\tilde{p}$ *is exactly indifferent between offering contract* α *and be perceived as* $\Pi(\alpha, \tilde{p})$ *and his equilibrium payoff* $H(\tilde{p})$

$$\int_0^\infty e^{-rt} \cdot \left(\tilde{p} + (1-\tilde{p}) \cdot e^{-\lambda t}\right) \cdot \alpha_t \cdot A\left(\pi_t\left(\Pi^I(\alpha, \tilde{p})\right)\right) dt = H(\tilde{p}).$$

The equilibrium beliefs $\Pi(\cdot)$ *then satisfy the D1 belief refinement if belief* $\Pi(\alpha)$ *for a contract* α *that is not offered in equilibrium places* 0 *probability* 0 *on types* $\tilde{p}$ *for which the indifference belief is not minimal, i.e., types for which* $\Pi^I(\alpha, \tilde{p}) > \min_{\hat{p}}\left\{\Pi^I(\alpha, \hat{p})\right\}$*.*

The payoff of the agent is strictly increasing in the client's belief for any fee structure α offered by the agent. For this reason, it is sufficient to compare the indifference belief threshold $\Pi^I(\alpha, \tilde{p})$ across types, rather than ranking sets of beliefs. Following Cho and Sobel (1990) this implies that the D1 belief refinement is equivalent to Independence of Never Weak Best Responses and Universal Divinity. In what follows, we show that there is a unique Perfect Bayesian Equilibrium for the agent-client signaling game. This equilibrium is separating in types and is the most efficient separating equilibrium.

We begin by establishing an auxiliary single-crossing result that is standard in security design.

Lemma A.16 (Single crossing)**.** *Consider two differentiable cumulative dividend processes* Y *and* $\hat{Y}$ *such that* $dY_t = y_t\, dt$ *and* $d\hat{Y}_t = \hat{y}_t$*. Moreover,* $\exists\, T > 0$ *s.t.* $y_t \geq \hat{y}_t$ *for* $t \leq T$ *and* $y_t \leq \hat{y}_t$ *for* $t \geq T$*. Suppose the agent of ex-ante type* $\tilde{p}_0 = p$ *is indifferent between cash flows* y *and* $\hat{y}$*. Then*

(i) every type $\hat{p} < p$ *(*$\hat{p} > p$*) weakly prefers dividend process* Y *to* $\hat{Y}$ *(*$\hat{Y}$ *to* Y*);*

(ii) such preference is strict whenever $y_t \neq \hat{y}_t$ on a set of positive Lebesgue measure.

Proof. For $p' < \tilde{p}_0$ function $\frac{p'+(1-p')\cdot e^{-\lambda t}}{\tilde{p}_0+(1-p)\cdot e^{-\lambda t}}$ is decreasing in t as can be seen from the following simplification

$$\frac{p' + (1-p')\cdot e^{-\lambda t}}{\tilde{p}_0 + (1-\tilde{p}_0)\cdot e^{-\lambda t}} = \frac{1-p'}{1-\tilde{p}_0} \cdot \frac{p' \cdot \frac{1-\tilde{p}_0}{1-p'} + (1-\tilde{p}_0)\cdot e^{-\lambda t}}{\tilde{p}_0 + (1-\tilde{p}_0)\cdot e^{-\lambda t}} = \frac{1-p'}{1-\tilde{p}_0} \cdot \left(\frac{p' \cdot \frac{1-\tilde{p}_0}{1-p'} - \tilde{p}_0}{\tilde{p}_0 + (1-\tilde{p}_0)\cdot e^{-\lambda t}} + 1 \right)$$
$$= \frac{1-p'}{1-\tilde{p}_0} \cdot \left(1 - \frac{\tilde{p}_0 - p'}{1-p'} \cdot \frac{1}{\tilde{p}_0 + (1-\tilde{p}_0)\cdot e^{-\lambda t}} \right).$$

The difference in expected values from cash flows y and $\hat{y}$ for type $p' < \tilde{p}_0$ is

$$\begin{aligned}
&\int_0^\infty e^{-rt} \cdot \left(p' + (1-p')\cdot e^{-\lambda t} \right) \cdot y_t \, dt - \int_0^\infty e^{-rt} \cdot \left(p' + (1-p')\cdot e^{-\lambda t} \right) \cdot \hat{y}_t \, dt \\
=& \int_0^\infty e^{-rt} \cdot \left(p' + (1-p')\cdot e^{-\lambda t} \right) \cdot (y_t - \hat{y}_t) \, dt \\
=& \int_0^\infty e^{-rt} \cdot \left(\tilde{p}_0 + (1-\tilde{p}_0)\cdot e^{-\lambda t} \right) \cdot \frac{p' + (1-p')\cdot e^{-\lambda t}}{\tilde{p}_0 + (1-\tilde{p}_0)\cdot e^{-\lambda t}} \cdot (y_t - \hat{y}_t) \, dt \\
=& \int_0^T e^{-rt} \cdot \left(\tilde{p}_0 + (1-\tilde{p}_0)\cdot e^{-\lambda t} \right) \cdot \frac{p' + (1-p')\cdot e^{-\lambda t}}{\tilde{p}_0 + (1-\tilde{p}_0)\cdot e^{-\lambda t}} \cdot \underbrace{(y_t - \hat{y}_t)}_{\leq 0} \, dt \\
+& \int_T^\infty e^{-rt} \cdot \left(\tilde{p}_0 + (1-\tilde{p}_0)\cdot e^{-\lambda t} \right) \cdot \frac{p' + (1-p')\cdot e^{-\lambda t}}{\tilde{p}_0 + (1-\tilde{p}_0)\cdot e^{-\lambda t}} \cdot \underbrace{(y_t - \hat{y}_t)}_{\geq 0} \, dt \\
\overset{(i)}{\leq}& \int_0^T e^{-rt} \cdot \left(\tilde{p}_0 + (1-\tilde{p}_0)\cdot e^{-\lambda t} \right) \cdot \frac{p' + (1-p')\cdot e^{-\lambda T}}{\tilde{p}_0 + (1-\tilde{p}_0)\cdot e^{-\lambda T}} \cdot (y_t - \hat{y}_t) \, dt \\
+& \int_T^\infty e^{-rt} \cdot \left(\tilde{p}_0 + (1-\tilde{p}_0)\cdot e^{-\lambda t} \right) \cdot \frac{p' + (1-p')\cdot e^{-\lambda T}}{\tilde{p}_0 + (1-\tilde{p}_0)\cdot e^{-\lambda T}} \cdot (y_t - \hat{y}_t) \, dt \\
=& \frac{p' + (1-p')\cdot e^{-\lambda T}}{\tilde{p}_0 + (1-\tilde{p}_0)\cdot e^{-\lambda T}} \cdot \int_0^\infty e^{-rt} \cdot \left(\tilde{p}_0 + (1-\tilde{p}_0)\cdot e^{-\lambda t} \right) \cdot (y_t - \hat{y}_t) \, dt = 0.
\end{aligned}$$

Inequality (i) follows from the fact that function $\frac{p'+(1-p')\cdot e^{-\lambda t}}{\tilde{p}_0+(1-\tilde{p}_0)\cdot e^{-\lambda t}}$ is decreasing in time t and we are multiplying the negative (positive) cash flow outcomes by a smaller (larger) number by evaluating it at time T. We can also see that as long as $y_t \neq \hat{y}_t$ on a set of positive Lebesgue measure, then inequality (i) is strict. The argument is reversed for $p' > p$ since function $\frac{p'+(1-p')\cdot e^{-\lambda t}}{\tilde{p}_0+(1-\tilde{p}_0)\cdot e^{-\lambda t}}$ is increasing in this case. ☐

Lemma A.17 (Indifference Belief)**.** *Suppose types $\tilde{p} \in [p_1, p_2]$ offer the same contract $\alpha = \{\alpha_t\}_{t\geq 0} \in \cap_{p\in[p_1,p_2]}\Delta(p)$ and are perceived as $\Pi(\alpha)$ in equilibrium. Consider an alternative contract $\hat{\alpha} \overset{def}{=} \{\alpha_t \cdot \mathbb{1}\{t \geq T\}\}_{t\geq 0}$. Then the indifference belief $\Pi^I(\hat{\alpha}, \tilde{p}) \leq \Pi^I(\hat{\alpha}, p_2)$ for all $\tilde{p} < p_2$ with the inequality being strict whenever $\hat{\alpha} \neq \alpha$ on a set of positive Lebesgue measure.*

Proof. Consider contract $\hat{\alpha}$. Since $\hat{\alpha} < \alpha$, the indifference belief $\Pi^I(\hat{\alpha}, p_2)$ required by type p_2 to be indifferent is weakly higher that $\Pi(\alpha)$. Consider the baseline consumption process $C_t = \alpha_t \cdot A(\pi_t(\Pi(\alpha)))$

and the alternative consumption process $\hat{C}_t = \hat{\alpha}_t \cdot A\left(\pi_t\left(\Pi^I(\hat{\alpha}, p_2)\right)\right)$. The compensation process $C_t \geq \hat{C}_t$ for $t \leq T$ and $C_t \leq \hat{C}_t$ for $t \geq T$. This implies that compensation processes C and $\hat{C}$ satisfy Lemma A.16 and types $\tilde{p} < \hat{p}$ weakly prefer contract α and belief $\Pi(\alpha)$ over contract $\hat{\alpha}$ and belief $\Pi^I(\hat{\alpha}, p_2)$. This implies that the indifference belief $\Pi(\hat{\alpha}, \tilde{p}) \geq \Pi^I(\hat{\alpha}, p_2)$ for all $\tilde{p} < p_2$.

Whenever $\hat{\alpha}_t < \alpha_t$ on a set of positive Lebesgue measure, it follows that $\Pi^I(\hat{\alpha}, p_2) > \Pi(\alpha)$ as type p_2 will require a strictly higher belief following time T to forgo compensation during time $[0, T]$. This implies that $C \neq \hat{C}$ on a set of positive measure, and, following Lemma A.16, it implies that types $\tilde{p} < p_2$ strictly prefer contract α and belief $\Pi(\alpha)$ to contract $\hat{\alpha}$ and belief $\Pi^I(\hat{\alpha}, p_2)$. This, in turn, implies that type $\tilde{p} < p_2$ requires a strictly higher belief $\Pi^I(\hat{\alpha}, \tilde{p}) > \Pi^I(\hat{\alpha}, p_2)$ to be indifferent between offering contracts $\hat{\alpha}$ and α. ∎

Lemma A.18 (Unique Separating Equilibrium)**.** *Every Perfect Bayesian Equilibrium of the subgame that satisfies the D1 belief refinement must be separating in types.*

Proof. Suppose there exists a pooling equilibrium in which types p_1 and p_2 choose the same contract α. The weak convexity of $H(\tilde{p})$ implies that contract α^p is weakly optimal for all types $\tilde{p} \in [p_1, p_2]$. Pooling on contract α implies that $\Pi(\alpha) \in (p_1, p_2)$. Define $\hat{t} \overset{def}{=} \inf\{t : \ \alpha_t > 0\}$. Right continuity of α_t implies that $\alpha_t = 0$ for all $t < \hat{t}$ and there exists an $\varepsilon > 0$ such that $\alpha_t > 0$ for $t \in \left(\hat{t}, \hat{t} + \varepsilon\right)$.

Consider a deviation contract $\alpha^T = \{\alpha_t \cdot \mathbb{1}\{t \geq T\}\}_{t \geq 0}$ for $T \in (\hat{t}, \hat{t} + \varepsilon)$. We aim to show that there exists an interval $\left(\hat{t}, \hat{t} + \varepsilon^*\right) \subseteq \left(\hat{t}, \hat{t} + \varepsilon\right)$ such that contracts α^T for $T \in (t_1, t_2)$ are not offered in equilibrium.

(i) Ruling out types $[\underline{p}, p_1]$. Contract α^T is strictly suboptimal for types $\tilde{p} \leq p_1$ along the equilibrium path since by choosing contract α they get more revenue from clients and are perceived as at least type p_1. Hence contract α^T is not used in equilibrium by types $\tilde{p} < p_1$.

(ii) Ruling out types $[p_2, \overline{p}]$. There also exists a $\hat{\varepsilon} > 0$ such that contracts α^T for $T \in [\hat{t}, \hat{t} + \hat{\varepsilon}]$ are not offered in equilibrium by types $\tilde{p} > p_2$. If this was not the case, then then types $[p_1, p_2]$ would deviate to a contract α^T for T sufficiently close to $\hat{t}$, and thus sacrifice very little fees, but be perceived as type $\Pi(\alpha^T) \geq p_2$. These arguments imply that there exists a an interval of times $(\hat{t}, \hat{t} + \hat{\varepsilon})$ such that contracts α^T for $T \in (\hat{t}, \hat{t} + \hat{\varepsilon})$ are not used in equilibrium by types $[\underline{p}, p_1] \cup [p_2, \overline{p}]$.

(iii) Ruling out types $[p_1, p_2]$. Suppose the set of times $\mathbb{T}$ for which contracts α^T are offered by types $[p_1, p_2]$ is dense in $(\hat{t}, \hat{t} + \hat{\varepsilon})$. Optimality implies that all types $[p_1, p_2]$ must be indifferent in offering contracts α^T for $T \in \mathbb{T}$. This leads to a contradiction since $\alpha_t > 0$ for $t \in (\hat{t}, \hat{t} + \hat{\varepsilon})$ meaning that there does not exist a belief mapping $\Pi(\alpha)$ that preserves indifference for all types $[p_1, p_2]$. This implies that there exists an interval $(\hat{t}, \hat{t} + \varepsilon^*) \subseteq (\hat{t}, \hat{t} + \hat{\varepsilon})$ such that contract α^T is not offered in equilibrium.

Contracts α^T for $t \in (\hat{t}, \hat{t} + \varepsilon^*)$ are not offered in equilibrium. Lemma A.17 implies that the indifference belief $\Pi(\alpha^T, \tilde{p})$ is strictly decreasing in $\tilde{p}$ over $\tilde{p} \in [p_1, p_2]$. Moreover, since contract α^T is suboptimal for types $\tilde{p} < p_1$, it follows that $\Pi(\alpha^T, \tilde{p}) \geq \Pi(\alpha^T, p_1)$ for $\tilde{p} \in [\underline{p}, p_1]$. Consequently, the D1 equilibrium refinement implies that the belief $\Pi(\alpha^T)$ must put 0 weight on types below p_2, implying that $\Pi(\alpha^T) \geq p_2$. This implies that type p_2 can generate a strict improvement upon his equilibrium payoff by switching to contract α^T for T sufficiently close to $\hat{t}$ and being perceived as $\Pi(\alpha^T) \geq p_2 > \Pi(\alpha)$. this contradicts pooling by types $[p_1, p_2]$ on contract α. □

Definition 3 (Deferred Compensation Equilibrium)**.** *Consider a family of contracts $\{\alpha^T\}_{T\geq 0}$, defined as $\alpha_t^T \overset{def}{=} \mathbb{1}\{t \geq T\}$. A deferred compensation equilibrium is one in which the agents only offer contracts from the set $\{\alpha^T\}_{T\geq 0}$ along the equilibrium path.*

Lemma A.19. *There exists a unique deferred compensation equilibrium that satisfies the D1 belief refinement.*

Proof. First, we characterize the unique deferred compensation equilibrium that has to satisfy the D1 belief refinement. Then we show that the constructed equilibrium indeed satisfies it, i.e., immune from deviations to alternative contracts.

The candidate equilibrium is one in which the agent of initial skill $\tilde{p}$ chooses contract $\alpha^{T(\tilde{p})}$ in which he commits to forgo all of the wages for a period $T(\tilde{p})$ and then collects the full value $A(\pi_t(\tilde{p}))$ of his services for $t > T(\tilde{p})$. The expected value to the agent of ability $\tilde{p}$ from offering contract $\alpha^{T(\hat{p})}$ and being perceived as type $\hat{p}$ is

$$H(\tilde{p}, \hat{p}) = \int_{T(\hat{p})}^{\infty} e^{-rt} \cdot \left(\tilde{p} + (1-\tilde{p}) \cdot e^{-\lambda t}\right) \cdot A(\pi_t(\hat{p}))\, dt.$$

It follows from Lemma A.18 that every Perfect Bayesian Equilibrium must be separating in types. It must be the case that $T(\tilde{p})$ is continuous and increasing in $\tilde{p}$ since either non-monotonicity, or discontinuities would contradict the incentive compatibility of the separating equilibrium. The equilibrium is separating as long as $\frac{\partial H}{\partial \hat{p}} H(\tilde{p}, \hat{p})\big|_{\hat{p}=p} = 0$. The strict optimality of $T(\tilde{p})$ then requires that

$$\frac{\partial}{\partial \hat{p}} H(\tilde{p}, \hat{p})\Big|_{\hat{p}=p} = \int_{T(\tilde{p})}^{\infty} e^{-rt} \left(\tilde{p} + (1-\tilde{p})e^{-\lambda t}\right) A(\pi_t(\tilde{p}))\pi_t'(\tilde{p})\, dt - \left(p + (1-p)e^{-\lambda T(\tilde{p})}\right) A\left(\pi_{T(\tilde{p})}(p)\right) \cdot T'(p) = 0.$$

Combined with the boundary condition $T(\underline{p}) = 0$, the above differential equation for $T(\tilde{p})$ uniquely pins down times $T(\tilde{p})$ for $p \geq \underline{p}$. This implies that there is at most one deferred compensation equilibrium that satisfies the D1 refinement.

Now, we show that the deferred compensation equilibrium constructed above indeed satisfies the D1 belief refinement. Consider an off-path contract $\alpha = (\alpha_t)_{t\geq 0}$ and refer to $\Pi^I(\alpha, \tilde{p})$ as the indifference belief of type $\tilde{p}$ that makes them indifferent between contract α and the equilibrium contract $\alpha^{T(\tilde{p})}$ in definition 3.

The indifference belief $\Pi^I(\alpha, \tilde{p})$ satisfies

$$\int_0^\infty e^{-rt} \cdot \left(\tilde{p} + (1-\tilde{p}) \cdot e^{-\lambda t}\right) \cdot \alpha_t \cdot A\left[\pi_t\left(\Pi^I(\alpha, \tilde{p})\right)\right] \, dt = \int_{T(\tilde{p})}^\infty e^{-rt} \cdot \left(\tilde{p} + (1-\tilde{p}) \cdot e^{-\lambda t}\right) \cdot A(\pi_t(\tilde{p})) \, dt.$$

Given the strict convexity of the agent's value function $H(p)$ in the unique deferred compensation equilibrium satisfying the D1 refinement, it follows that $\arg\min_p \Pi^I(\alpha, \tilde{p}) = p^*$ is a singleton. Contract α leads to the violation of the D1 belief refinement of the deferred compensation equilibrium if there exists a type $\hat{p}$ that would benefit from offering contract α and be perceived as type p^*, rather than obtain his equilibrium payoff, i.e.,

$$\int_0^\infty e^{-rt} \cdot \left(\hat{p} + (1-\hat{p}) \cdot e^{-\lambda t}\right) \cdot \alpha_t \cdot A\Big(\pi_t(p^*)\Big) \, dt > \int_{T(\hat{p})}^\infty e^{-rt} \cdot \left(\hat{p} + (1-\hat{p}) \cdot e^{-\lambda t}\right) \cdot A(\pi_t(\hat{p})) \, dt.$$

Given the monotonicity of the agent's payoff in client beliefs, the above inequality can be restated as

$$\arg\min_p \Pi^I(\alpha, \tilde{p}) = p^* > \Pi^I(\alpha, p^*) = \min_p \Pi^I(\alpha, \tilde{p}).$$

Consider a deferred compensation contract $\hat{\alpha}$, defined by $\hat{\alpha}_t = \mathbb{1}\left\{t \geq \hat{T}\right\}$ such that type p^* is exactly indifferent between $\hat{\alpha}$ and α if he is perceived as $\Pi^I(\alpha, p^*)$. By construction, contract $\hat{\alpha}$ single crosses contract α from below. Following Lemma A.16, it follows that $\Pi^I(\hat{\alpha}, \tilde{p}) > \Pi^I(\alpha, \tilde{p})$ for $p < p^*$ and $\Pi^I(\hat{\alpha}, \tilde{p}) < \Pi^I(\alpha, \tilde{p})$ for $p > p^*$. This implies

$$\arg\min_p \Pi^I(\hat{\alpha}, \tilde{p}) \geq \arg\min_p \Pi^I(\alpha, \tilde{p}) = p^* > \Pi^I(\alpha, p^*) = \min_p \Pi^I(\alpha, \tilde{p}) = \Pi^I(\hat{\alpha}, p^*) \geq \min_p \Pi^I(\hat{\alpha}, \tilde{p}).$$

Consequently, if contract α violates the D1 equilibrium refinement, then the deferred compensation contract $\hat{\alpha}$ must also violate it. This, however, cannot be the case as the deferred compensation contract $\hat{\alpha}$ is a feasible deferred compensation contract that is either offered and chosen in equilibrium or is off-path, meaning that it is dominated by the longest-deferred compensation contract, offered by the highest type $\bar{p}$. Consequently, there is a unique equilibrium. □

Lemma A.20. *The deferred compensation equilibrium Pareto dominates any other separating equilibrium for the agents.*

Proof. Consider an arbitrary separating equilibrium value function $\hat{H}(\tilde{p}, \hat{p})$. The first order condition is

$$\frac{\partial}{\partial \hat{p}} \hat{H}(\tilde{p}, \hat{p}-) \geq 0 \geq \frac{\partial}{\partial \hat{p}} \hat{H}(\tilde{p}, \hat{p}+).$$

For each $\hat{p}$ have

$$\frac{\partial}{\partial p} \hat{H}(\tilde{p}, \hat{p}) = \int_0^\infty e^{-rt} \cdot \left(1 - e^{-\lambda t}\right) \cdot \alpha_t(\hat{p}) \cdot A(\pi_t(\hat{p})) \, dt.$$

In a separating equilibrium it must be the case that $\alpha_t(\underline{p}) \equiv 1$, implying that $H(\underline{p}, \underline{p}) = \hat{H}(\underline{p}, \underline{p})$. Suppose there exists $\tilde{p} > \underline{p}$ such that $H(\tilde{p}, \tilde{p}) = \hat{H}(\tilde{p}, \tilde{p})$. Then we will show that $\frac{d}{dp} H(\tilde{p}+, \tilde{p}+)| \geq \frac{d}{dp} \hat{H}(\tilde{p}+, \tilde{p}+)$ with the inequality being strict whenever $\hat{\alpha}_t(\tilde{p}) \neq \mathbb{1}\{t \geq T(\tilde{p})\}$ for every $t \geq 0$. Equality $H(\tilde{p}, \tilde{p}) = \hat{H}(\tilde{p}, \tilde{p})$ can be written as i.e.,

$$\begin{aligned} H(\tilde{p}, \tilde{p}) &= \int_{T(\tilde{p})}^{\infty} e^{-rt} \left(p + (1-p) e^{-\lambda t} \right) \cdot A(\pi_t(\tilde{p}))\, dt \\ &= \int_0^{\infty} e^{-rt} \left(p + (1-p) e^{-\lambda t} \right) \cdot \alpha_t(\tilde{p}) \cdot A(\pi_t(\tilde{p}))\, dt = \hat{H}(\tilde{p}, \tilde{p}). \end{aligned} \tag{A.71}$$

Following Lemma A.16 it follows that whenever $\alpha_t(p) \neq \mathbb{1}\{t \geq T(\tilde{p})\}$ for all $t \geq 0$ then for all $\hat{p} < p$

$$\int_{T(\tilde{p})}^{\infty} e^{-rt} \left(\hat{p} + (1-\hat{p}) e^{-\lambda t} \right) \cdot A(\pi_t(\tilde{p}))\, dt < \int_0^{\infty} e^{-rt} \left(\hat{p} + (1-\hat{p}) e^{-\lambda t} \right) \cdot \alpha_t(\tilde{p}) \cdot A(\pi_t(\tilde{p}))\, dt. \tag{A.72}$$

The right derivative of the deferred equilibrium value function $H(\tilde{p}, p)$ at $\tilde{p} = p$ is

$$\frac{d}{dp} H(p+, p+) = \frac{\partial}{\partial p} H(p+, p) + \underbrace{\frac{\partial}{\partial \hat{p}} H(\tilde{p}, p+)}_{=0} = \frac{\partial}{\partial p} H(p+, p) = \int_{T(\tilde{p})}^{\infty} e^{-rt} \cdot \left(1 - e^{-\lambda t} \right) \cdot A(\pi_t(\tilde{p}))\, dt.$$

The right derivative at $\tilde{p}$ of the candidate equilibrium is

$$\frac{d}{dp} \hat{H}(\tilde{p}, \tilde{p}) = \frac{\partial}{\partial p} \hat{H}(\tilde{p}, \tilde{p}) + \underbrace{\frac{\partial}{\partial \hat{p}} \hat{H}(\tilde{p}, p+)}_{\leq 0} \leq \frac{\partial}{\partial p} \hat{H}(\tilde{p}, \tilde{p}) = \int_0^{\infty} e^{-rt} \cdot \left(1 - e^{-\lambda t} \right) \cdot \alpha_t(\tilde{p}) \cdot A(\pi_t(\tilde{p}))\, dt.$$

Then, the comparison becomes

$$\frac{d}{dp} \hat{H}(\tilde{p}, \tilde{p}) \leq \int_0^{\infty} e^{-rt} \left(1 - e^{-\lambda t} \right) \cdot \alpha_t(\tilde{p}) \cdot A(\pi_t(\tilde{p}))\, dt \overset{(i)}{<} \int_{T(\tilde{p})}^{\infty} e^{-rt} \left(1 - e^{-\lambda t} \right) \cdot A(\pi_t(\tilde{p}))\, dt = \frac{d}{dp} \hat{H}(\tilde{p}, \tilde{p}),$$

where inequality (i) follows from combining (A.71) and (A.72). □

Lemma A.21 (Unique D1 Equilibrium)**.** *The deferred compensation contract is the unique Perfect Bayesian Equilibrium satisfying the D1 refinement.*

Proof. Consider a candidate equilibrium value function $\hat{H}(\tilde{p}, \hat{p})$. For the family of contracts $\Delta(\tilde{p})$ offered in equilibrium by type $\tilde{p}$ define times $\hat{T}(\tilde{p})$ such that type $\tilde{p}$ is exactly indifferent between offering contract $\alpha(\tilde{p}) \in \Delta(\tilde{p})$ and the deferred compensation contract $\alpha^{\hat{T}(\tilde{p})}$. Formally, $\hat{T}(\tilde{p})$ solves

$$\int_0^{\infty} e^{-rt} \cdot \left(\tilde{p} + (1-\tilde{p}) e^{-\lambda t} \right) \cdot \alpha_t(\tilde{p}) \cdot A(\pi_t(\tilde{p}))\, dt = \int_{\hat{T}(\tilde{p})}^{\infty} e^{-rt} \cdot \left(\tilde{p} + (1-\tilde{p}) e^{-\lambda t} \right) \cdot A(\pi_t(\tilde{p}))\, dt.$$

From Lemma A.16 it follows that $\Pi^I \left(\alpha^{\hat{T}(\tilde{p})}, p \right) \geq \Pi^I(\alpha(p), \tilde{p}) = p$ for every $p \leq \tilde{p}$. This, in turn, implies that $\Pi\left(\alpha^{\hat{T}(p)}\right) \geq p$ for every $p \in [\underline{p}, 1]$. If it was ever the case that $\Pi\left(\alpha^{\hat{T}(p)}\right) > p$, it would contradict the optimality of choosing $\alpha(p)$ over $\alpha^{\hat{T}(p)}$ for type $\tilde{p}$. Hence it must be the case that $\Pi\left(\alpha^{\hat{T}(\tilde{p})}\right) = \tilde{p}$ for every $\tilde{p} \in [\underline{p}, 1]$. Consequently, if contracts $\{\Delta(\tilde{p})\}_{p\in[\underline{p},1]}$ constitute an equilibrium with beliefs $\Pi(\cdot)$, then it must be the case that $\alpha^{\hat{T}(\tilde{p})} \in \Delta(\tilde{p})$ for each $\tilde{p} \in [\underline{p}, 1]$. Following Lemma A.19 the deferred compensation equilibrium is unique, implying that the value function $\hat{H}(\tilde{p}, \tilde{p})$ of our starting equilibrium has to coincide with the value function of $H(\tilde{p}, p)$ of the deferred compensation equilibrium. Consequently, $\Delta(\tilde{p}) = \{\alpha^{T(\tilde{p})}\}$ for each $p \in [\underline{p}, 1]$. □

Proof of Observation 4 (intermediary equilibrium with independent signaling)

To simplify the derivations below, we assume, without loss, that $L = 0$. Similar to our argument in the first paragraph of the proof of Observation 3, we assume, without loss, that $L = 0$. To account for $L > 0$ we need only consider an initial truncation of types. Moreover, it follows from Lemma 3 that churned agents prefer to stay in the industry and, consequently, in the subgame equilibria that take place after the start of the churning period, the positive outside option $L > 0$ of the agent stops being binding. To avoid the incremental notation associated with $L > 0$, we assume it is equal to 0 in the proof below, but it can be directly extended, similar to how we do so in Section (A.1) when generalizing the monotonicity requirement for equilibrium uniqueness.

Denote the expected value to the agent of ability $\tilde{p}$ from choosing deferred compensation contract $\alpha^{T(\tilde{p})}$. The expected on-path value to the agent of type $\tilde{p}$ at the start of the deferred compensation equilibrium is denoted by

$$H(\tilde{p}) \stackrel{def}{=} H(\tilde{p}, \tilde{p}) = \int_{\hat{T}(\tilde{p})}^{\infty} e^{-rt} \cdot \left(\tilde{p} + (1-\tilde{p})e^{-\lambda t}\right) \cdot A(\pi_t(\tilde{p}))\, dt.$$

It is convenient to denote $u_\theta(k) \stackrel{def}{=} U(\theta, k)$. By Law of Iterated Expectation, we can write $U(\tilde{p}, k) = \tilde{p} \cdot u_1(k) + (1-\tilde{p}) \cdot u_0(k)$.

Lemma A.22. *Suppose the set of initial types is distributed continuously over support* $[l, 1]$. *The deferred compensation equilibrium satisfies*

$$H'(\tilde{p}) = e^{-rT(\tilde{p})} \cdot u_1\left(\pi_{T(\tilde{p})}(\tilde{p})\right) - e^{-(r+\lambda)T(\tilde{p})} \cdot u_0\left(\pi_{T(\tilde{p})}(\tilde{p})\right) > 0 \tag{A.73}$$

and

$$H''(\tilde{p}) = \left[u_1'\left(\pi_{T(\tilde{p})}(\tilde{p})\right) - u_0'\left(\pi_{T(\tilde{p})}(\tilde{p})\right)\right] \cdot \frac{e^{-(r+2\lambda)T(\tilde{p})}}{\left(\tilde{p} + (1-\tilde{p}) \cdot e^{-\lambda T(\tilde{p})}\right)^3} \tag{A.74}$$

Proof. Define belief process $k = \{k_t\}_{t\in[0,T(1)]}$ via the differential equation

$$\dot{k}_t = \lambda k_t(1-k_t) + \frac{A(k_t)}{\partial_2 U(k_t,k_t)}, \qquad k_0 = l. \tag{A.75}$$

It follows that $k_{T(\tilde{p})} = \pi_{T(\tilde{p})}(\tilde{p})$ for each $p \in [\underline{p}, 1]$. The expected value to the agent of ability $\tilde{p}$ in the deferred compensation equilibrium can then be written as

$$H(\tilde{p}) = \tilde{p}\cdot e^{-rT(\tilde{p})}\cdot u_1(k_{T(\tilde{p})}) + (1-\tilde{p})\cdot e^{-(r+\lambda)T(\tilde{p})}\cdot u_0(k_{T(\tilde{p})}). \tag{A.76}$$

Use Envelope theorem to differentiate (A.76) to obtain

$$H'(\tilde{p}) = e^{-rT(\tilde{p})}\cdot u_1\left(k_{T(\tilde{p})}\right) - e^{-(r+\lambda)T(\tilde{p})}\cdot u_0\left(k_{T(\tilde{p})}\right), \tag{A.77}$$

which implies (A.73) since $k_{T(\tilde{p})} = \pi_{T(\tilde{p})}(\tilde{p})$.

Second derivative. By construction of process k in (A.75) it follows that $\pi_{T(\tilde{p})}(\tilde{p}) \equiv k_{T(\tilde{p})}$ for every $\tilde{p}$. Differentiating this expression with respect to $\tilde{p}$ obtain an identity

$$\begin{aligned}
\pi'_{T(\tilde{p})}(\tilde{p}) + \dot{\pi}_{T(\tilde{p})}(\tilde{p})\cdot T'(\tilde{p}) &= \dot{k}_{T(\tilde{p})}\cdot T'(\tilde{p})\\
\pi'_{T(\tilde{p})}(\tilde{p}) + \lambda\pi_{T(\tilde{p})}(\tilde{p})\Big(1-\pi_{T(\tilde{p})}(\tilde{p})\Big)T'(\tilde{p}) &= \left[\lambda k_{T(\tilde{p})}\left(1-k_{T(\tilde{p})}\right) + \frac{A\left(k_{T(\tilde{p})}\right)}{\partial_2 U\left(k_{T(\tilde{p})},k_{T(\tilde{p})}\right)}\right]T'(\tilde{p})\\
\frac{e^{-\lambda T(\tilde{p})}}{(\tilde{p}+(1-\tilde{p})\cdot e^{-\lambda T(\tilde{p})})^2} &= \frac{A(k_{T(\tilde{p})}}{\partial_2 U\left(k_{T(\tilde{p})},k_{T(\tilde{p})}\right)}\cdot T'(\tilde{p})\\
\frac{k_{T(\tilde{p})}\cdot(1-k_{T(\tilde{p})})}{p\cdot(1-\tilde{p})} &= \frac{\alpha_t\cdot A(k_{T(\tilde{p})}}{\partial_2 U\left(k_{T(\tilde{p})},k_{T(\tilde{p})}\right)}\cdot T'(\tilde{p}).\\
\frac{k_{T(\tilde{p})}\cdot(1-k_{T(\tilde{p})})}{\tilde{p}\cdot(1-\tilde{p})}\cdot\frac{\partial_2 U\left(k_{T(\tilde{p})},k_{T(\tilde{p})}\right)}{\alpha_t\cdot A\left(k_{T(\tilde{p})}\right)} &= T'(\tilde{p}).
\end{aligned} \tag{A.78}$$

Differentiate (A.77) to obtain the second derivative of $H(\tilde{p})$ as

$$\begin{aligned}
H''(p) =& - r\cdot e^{-rT(\tilde{p})}\cdot u_1\left(k_{T(\tilde{p})}\right)\cdot T'(\tilde{p}) + e^{-rT(\tilde{p})}\cdot u_1'\left(k_{T(\tilde{p})}\right)\cdot \dot{k}_{T(\tilde{p})}\cdot T'(\tilde{p})\\
&+ (r+\lambda)\cdot e^{-(r+\lambda)T(\tilde{p})}\cdot u_0\left(k_{T(\tilde{p})}\right)\cdot T'(\tilde{p}) - e^{-(r+\lambda)T(\tilde{p})}\cdot u_0'\left(k_{T(\tilde{p})}\right)\cdot \dot{k}_{T(\tilde{p})}\cdot T'(\tilde{p})\\
\overset{t=T(\tilde{p})}{=}& e^{-rt}\cdot\left[-ru_1(k_t) + u_1'(k_t)\cdot\left(\lambda k_t(1-k_t) + \frac{A(k_t)}{\partial_2 U(k_t,k_t)}\right)\right]\cdot T'(\tilde{p})\\
&- e^{-(r+\lambda)t}\cdot\left[-(r+\lambda)u_0(k_t) + u_0'(k_t)\cdot\left(\lambda k_t(1-k_t) + \frac{A(k_t)}{\partial_2 U(k_t,k_t)}\right)\right]\cdot T'(\tilde{p})\\
=& e^{-rt}\left[-A(k_t) + u_1'(k_t)\cdot\frac{A(k_t)}{\partial_2 U(k_t,k_t)}\right]\cdot T'(\tilde{p})
\end{aligned} \tag{A.79}$$

$$
\begin{aligned}
&- e^{-(r+\lambda)t}\left[-A(k_t)+u_0'(k_t)\cdot\frac{A(k_t)}{\partial_2 U(k_t,k_t)}\right]\cdot T'(\tilde p)\\
=&e^{-rt}\left[-1+\frac{u_1'(k_t)}{\partial_2 U(k_t,k_t)}\right]\cdot A(k_t)\cdot T'(\tilde p)-e^{-(r+\lambda)t}\left[-1+\frac{u_0'(k_t)}{\partial_2 U(k_t,k_t)}\right]\cdot A(k_t)\cdot T'(\tilde p)\\
=&\left[e^{-rt}\cdot\left(\frac{u_1'(k_t)}{\partial_2 U(k_t,k_t)}-1\right)+e^{-(r+\lambda)t}\cdot\left(1-\frac{u_0'(k_t)}{\partial_2 U(k_t,k_t)}\right)\right]\cdot A(k_t)\cdot T'(\tilde p)\\
=&\left[e^{-rt}\cdot\left(\frac{u_1'(k_t)}{\partial_2 U(k_t,k_t)}-1\right)+e^{-(r+\lambda)t}\cdot\left(1-\frac{u_0'(k_t)}{\partial_2 U(k_t,k_t)}\right)\right]\cdot\frac{k_t(1-k_t)}{p(1-p)}\cdot\partial_2 U(k_t,k_t)\\
=&\left[e^{-rt}\cdot(1-k_t)\cdot(u_1'(k_t)-u_0'(k_t))+e^{-(r+\lambda)t}\cdot k_t\cdot(u_1'(k_t)-u_0'(k_t))\right]\cdot\frac{k_t(1-k_t)}{p(1-p)}\\
=&(u_1'(k_t)-u_0'(k_t))\cdot\left[e^{-rt}\cdot(1-k_t)+e^{-(r+\lambda)t}\cdot k_t\right]\cdot\frac{k_t(1-k_t)}{p(1-p)}.
\end{aligned}
$$

Equality (A.79) above follows from (18) evaluated for $u_\theta(k)$ at $\tilde p=0$ and $\tilde p=1$. Note that $k_t=\pi_t(\tilde p)$ at $t=T(\tilde p)$ implies that

$$
\begin{aligned}
H''(p)&=\left[u_1'\Big(\pi_{T(\tilde p)}(\tilde p)\Big)-u_0'\Big(\pi_{T(\tilde p)}(\tilde p)\Big)\right]\cdot\frac{e^{-(r+\lambda)T(\tilde p)}}{\tilde p+(1-\tilde p)\cdot e^{-\lambda T(\tilde p)}}\cdot\frac{e^{-\lambda T(\tilde p)}}{\big(p+(1-p)e^{-\lambda T(\tilde p)}\big)^2}\\
&=\left[u_1'\Big(\pi_{T(\tilde p)}(\tilde p)\Big)-u_0'\Big(\pi_{T(\tilde p)}(\tilde p)\Big)\right]\cdot\frac{e^{-(r+2\lambda)T(\tilde p)}}{\big(\tilde p+(1-\tilde p)\cdot e^{-\lambda T(\tilde p)}\big)^3}.
\end{aligned}
$$

At $p=l$ we have $T(l)=0$ and $k_{T(l)}=k_0=l$ implying that $H''(l)=u_1'(l)-u_0'(l)$. □

Lemma A.23 (Convexity upper bound)**.** *The second derivative $H''(\tilde p)$ is bounded above by*

$$
H''(\tilde p)\le u_1'(\tilde p)-u_0'(\tilde p). \tag{A.80}
$$

Proof. We can express

$$
\begin{aligned}
u_1'(\pi_t(\tilde p))-u_0'(\pi_t(\tilde p))&=\int_0^\infty e^{-rs}\left(1-e^{-\lambda s}\right)\cdot A'\Big(\pi_s(\pi_t(\tilde p))\Big)\cdot\frac{e^{-\lambda s}}{\big(\pi_t(\tilde p)+(1-\pi_t(\tilde p))\cdot e^{-\lambda s}\big)^2}\,ds\\
=&\int_0^\infty e^{-rs}\left(1-e^{-\lambda s}\right)\cdot A'\Big(\pi_{T+s}(\tilde p)\Big)\cdot\frac{e^{-\lambda s}\cdot\big(\tilde p+(1-\tilde p)e^{-\lambda T}\big)^2}{\big(\tilde p+(1-\tilde p)\cdot e^{-\lambda(T+s)}\big)^2}\,ds\\
=&\Big(\tilde p+(1-\tilde p)e^{-\lambda T}\Big)^2\cdot\int_T^\infty e^{-r(s-T)}\left(1-e^{-\lambda(s-T)}\right)\cdot A'\Big(\pi_s(\tilde p)\Big)\cdot\frac{e^{-\lambda(s-T)}}{\big(\tilde p+(1-\tilde p)\cdot e^{-\lambda s}\big)^2}\,ds
\end{aligned}
$$

Substituting the above expression in (A.74) obtain

$$
H''(p)=\left[u_1'\Big(\pi_{T(\tilde p)}(\tilde p)\Big)-u_0'\Big(\pi_{T(\tilde p)}(\tilde p)\Big)\right]\cdot\frac{e^{-(r+2\lambda)T(\tilde p)}}{\big(\tilde p+(1-\tilde p)\cdot e^{-\lambda T(\tilde p)}\big)^3}
$$

$$= \frac{e^{-\lambda T(\tilde{p})}}{\tilde{p} + (1-\tilde{p}) \cdot e^{-\lambda T(\tilde{p})}} \int_{T(\tilde{p})}^{\infty} e^{-rs} \cdot \left(1 - e^{-\lambda(s - T(\tilde{p}))}\right) \cdot A'\Big(\pi_s(\tilde{p})\Big) \cdot \frac{e^{-\lambda s}}{\left(\tilde{p} + (1-\tilde{p}) \cdot e^{-\lambda s}\right)^2} \, ds$$
$$\leq \frac{e^{-\lambda T(\tilde{p})}}{\tilde{p} + (1-\tilde{p}) \cdot e^{-\lambda T(\tilde{p})}} \int_{T(\tilde{p})}^{\infty} e^{-rs} \cdot \left(1 - e^{-\lambda s}\right) \cdot A'\Big(\pi_s(\tilde{p})\Big) \cdot \frac{e^{-\lambda s}}{\left(\tilde{p} + (1-\tilde{p}) \cdot e^{-\lambda s}\right)^2} \, ds$$
$$\leq \int_{0}^{\infty} e^{-rs} \cdot \left(1 - e^{-\lambda s}\right) \cdot A'\Big(\pi_s(\tilde{p})\Big) \cdot \frac{e^{-\lambda s}}{\left(\tilde{p} + (1-\tilde{p}) \cdot e^{-\lambda s}\right)^2} \, ds = u_1'(\tilde{p}) - u_0'(\tilde{p}).$$

□

Deferred Compensation Equilibrium with a Finitely Lived Client

Suppose each client lives for an exponential time with parameter $\chi > 0$. Time χ translates into an increased discount rate $r + \chi$ and we can express the expected value to the agent of true ability $\theta \in \{0, 1\}$ from being perceived as type p by the current client as

$$u_{\chi,\theta}(p) \stackrel{def}{=} \int_0^{\infty} e^{-(r+\chi)t - \lambda(1-\theta)t} \cdot A(\pi_t(\tilde{p})) \, dt. \tag{A.81}$$

The results from Observation 4 continue to hold. Suppose the set of ex-ante types is restricted to a support $[l, 1]$. Define by $\alpha^{T_\chi(\tilde{p},l)}$ the deferred compensation contract chosen by the agent of type $\tilde{p}$ in equilibrium. We can define $H_\chi(\tilde{p}, l)$ as the on-path expected value to the agent of type p from choosing contract $\alpha^{T_\chi(\tilde{p},l)}$ and signaling to a single client. It is given by

$$H_\chi(\tilde{p}, l) \stackrel{def}{=} \tilde{p} \cdot e^{-(r+\chi) \cdot T_\chi(p,l)} \cdot u_{\chi,1}\Big(\pi_{T_\chi(\tilde{p},l)}(\tilde{p})\Big) + (1-\tilde{p}) \cdot e^{-(r+\chi+\lambda) \cdot T_\chi(\tilde{p},l)} \cdot u_{\chi,0}\Big(\pi_{T_\chi(\tilde{p},l)}(\tilde{p})\Big).$$

It follows from Lemma A.22 that

$$\frac{\partial}{\partial \tilde{p}} H_\chi(\tilde{p}, l)\Big|_{\tilde{p}=l} = u_{\chi,1}(l) - u_{\chi,0}(l), \qquad \frac{\partial^2}{\partial \tilde{p}^2} H_\chi(\tilde{p}, l) \leq u_{\chi,1}'(\tilde{p}) - u_{\chi,0}'(\tilde{p}). \tag{A.82}$$

First and Second Derivative Limits as $\chi \to \infty$

Denote by ξ_k the survival time of clients 1 through k, implying that ξ_k is equal to the sum of exponential random variables, each with parameter χ. This implies that ξ_k follows Erlang distribution with parameters χ and k. For $k = 0$ set $\xi_0 \equiv 0$. The expected value to the agent at $t = 0$ from contracting directly with a sequence of clients, writing a new deferred compensation contract when clients turnover, is given by the sum of expected values from contracting with the infinite sequence of clients

$$H_\chi^*(\tilde{p}, l) = \sum_{k=0}^{\infty} \mathrm{E}\left[e^{-r\xi_k} \cdot \left(\tilde{p} + (1-\tilde{p}) \cdot e^{-\lambda \xi_k}\right) \cdot H_\chi\Big(\pi_{\xi_k}(\tilde{p}), \pi_{\xi_k}(l)\Big)\right]. \tag{A.83}$$

Lemma A.24 (Signaling to Finite Clients Value Function Properties)**.** *For any χ it follows that*

$$\frac{\partial}{\partial \tilde{p}} H^*_\chi(\tilde{p}, l)\Big|_{p=l} = u_1(l) - u_0(l). \tag{A.84}$$

Moreover, $\lim_{\chi\to\infty} \frac{\partial^2}{\partial \tilde{p}^2} H^*_\chi(\tilde{p}, l) = 0$ *uniformly in p and l.*

Proof. The first derivative of (A.83) with respect to $\tilde{p}$ is

$$\begin{aligned} \frac{\partial}{\partial \tilde{p}} H^*_\chi(\tilde{p}, l) = &\sum_{k=0}^{\infty} \mathrm{E}\left[e^{-r\xi_k} \cdot \left(1 - e^{-\lambda\xi_k}\right) \cdot H_\chi\Big(\pi_{\xi_k}(\tilde{p}), \pi_{\xi_k}(l)\Big)\right] \\ &+ \sum_{k=0}^{\infty} \mathrm{E}\left[e^{-r\xi_k} \cdot \left(\tilde{p} + (1-\tilde{p}) \cdot e^{-\lambda\xi_k}\right) \cdot \frac{\partial}{\partial \tilde{p}} H_\chi\Big(\pi_{\xi_k}(\tilde{p}), \pi_{\xi_k}(l)\Big)\right]. \end{aligned} \tag{A.85}$$

At $\tilde{p} = l$ we have $H_\chi(l, l) = l \cdot u_{\chi,1}(l) + (1-l) \cdot u_{\chi,0}(l)$. Combining this with (A.82) we can express (A.85) at $\tilde{p} = l$ as

$$\begin{aligned} \frac{\partial}{\partial \tilde{p}} H^*_\chi(\tilde{p}, l)\Big|_{\tilde{p}=l} = &\sum_{k=0}^{\infty} \mathrm{E}\left[e^{-r\xi_k} \cdot \left(1 - e^{-\lambda\xi_k}\right) \cdot \left[\pi_{\xi_k}(l) \cdot u_{\chi,1}\Big(\pi_{\xi_k}(l)\Big) + (1 - \pi_{\xi_k}(l)) \cdot u_{\chi,0}\Big(\pi_{\xi_k}(l)\Big)\right]\Big)\right] \\ &+ \sum_{k=0}^{\infty} \mathrm{E}\left[e^{-r\xi_k} \cdot \left(l + (1-l) \cdot e^{-\lambda\xi_k}\right) \cdot \left[u_{\chi,1}\Big(\pi_{\xi_k}(l)\Big) - u_{\chi,0}\Big(\pi_{\xi_k}(l)\Big)\right] \cdot \pi'_{\xi_k}(l)\right] \\ = &\sum_{k=0}^{\infty} \mathrm{E}\left[e^{-r\xi_k} \cdot u_{\xi,1}(\pi_{\xi_k}(l)) - e^{-(r+\lambda)\xi_k} \cdot u_{\xi,0}(\pi_{\xi_k}(l))\right] \\ = &\sum_{k=0}^{\infty} \mathrm{E}\left[e^{-r\xi_k} \cdot \int_0^{\xi_{k+1}-\xi_k} e^{-rs} \cdot A\Big(\pi_s((\pi_{\xi_k}(l))\Big)\, ds\right] \\ &- \sum_{k=0}^{\infty} \mathrm{E}\left[e^{-(r+\lambda)\xi_k} \cdot \int_0^{\xi_{k+1}-\xi_k} e^{-(r+\lambda)s} \cdot A\Big(\pi_s((\pi_{\xi_k}(l))\Big)\, ds\right] \\ = &\sum_{k=0}^{\infty} \mathrm{E}\left[e^{-r\xi_k} \cdot \int_0^{\xi_{k+1}-\xi_k} e^{-rs} \cdot A\Big(\pi_{\xi_k+s}(l)\Big)\, ds\right] \\ &- \sum_{k=0}^{\infty} \mathrm{E}\left[e^{-(r+\lambda)\xi_k} \cdot \int_0^{\xi_{k+1}-\xi_k} e^{-(r+\lambda)s} \cdot A\Big(\pi_{\xi_k+s}(l)\Big)\, ds\right] \\ = &\sum_{k=0}^{\infty} \mathrm{E}\left[\int_{\xi_k}^{\xi_{k+1}} e^{-rs} \cdot A\Big(\pi_s(l)\Big)\, ds - e^{-(r+\lambda)\xi_k} \int_{\xi_k}^{\xi_{k+1}} e^{-(r+\lambda)s} \cdot A\Big(\pi_s(l)\Big)\, ds\right] \\ = &\,\mathrm{E}\left[\int_0^{\infty} e^{-rs} \cdot A\Big(\pi_s(l)\Big)\, ds - \int_0^{\infty} e^{-(r+\lambda)s} \cdot A\Big(\pi_s(l, s)\Big)\, ds\right] = u_1(l) - u_0(l), \end{aligned}$$

which proves (A.84). The second derivative of (A.83) with respect to $\tilde{p}$ is

$$\begin{aligned}
\frac{\partial^2}{\partial \tilde{p}^2} H_\chi^*(\tilde{p}, l) &= 2 \cdot \sum_{k=0}^{\infty} \mathrm{E}\left[e^{-r\xi_k} \cdot \left(1 - e^{-\lambda \xi_k}\right) \cdot \frac{\partial}{\partial \tilde{p}} H_\chi\left(\pi_{\xi_k}(\tilde{p}), \pi_{\xi_k}(l)\right)\right] \\
&\quad + \sum_{k=0}^{\infty} \mathrm{E}\left[e^{-r\xi_k} \cdot \left(\tilde{p} + (1-\tilde{p}) \cdot e^{-\lambda\xi_k}\right) \cdot \frac{\partial^2}{\partial \tilde{p}^2} H_\chi\left(\pi_{\xi_k}(\tilde{p}), \pi_{\xi_k}(l)\right)\right] \\
&= 2 \cdot \sum_{k=0}^{\infty} \mathrm{E}\left[e^{-r\xi_k} \cdot \left(1 - e^{-\lambda \xi_k}\right) \cdot \partial_1 H_\chi\left(\pi_{\xi_k}(\tilde{p}), \pi_{\xi_k}(l)\right) \cdot \pi'_{\xi_k}(\tilde{p})\right] \\
&\quad + \sum_{k=0}^{\infty} \mathrm{E}\left[e^{-r\xi_k} \cdot \left(\tilde{p} + (1-\tilde{p}) \cdot e^{-\lambda\xi_k}\right) \cdot \partial_1^2 H_\chi\left(\pi_{\xi_k}(\tilde{p}), \pi_{\xi_k}(l)\right) \cdot \pi'_{\xi_k}(\tilde{p})^2\right] \\
&\quad + \sum_{k=0}^{\infty} \mathrm{E}\left[e^{-r\xi_k} \cdot \left(\tilde{p} + (1-\tilde{p}) \cdot e^{-\lambda\xi_k}\right) \cdot \partial_1 H_\chi\left(\pi_{\xi_k}(\tilde{p}), \pi_{\xi_k}(l)\right) \cdot \pi''_{\xi_k}(\tilde{p})\right] \\
&= \sum_{k=0}^{\infty} \mathrm{E}\left[e^{-r\xi_k} \partial_1 H_\chi\left(\pi_{\xi_k}(\tilde{p}), \pi_{\xi_k}(l)\right)\left(2\left(1 - e^{-\lambda\xi_k}\right) \cdot \pi'_{\xi_k}(\tilde{p}) + \left(p + (1-p)e^{-\lambda\xi_k}\right) \cdot \pi''_{\xi_k}(\tilde{p})\right)\right] \\
&\quad + \sum_{k=0}^{\infty} \mathrm{E}\left[e^{-r\xi_k} \cdot \left(\tilde{p} + (1-\tilde{p}) \cdot e^{-\lambda\xi_k}\right) \cdot \partial_1^2 H_\chi\left(\pi_{\xi_k}(\tilde{p}), \pi_{\xi_k}(l)\right) \cdot \pi'_{\xi_k}(\tilde{p})^2\right] \\
&\overset{(i)}{=} \sum_{k=0}^{\infty} \mathrm{E}\left[e^{-r\xi_k} \left(\tilde{p} + (1-\tilde{p})e^{-\lambda\xi_k}\right) \cdot \partial_1^2 H_\chi\left(\pi_{\xi_k}(\tilde{p}), \pi_{\xi_k}(l)\right) \cdot \pi'_{\xi_k}(\tilde{p})^2\right] \\
&= \sum_{k=0}^{\infty} \mathrm{E}\left[e^{-r\xi_k} \cdot \frac{e^{-\lambda\xi_k}}{\tilde{p} + (1-\tilde{p})e^{-\lambda\xi_k}} \cdot \partial_1^2 H_\chi\left(\pi_{\xi_k}(\tilde{p}), \pi_{\xi_k}(l)\right) \cdot \pi'_{\xi_k}(\tilde{p})\right]. \qquad \text{(A.86)}
\end{aligned}$$

where equality (i) follows from

$$\pi'_t(\tilde{p}) = \frac{e^{-\lambda t}}{(\tilde{p} + (1-\tilde{p})e^{-\lambda t})^2}, \qquad \pi''_t(\tilde{p}) = -2 \cdot \frac{e^{-\lambda t}(1 - e^{-\lambda t})}{(\tilde{p} + (1-\tilde{p})e^{-\lambda t})^3}. \qquad \text{(A.87)}$$

and

$$2\left(1 - e^{-\lambda t}\right) \cdot \pi'_t(\tilde{p}) + \left(\tilde{p} + (1-\tilde{p})e^{-\lambda t}\right) \cdot \pi''_t(\tilde{p}) = \frac{2\left(1 - e^{-\lambda t}\right) \cdot e^{-\lambda t}}{(\tilde{p} + (1-\tilde{p})e^{-\lambda t})^2} - 2\frac{e^{-\lambda t}(1 - e^{-\lambda t})}{(\tilde{p} + (1-\tilde{p})e^{-\lambda t})^2} = 0.$$

Substituting the upper bound on $\frac{\partial^2}{\partial \tilde{p}^2} H\chi(\tilde{p}, l)$ in (A.80) into (A.86) obtain

$$\frac{\partial^2}{\partial \tilde{p}^2} H_\chi^*(\tilde{p}, l) \leq \sum_{k=0}^{\infty} \mathrm{E}\left[e^{-r\xi_k} \cdot \frac{e^{-\lambda\xi_k}}{\tilde{p} + (1-\tilde{p})e^{-\lambda\xi_k}} \left(u'_{\chi,1}(\pi_{\xi_k}(\tilde{p})) - u'_{\chi,0}(\pi_{\xi_k}(\tilde{p}))\right) \cdot \pi'_{\xi_k}(\tilde{p})\right]. \qquad \text{(A.88)}$$

We can express $u'_{\chi,\theta}(\pi_t(\tilde{p}))$ in its integral form as

$$
\begin{aligned}
u'_{\chi,\theta}(\pi_t(\tilde{p})) &= \int_0^\infty e^{-(r+\chi+\lambda(1-\theta))s} \cdot A'\Big(\pi_s(\pi_t(\tilde{p}))\Big) \cdot \frac{e^{-\lambda s}}{(\pi_t(\tilde{p}) + (1-\pi_t(\tilde{p})e^{-\lambda s})^2} \, ds \\
&= \int_0^\infty e^{-(r+\chi+\lambda(1-\theta))s} \cdot A'\Big(\pi_{t+s}(\tilde{p})\Big) \cdot \frac{e^{-\lambda s}}{\Big(\frac{\tilde{p}}{\tilde{p}+(1-\tilde{p})e^{-\lambda t}} + \frac{(1-\tilde{p})e^{-\lambda t}}{\tilde{p}+(1-\tilde{p})e^{-\lambda t}} e^{-\lambda s}\Big)^2} \, ds \\
&= \int_0^\infty e^{-(r+\chi+\lambda(1-\theta))s} \cdot A'\Big(\pi_{t+s}(\tilde{p})\Big) \cdot \frac{e^{-\lambda s} \cdot \big(\tilde{p} + (1-\tilde{p})e^{-\lambda t}\big)^2}{\big(\tilde{p} + (1-\tilde{p})e^{-\lambda(t+s)}\big)^2} \, ds.
\end{aligned}
$$

Then

$$
\begin{aligned}
e^{-rt} u'_{\chi,\theta}(\pi_t(\tilde{p}))\pi'_t(\tilde{p}) &= e^{-rt} \left[\int_0^\infty e^{-(r+\chi+\lambda(1-\theta))s} A'\Big(\pi_{t+s}(\tilde{p})\Big) \frac{e^{-\lambda s} \big(\tilde{p} + (1-\tilde{p})e^{-\lambda t}\big)^2}{\big(\tilde{p} + (1-\tilde{p})e^{-\lambda(t+s)}\big)^2} ds \right] \frac{e^{-\lambda t}}{(\tilde{p} + (1-\tilde{p})e^{-\lambda t})^2} \\
&= \int_0^\infty e^{-(\chi+\lambda(1-\theta))s} \cdot e^{-r(t+s)} \cdot A'(\pi_{t+s}(\tilde{p})) \cdot \frac{e^{-\lambda(t+s)}}{\big(\tilde{p} + (1-\tilde{p})e^{-\lambda(t+s)}\big)^2} \, ds \\
&= \int_t^\infty e^{-(\chi+\lambda(1-\theta))(s-t)} \cdot e^{-rs} \cdot A'(\pi_s(\tilde{p})) \cdot \frac{e^{-\lambda s}}{(\tilde{p} + (1-\tilde{p})e^{-\lambda s})^2} \, ds.
\end{aligned}
$$

The upper bound on $\frac{\partial^2}{\partial \tilde{p}^2} H^*_\chi(\tilde{p}, l)$ in (A.88) can then be simplified to

$$
\sum_{k=0}^{\infty} \mathrm{E} \left[\frac{e^{-\lambda \xi_k}}{\tilde{p} + (1-\tilde{p})e^{-\lambda \xi_k}} \cdot \int_{\xi_k}^\infty e^{-\chi(s-\xi_k)} \cdot \Big(1 - e^{-\lambda(s-\xi_k)}\Big) \cdot e^{-rs} \cdot A'(\pi_s(\tilde{p})) \cdot \frac{e^{-\lambda s}}{(\tilde{p} + (1-\tilde{p})e^{-\lambda s})^2} \, ds \right]
$$

Consider term $k = 0$. In this case $\xi_0 \equiv 0$ and we obtain

$$
\int_0^\infty e^{-(r+\chi)s} \cdot \Big(1 - e^{-\lambda s}\Big) \cdot A'(\pi_s(\tilde{p})) \cdot \frac{e^{-\lambda s}}{(\tilde{p} + (1-\tilde{p}) \cdot e^{-\lambda s})^2} \, ds. \tag{A.89}
$$

Consider terms $k = 1, 2, \ldots$. In this case

$$
\begin{aligned}
&\mathrm{E} \left[\frac{e^{-\lambda \xi_k}}{\tilde{p} + (1-\tilde{p})e^{-\lambda \xi_k}} \cdot \int_{\xi_k}^\infty e^{-\chi(s-\xi_k)} \cdot \Big(1 - e^{-\lambda(s-\xi_k)}\Big) \cdot e^{-rs} \cdot A'(\pi_s(\tilde{p})) \cdot \frac{e^{-\lambda s}}{(\tilde{p} + (1-\tilde{p})e^{-\lambda s})^2} \, ds \right] \\
&= \int_0^\infty \phi(t,k) \cdot \frac{e^{-\lambda t}}{\tilde{p} + (1-\tilde{p})e^{-\lambda t}} \cdot \int_t^\infty e^{-\chi(s-t)} \cdot \Big(1 - e^{-\lambda(s-t)}\Big) \cdot e^{-rs} \cdot A'(\pi_s(\tilde{p})) \cdot \frac{e^{-\lambda s}}{(\tilde{p} + (1-\tilde{p})e^{-\lambda s})^2} \, ds \, dt \\
&= \int_0^\infty e^{-rs} \cdot A'(\pi_s(\tilde{p})) \cdot \frac{e^{-\lambda s}}{(\tilde{p} + (1-\tilde{p})e^{-\lambda s})^2} \cdot \int_0^s \phi(t,k) \cdot \frac{e^{-\lambda t}}{\tilde{p} + (1-\tilde{p})e^{-\lambda t}} \cdot e^{-\chi(s-t)} \cdot \Big(1 - e^{-\lambda(s-t)}\Big) \, dt \, ds
\end{aligned}
$$

Density function $\phi(t,k)$ follows Erlang distribution given by

$$\phi(t,k) \overset{def}{=} \frac{\chi^k \cdot t^{k-1} \cdot e^{-\chi t}}{(k-1)!}.$$

Substituting $\phi(t,k)$ for $k \in \mathbb{N}$ obtain

$$\int_0^\infty e^{-rs} \cdot A'(\pi_s(\tilde{p})) \cdot \frac{e^{-\lambda s}}{\left(\tilde{p} + (1-\tilde{p})e^{-\lambda s}\right)^2} \cdot \int_0^s \phi(t,k) \cdot \frac{e^{-\lambda t}}{\tilde{p} + (1-\tilde{p})e^{-\lambda t}} \cdot e^{-\chi(s-t)} \cdot \left(1 - e^{-\lambda(s-t)}\right) dt\, ds$$

$$= \int_0^\infty e^{-rs} \cdot A'(\pi_s(\tilde{p})) \cdot \frac{e^{-\lambda s}}{\left(\tilde{p} + (1-\tilde{p})e^{-\lambda s}\right)^2} \cdot \int_0^s \frac{\chi^k \cdot t^{k-1} \cdot e^{-\chi t}}{(k-1)!} \cdot \frac{e^{-\lambda t}}{\tilde{p} + (1-\tilde{p})e^{-\lambda t}} \cdot e^{-\chi(s-t)} \cdot \left(1 - e^{-\lambda(s-t)}\right) dt\, ds$$

$$= \int_0^\infty e^{-(r+\chi)s} \cdot A'(\pi_s(\tilde{p})) \cdot \frac{e^{-\lambda s}}{\left(\tilde{p} + (1-\tilde{p})e^{-\lambda s}\right)^2} \cdot \int_0^s \frac{\chi^k \cdot t^{k-1}}{(k-1)!} \cdot \frac{e^{-\lambda t} - e^{-\lambda s}}{\tilde{p} + (1-\tilde{p})e^{-\lambda t}}\, dt\, ds$$

$$= \int_0^\infty \chi \cdot e^{-(r+\chi)s} \cdot A'(\pi_s(\tilde{p})) \cdot \frac{e^{-\lambda s}}{\left(\tilde{p} + (1-\tilde{p})e^{-\lambda s}\right)^2} \cdot \int_0^s \frac{(\chi \cdot t)^{k-1}}{(k-1)!} \cdot \frac{e^{-\lambda t} - e^{-\lambda s}}{\tilde{p} + (1-\tilde{p})e^{-\lambda t}}\, dt\, ds$$

The sum of terms for $k = 1, \ldots,$ is

$$\sum_{k=1}^\infty \int_0^\infty \chi \cdot e^{-(r+\chi)s} \cdot A'(\pi_s(\tilde{p})) \cdot \frac{e^{-\lambda s}}{\left(\tilde{p} + (1-\tilde{p})e^{-\lambda s}\right)^2} \cdot \int_0^s \frac{(\chi \cdot t)^{k-1}}{(k-1)!} \cdot \frac{e^{-\lambda t} - e^{-\lambda s}}{\tilde{p} + (1-\tilde{p})e^{-\lambda t}}\, dt\, ds$$

$$= \int_0^\infty \chi \cdot e^{-(r+\chi)s} \cdot A'(\pi_s(\tilde{p})) \cdot \frac{e^{-\lambda s}}{\left(\tilde{p} + (1-\tilde{p})e^{-\lambda s}\right)^2} \cdot \int_0^s \left(\sum_{k=1}^\infty \frac{(\chi \cdot t)^{k-1}}{(k-1)!}\right) \cdot \frac{e^{-\lambda t} - e^{-\lambda s}}{\tilde{p} + (1-\tilde{p})e^{-\lambda t}}\, dt\, ds$$

$$= \int_0^\infty \chi \cdot e^{-(r+\chi)s} \cdot A'(\pi_s(\tilde{p})) \cdot \frac{e^{-\lambda s}}{\left(\tilde{p} + (1-\tilde{p})e^{-\lambda s}\right)^2} \cdot \int_0^s e^{\chi t} \cdot \frac{e^{-\lambda t} - e^{-\lambda s}}{\tilde{p} + (1-\tilde{p})e^{-\lambda t}}\, dt\, ds$$

$$= \int_0^\infty e^{-(r+\chi)s} \cdot A'(\pi_s(\tilde{p})) \cdot \frac{e^{-\lambda s}}{\left(\tilde{p} + (1-\tilde{p})e^{-\lambda s}\right)^2} \cdot \int_0^s \chi \cdot e^{\chi t} \cdot \frac{e^{-\lambda t} - e^{-\lambda s}}{\tilde{p} + (1-\tilde{p})e^{-\lambda t}}\, dt\, ds. \tag{A.90}$$

Summing up (A.89) and (A.90) obtain

$$\begin{aligned}
\frac{\partial^2}{\partial \tilde{p}^2} H^*_\chi(\tilde{p}, l) \leq & \int_0^\infty e^{-(r+\chi)s} \cdot \left(1 - e^{-\lambda s}\right) \cdot A'(\pi_s(\tilde{p})) \cdot \frac{e^{-\lambda s}}{(\tilde{p} + (1-\tilde{p}) \cdot e^{-\lambda s})^2}\, ds \\
& + \int_0^\infty e^{-(r+\chi)s} \cdot A'(\pi_s(\tilde{p})) \cdot \frac{e^{-\lambda s}}{\left(\tilde{p} + (1-\tilde{p})e^{-\lambda s}\right)^2} \cdot \int_0^s \chi \cdot e^{\chi t} \cdot \frac{e^{-\lambda t} - e^{-\lambda s}}{\tilde{p} + (1-\tilde{p})e^{-\lambda t}}\, dt\, ds \\
& = \int_0^\infty e^{-rs} \left(1 - e^{-\lambda s}\right) \frac{A'(\pi_s(\tilde{p})) \cdot e^{-\lambda s}}{(\tilde{p} + (1-\tilde{p})e^{-\lambda s})^2} \left[e^{-\chi s} + \int_0^s \chi \cdot e^{\chi(t-s)} \cdot \frac{\frac{e^{-\lambda t} - e^{-\lambda s}}{1 - e^{-\lambda s}}}{\tilde{p} + (1-\tilde{p})e^{-\lambda t}}\, dt\right] ds.
\end{aligned}$$

Note that

$$
\begin{aligned}
&\int_0^s \chi \cdot e^{\chi(t-s)} \cdot \frac{\frac{e^{-\lambda t}-e^{-\lambda s}}{1-e^{-\lambda s}}}{\tilde{p}+(1-\tilde{p})e^{-\lambda t}}\, dt \leq \frac{1}{\tilde{p}} \cdot \int_0^s \chi \cdot e^{\chi(t-s)} \cdot \frac{e^{-\lambda t}-e^{-\lambda s}}{1-e^{-\lambda s}}\, dt \\
=&\frac{1}{\tilde{p}} \cdot \frac{1}{1-e^{-\lambda s}} \cdot \left[e^{-\chi s} \cdot \int_0^s \chi e^{(\chi-\lambda)t}\, dt - e^{-(\chi+\lambda)s} \int_0^s \chi e^{\chi t}\, ds\right] \\
=&\frac{1}{\tilde{p}} \cdot \frac{1}{1-e^{-\lambda s}} \cdot \left[e^{-\chi s} \cdot \frac{\chi}{\chi-\lambda} \cdot e^{(\chi-\lambda)t}\Big|_{t=0}^{t=s} - e^{-(\chi+\lambda)s} \cdot e^{\chi t}\Big|_{t=0}^{t=s}\right] \\
=&\frac{1}{\tilde{p}} \cdot \frac{1}{1-e^{-\lambda s}} \cdot \left[e^{-\chi s} \cdot \frac{\chi}{\chi-\lambda} \cdot \left(e^{(\chi-\lambda)s}-1\right) - e^{-(\chi+\lambda)s} \cdot (e^{\chi s}-1)\right] \\
=&\frac{1}{\tilde{p}} \cdot \frac{1}{1-e^{-\lambda s}} \cdot \left[\frac{\chi}{\chi-\lambda} \cdot \left(e^{-\lambda s}-e^{-\chi s}\right) - \left(e^{-\lambda s}-e^{-(\chi+\lambda)s}\right)\right] \\
=&\frac{1}{\tilde{p}} \cdot \frac{1}{1-e^{-\lambda s}} \cdot \left[\frac{\lambda}{\chi-\lambda} \cdot e^{-\lambda s} - \frac{\chi}{\chi-\lambda} \cdot e^{-\chi s} + e^{-(\chi+\lambda)s}\right]. \qquad \text{(A.91)}
\end{aligned}
$$

Substitute (A.91) into the upper bound for $\frac{\partial^2}{\partial \tilde{p}^2} H_\chi^*(\tilde{p}, l)$ to obtain

$$
\int_0^\infty e^{-rs} \frac{A'(\pi_s(\tilde{p})) \cdot e^{-\lambda s}}{(\tilde{p}+(1-\tilde{p})e^{-\lambda s})^2} \cdot \left[e^{-\chi s} \cdot \left(1-e^{-\lambda s}\right) + \frac{1}{\tilde{p}} \cdot \left[\frac{\lambda}{\chi-\lambda} e^{-\lambda s} - \frac{\chi}{\chi-\lambda} e^{-\chi s} + e^{-(\chi+\lambda)s}\right]\right] ds.
$$

This upper bound converges to 0 as $\chi \to \infty$. It is independent of l and uniformly bounded in $\tilde{p}$ for $\tilde{p} \geq \underline{p} > 0$. Consequently $\frac{\partial^2}{\partial \tilde{p}^2} H_\chi^*(\tilde{p}, l)$ converges uniformly to 0 in $\tilde{p}$ and l as $\chi \to \infty$.

□

Intermediary Equilibrium Value Function Properties

Denote by $T(\tilde{p}_t)$ the optimal time when the agent of posterior ability $\tilde{p}_t$ leaves the intermediary along the path of good performance. Consider the joint surplus value between the agent and the intermediary given by

$$
\begin{aligned}
W_t(\tilde{p}_t) &\overset{def}{=} \sup_\tau \left\{\int_t^\tau e^{-r(s-t)} \cdot \left(\tilde{p}_t + (1-\tilde{p}_t) \cdot e^{-\lambda(s-t)}\right) \cdot (A(q_s) - rV)\, ds + e^{-r(\tau-t)} \cdot U(\tilde{p}_\tau, k_\tau) + V\right\} \\
&= \sup_T \left\{\int_t^{T\wedge\eta} e^{-r(s-t)} \cdot \left(\tilde{p}_t + (1-\tilde{p}_t) \cdot e^{-\lambda(s-t)}\right) \cdot (A(q_s) - rV)\, ds + e^{-r(T\wedge\eta-t)} \cdot U(\tilde{p}_{T\wedge\eta}, k_{T\wedge\eta}) + V\right\}.
\end{aligned}
$$

It follows from the optimality of separation times it follows that

$$
W_t(\tilde{p}_t, k_t) \geq U(\tilde{p}_t, k_t) = \tilde{p}_t \cdot U(1, k_t) + (1-\tilde{p}_t) \cdot U(0, k_t).
$$

As we have shown in constructing the equilibrium in Proposition 2, the intermediary finds higher-skilled agents more valuable. This implies that the intermediary's equilibrium share of the welfare is the difference between total welfare and the agent's outside option: $W_t(\tilde{p}_t) - U(\tilde{p}_t, k_t)$. Since it is increasing in $\tilde{p}_t$, it follows that $\frac{\partial}{\partial \tilde{p}} W_t(\tilde{p}_t) \geq u_1(k_t) - u_0(k_t)$.

Lemma A.25 (Strict Value Function Convexity)**.** *The second derivative of $W_t(\tilde{p}_t, k)$ with respect to $\tilde{p}_t$ can be expressed as*

$$\frac{\partial^2}{\partial \tilde{p}^2} W_t(\tilde{p}_t, k) = \Big[u_1'(k_t) - u_0'(k_t)\Big] \cdot e^{-r(T(\tilde{p}_t)-t)} \cdot \frac{e^{-\lambda(T(\tilde{p}_t)-t)}}{\left(\tilde{p}_t + (1-\tilde{p}_t) \cdot e^{-\lambda(T(\tilde{p}_t)-t)}\right)^3}, \tag{A.92}$$

where $T(\tilde{p}_t)$ is the equilibrium time when time t posterior type $\tilde{p}_t$ separates from the intermediary.

Proof. The first derivative of W_t with respect to p is given by the Envelope theorem as

$$\begin{aligned} W_t(\tilde{p}_t, k_t) \overset{def}{=} & \int_t^{T(\tilde{p}_t)} e^{-r(s-t)} \cdot \left(1 - \cdot e^{-\lambda(s-t)}\right) \cdot (A(q_s) - rV)\; ds \\ & + e^{-r(T(\tilde{p}_t)-t)} \cdot \left[u_1\left(k_{T(\tilde{p}_t)}\right) - e^{-\lambda(T(\tilde{p}_t)-t)} \cdot u_0\left(k_{T(\tilde{p}_t)}\right)\right]. \end{aligned} \tag{A.93}$$

Without loss set $t = 0$ and drop the subscript t, thus referring to ex-ante types. Note that $\pi_{T(\tilde{p})}(\tilde{p}) \equiv k_{T(\tilde{p})}$ for all p. Differentiating this identity with respect to $\tilde{p}$ obtain

$$\begin{aligned} \pi'_{T(\tilde{p})}(\tilde{p}) + \dot{\pi}_{T(\tilde{p})}(\tilde{p}) \cdot T'(\tilde{p}) = \dot{k}_{T(\tilde{p})} \cdot T'(\tilde{p}) & \\ \pi'_{T(\tilde{p})}(\tilde{p}) &= \frac{A(k_{T(\tilde{p})} + rV - A\left(q_{T(\tilde{p})}\right)}{\partial_2 U\left(k_{T(\tilde{p})}, k_{T(\tilde{p})}\right)} \cdot T'(\tilde{p}) \\ T'(\tilde{p}) &= \pi'_{T(\tilde{p})}(\tilde{p}) \cdot \frac{A(k_{T(\tilde{p})} + rV - A\left(q_{T(\tilde{p})}\right)}{\partial_2 U\left(k_{T(\tilde{p})}, k_{T(\tilde{p})}\right)}. \end{aligned} \tag{A.94}$$

Differentiating (A.93) obtain

$$\begin{aligned} \frac{\partial^2}{\partial \tilde{p}^2} W(\tilde{p}, k) &= e^{-rT(\tilde{p})} \cdot \left(1 - e^{-\lambda T(\tilde{p})}\right) \cdot \left(A\left(q_{T(\tilde{p})}\right) - rV\right) \cdot T'(\tilde{p}) \\ &+ e^{-rT(\tilde{p})} \cdot T'(\tilde{p}) \cdot \left[-ru_1\left(k_{T(\tilde{p})}\right) + u_1'\left(k_{T(\tilde{p})}\right) \cdot \dot{k}_{T(\tilde{p})}\right] \\ &- e^{-(r+\lambda)T(\tilde{p})} \cdot T'(\tilde{p}) \cdot \left[-(r+\lambda)u_0\left(k_{T(\tilde{p})}\right) + u_0'\left(k_{T(\tilde{p})}\right) \cdot \dot{k}_{T(\tilde{p})}\right] \\ &= e^{-rT(\tilde{p})} \cdot T'(\tilde{p}) \cdot \left[A\left(q_{T(\tilde{p})}\right) - rV - ru_1\left(k_{T(\tilde{p})}\right) + u_1'\left(k_{T(\tilde{p})}\right) \cdot \dot{k}_{T(\tilde{p})}\right] \\ &- e^{-(r+\lambda)T(\tilde{p})} \cdot T'(\tilde{p}) \cdot \left[A\left(q_{T(\tilde{p})}\right) - rV - (r+\lambda)u_0\left(k_{T(\tilde{p})}\right) + u_0'\left(k_{T(\tilde{p})}\right) \cdot \dot{k}_{T(\tilde{p})}\right] \\ &= e^{-rT(\tilde{p})} \cdot T'(\tilde{p}) \cdot \left[A\left(q_{T(\tilde{p})}\right) - rV - A\left(k_{T(\tilde{p})}\right) + u_1'\left(k_{T(\tilde{p})}\right) \cdot \gamma_{T(\tilde{p})}\right] \\ &- e^{-(r+\lambda)T(\tilde{p})} \cdot T'(\tilde{p}) \cdot \left[A\left(q_{T(\tilde{p})}\right) - rV - A\left(k_{T(\tilde{p})}\right) + u_0'\left(k_{T(\tilde{p})}\right) \cdot \gamma_{T(\tilde{p})}\right] \\ &= e^{-rT(\tilde{p})} \cdot T'(\tilde{p}) \cdot \left[A\left(k_{T(\tilde{p})}\right) + rV - A\left(q_{T(\tilde{p})}\right)\right] \cdot \left[\frac{u_1'(k_{T(\tilde{p})}}{\partial_2 U\left(k_{T(\tilde{p})}, k_{T(\tilde{p})}\right)} - 1\right] \end{aligned}$$

$$
\begin{aligned}
&- e^{-(r+\lambda)T(\tilde{p})} \cdot T'(\tilde{p}) \cdot \left[A\left(k_{T(\tilde{p})}\right) + rV - A\left(q_{T(\tilde{p})}\right)\right] \cdot \left[\frac{u_0'(k_{T(\tilde{p})}}{\partial_2 U\left(k_{T(\tilde{p})}, k_{T(\tilde{p})}\right)} - 1\right] \\
&= e^{-rT(\tilde{p})} \cdot \pi_{T(\tilde{p})}'(\tilde{p}) \cdot \left[u_1'\left(k_{T(\tilde{p})}\right) - \partial_2 U\left(k_{T(\tilde{p})}, k_{T(\tilde{p})}\right)\right] \\
&- e^{-(r+\lambda)T(\tilde{p})} \cdot \pi_{T(\tilde{p})}'(\tilde{p}) \cdot \left[u_0'\left(k_{T(\tilde{p})}\right) - \partial_2 U\left(k_{T(\tilde{p})}, k_{T(\tilde{p})}\right)\right] \\
&= e^{-rT(\tilde{p})} \cdot \pi_{T(\tilde{p})}'(\tilde{p}) \cdot (1 - k_{T(\tilde{p})}) \cdot \left[u_1'\left(k_{T(\tilde{p})}\right) - u_0'\left(k_{T(\tilde{p})}\right)\right] \\
&+ e^{-(r+\lambda)T(\tilde{p})} \cdot \pi_{T(\tilde{p})}'(\tilde{p}) \cdot k_{T(\tilde{p})} \cdot \left[u_1'\left(k_{T(\tilde{p})}\right) - u_0'\left(k_{T(\tilde{p})}\right)\right] \\
&\overset{t=T(\tilde{p})}{=} \left[u_1'(k_t) - u_0'(k_t)\right] \cdot e^{-rt} \cdot \pi_t'(\tilde{p}) \cdot \left[1 - \pi_t(\tilde{p}) + \pi_t(\tilde{p}) \cdot e^{-\lambda t}\right] \\
&= \left[u_1'(k_t) - u_0'(k_t)\right] \cdot e^{-rt} \cdot \frac{e^{-\lambda t}}{\left(\tilde{p} + (1-\tilde{p}) \cdot e^{-\lambda t}\right)^3}. \qquad \text{(A.95)}
\end{aligned}
$$

Since separation occurs in finite time, which follows from Lemma A.26 below, the expression in (A.95) is bounded from below. □

Sufficiency of Intermediary Signaling

Corollary A.1. *There exists a $\bar{\chi} > 0$ such that for all $\chi > \bar{\chi}$ the intermediary's value added $W_t(\tilde{p}_t, k_t) - V$ exceeds the expected value $H_\chi^*(\tilde{p}_t, k_t)$ that the agent can obtain via independent signaling, implying that the intermediary's retention strategy coincides with that constructed in Proposition 2.*

Proof. The optimality of the churning time $T(\tilde{p})$ implies that

$$W_t(k_t, k_t) - V \geq k \cdot u_1(k_t) + (1 - k_t) \cdot u_0(k_t) = H_\chi^*(k_t, k_t).$$

For any t, $\tilde{p}_t$, and k_t it follows from Lemma A.1 and Lemma A.24 that

$$\frac{\partial}{\partial p} W_t(\tilde{p}_t, k_t)\Big|_{\tilde{p}_t = k_t} \geq u_1(k_t) - u_0(k_t) = \frac{\partial}{\partial p} H_\chi^*(\tilde{p}_t, k_t)\Big|_{\tilde{p}_t = k_t}.$$

It follows from Lemma A.24 that $\frac{\partial^2}{\partial p^2} H_\chi^*(\tilde{p}_t, k_t)$ converges to 0 uniformly in $\tilde{p}_t$ and k_t. At the same time, it follows from (A.92) that $\frac{\partial^2}{\partial p^2} W_t(\tilde{p}_t, k_t)$ is strictly bounded from below whenever $T(\tilde{p}_t)$ is finite. Thus, there exists a finite $\bar{\chi}$ such that $\frac{\partial}{\partial p} W_t(\tilde{p}_t, k_t) \geq \frac{\partial}{\partial p} H_\chi^*(\tilde{p}_t, k_t)$ for $\tilde{p}_t \geq k_t$. □

A.1 Extension: Equilibrium Uniqueness under Only Off-Path Monotonicity

In this section, we relax the global belief monotonicity condition *(i)* of Observation 2 to off-path belief monotonicity. While such refinement can be further weakened, we adopt it here for clarity. For ease of notation below, belief p denotes the ex-ante type $\tilde{p}_0$.

Observation A.1 (Monotone equilibrium uniqueness relaxed)**.** *Suppose the clients' belief process* $(l_t)_{t\geq 0}$ *is upper semi-continuous and is*

(i) weakly increasing in initial types over time during quiet periods, i.e., $l_t \overset{def}{=} \pi_{-t}(k_t)$ *is weakly increasing in* $t \in [t_1, t_2]$ *for every* $(t_1, t_2) \notin$ *support*(τ)

(ii) ***never dissuaded once convinced:*** l_t *assigns zero probability to private types* $\tilde{p}_t$ *who separate with certainty before time* t *in equilibrium.*

Then, the equilibrium constructed in Proposition 2 is unique.

We add a rather weak technical requirement that the belief process $(l_t)_{t\geq 0}$ is upper semi-continuous – this rules out negative belief jumps at the end of a given churning period, which can be self-fulfilling.[4]

Similar to l_t, denote by $m_t \overset{def}{=} \pi_{-t}(q_t) = \mathrm{E}\left[\tilde{p}_0 \,|\, t < \tau, t < \eta\right]$. We can rewrite (36) and express the set of ex-ante times when it is optimal to let the agent of ex-ante ability p go even if he performs well $\mathbb{T}^I(p)$ as

$$\begin{aligned}\mathbb{T}^I(p) &\overset{def}{=} \arg\max_t \Bigg\{ \mathrm{E}_{\tilde{p}_0 = p}\left[\int_0^{t\wedge\eta} e^{-rs}\cdot\left[A(\pi_s(m_s)) - rV\right]\, ds + \int_{t\wedge\eta}^{t} e^{-rs}\cdot rL\, ds\right] + V \\ &\qquad + \mathrm{E}_{\tilde{p}_0=p}\left[e^{-rt}\cdot U^*(\pi_t(p), \pi_t(l_t))\cdot \mathbb{1}\{t<\eta\} + e^{-rt}\cdot L\cdot \mathbb{1}\{t\geq\eta\}\right]\Bigg\} \\ &= \arg\max_t \Bigg\{ \mathrm{E}_{\tilde{p}_0=p}\left[\int_0^{t\wedge\eta} e^{-rs}\cdot\left[A(\pi_s(m_s)) - rV - rL\right]\, ds + V\right] \\ &\qquad + \mathrm{E}_{\tilde{p}_0=p}\left[e^{-rt}\cdot\max\Big\{U(\pi_t(p), \pi_t(l_t)) - L,\, 0\Big\}\cdot\mathbb{1}\{t<\eta\}\right] + L\Bigg\}.\end{aligned} \tag{A.96}$$

Since the optimal turnover times $\mathbb{T}^I(p)$ are efficient in the presence of flexible wages, the set of times $\mathbb{T}^I(p)$ maximize both the joint welfare of the intermediary-agent pair, as well as the intermediary's residual payoff given the outside option $U^*(p, k_0)$ of the agent. effectively maximizes the joint welfare of the intermediary-agent pair.

Lemma A.26 (Upper bound on separation times)**.** *Suppose* $r\cdot V > 0$*. Define*

$$\bar{T} \overset{def}{=} \inf\left\{t:\ 1 - A\left(\pi_t\left(\underline{p}\right)\right) < \frac{rV}{2}\right\} + \frac{\ln(2)}{r}. \tag{A.97}$$

[4] The assumption of upper semi-continuity can be weakened, we adopt it to shorten the proof.

Then, in any equilibrium, all agents find it strictly optimal to separate from the intermediary by time $\bar{T}$ in any equilibrium, i.e., in any equilibrium $\bar{t} \stackrel{def}{=} support(\tau) \le \sup_p \mathbb{T}^I(p) \le \bar{T}$.

Proof. Define $T_1 \stackrel{def}{=} \inf\{t : 1 - A(\pi_t(\underline{p})) < rV/2\}$. Suppose there exists a $t \in \mathbb{T}$ such that $t > \bar{T} > \bar{T}_1$. The expected payoff to the intermediary and the agent pair from stopping at time t is equal to

$$\begin{aligned} W_0(p;t) \stackrel{def}{=} &\int_0^t e^{-rs} \cdot \left(p + (1-p) \cdot e^{-\lambda s}\right) \cdot [A(\pi_s(m_s)) - rV]\; ds + V \\ &+ \int_t^{+\infty} e^{-rs} \cdot \left(p + (1-p) \cdot e^{-\lambda s}\right) \cdot A(\pi_s(l_t))\, ds. \end{aligned} \tag{A.98}$$

Compare this value with the value of waiting until time T_1 and separating at the lowest possible belief $\underline{p}$:

$$\begin{aligned} \underline{W}_0(p;T_1) \stackrel{def}{=} &\int_0^{T_1} e^{-rs} \cdot \left(p + (1-p) \cdot e^{-\lambda s}\right) \cdot [A(\pi_s(m_s)) - rV]\; ds + V \\ &+ \int_{T_1}^{+\infty} e^{-rs} \cdot \left(p + (1-p) \cdot e^{-\lambda s}\right) \cdot A(\pi_s(\underline{p}))\, ds. \end{aligned}$$

The difference in expected values between separating at time T_1 while being perceived as the worst possible type and separating at time $t \ge \bar{T} > \bar{T}_1$ along the path of good performance is

$$\begin{aligned} W_0(p;t) - \underline{W}_0(p;T_1) = &\int_{T_1}^t e^{-rs} \cdot \left(p + (1-p) \cdot e^{-\lambda s}\right) \cdot [A(\pi_s(m_s)) - rV]\; ds \\ &+ \int_t^{+\infty} e^{-rs} \cdot \left(p + (1-p) \cdot e^{-\lambda s}\right) \cdot A(\pi_s(l_t))\, ds \\ &- \int_{T_1}^{+\infty} e^{-rs} \cdot \left(p + (1-p) \cdot e^{-\lambda s}\right) \cdot A(\pi_s(\underline{p}))\, ds \\ = &\int_{T_1}^t e^{-rs} \cdot \left(p + (1-p) \cdot e^{-\lambda s}\right) \cdot \left[A(\pi_s(m_s)) - rV - A(\pi_s(\underline{p}))\right]\; ds \\ &+ \int_t^{+\infty} e^{-rs} \cdot \left(p + (1-p) \cdot e^{-\lambda s}\right) \cdot \left[A(\pi_s(l_t)) - A(\pi_s(\underline{p}))\right]\; ds \\ \le &\int_{T_1}^t e^{-rs} \cdot \left(p + (1-p) \cdot e^{-\lambda s}\right) \cdot \left[A(1) - rV - A(\pi_s(\underline{p}))\right]\; ds \\ &+ \int_t^{+\infty} e^{-rs} \cdot \left(p + (1-p) \cdot e^{-\lambda s}\right) \cdot \left[A(1) - A(\pi_s(\underline{p}))\right]\; ds \\ \stackrel{(i)}{<} &- \int_{T_1}^t e^{-rs} \cdot \left(p + (1-p) \cdot e^{-\lambda s}\right) \cdot \frac{rV}{2}\, ds \\ &+ \int_t^{+\infty} e^{-rs} \cdot \left(p + (1-p) \cdot e^{-\lambda s}\right) \cdot \frac{rV}{2}\, ds \\ = &p \cdot \frac{V}{2} \cdot \left(e^{-rt} - e^{-rT_1}\right) + (1-p) \cdot \frac{rV}{2(r+\lambda)} \cdot \left(e^{-(r+\lambda)t} - e^{-(r+\lambda)T_1}\right) \\ &+ p \cdot \frac{V}{2} \cdot e^{-rt} + (1-p) \cdot \frac{rV}{2(r+\lambda)} \cdot e^{-(r+\lambda)t} \end{aligned}$$

$$
\begin{aligned}
&= p \cdot \frac{V}{2} \cdot \left(2e^{-rt} - e^{-rT_1}\right) + (1-p) \cdot \frac{rV}{2(r+\lambda)} \cdot \left(2e^{-(r+\lambda)t} - e^{-(r+\lambda)T_1}\right) \\
&\overset{(ii)}{\leq} p \cdot \frac{V}{2} \cdot \left(2e^{-r(T_1+\ln(2)/r)} - e^{-rT_1}\right) \\
&+ (1-p) \cdot \frac{rV}{2(r+\lambda)} \cdot \left(2e^{-(r+\lambda)(T_1+\ln(2)/r)} - e^{-(r+\lambda)T_1}\right) \\
&\leq p \cdot \frac{V}{2} \cdot \left(e^{-rT_1} - e^{-rT_1}\right) + (1-p) \cdot \frac{rV}{2(r+\lambda)} \cdot \left(e^{-(r+\lambda)T_1} - e^{-(r+\lambda)T_1}\right) = 0,
\end{aligned}
$$

where the strict inequality (i) follows from the definition of $\bar{T}_1$ and inequality (ii) holds because $t > \bar{T} = T_1 + \frac{\ln(2)}{r}$. Hence, waiting past $\bar{T}$ to separate from the intermediary is strictly suboptimal. □

Lemma A.27 (Monotonicity of separating types)**.** *Suppose belief process $l = (l_t)_{t\geq 0}$ is upper semi continuous. For every $t \in support(\tau) \cap (0, +\infty)$ it must be the case that $l_t \leq m_t$ with the inequality being strict unless $t = \bar{t} \overset{def}{=} \sup\{support(\tau)\}$, implying that process $m = (m_t)_{t\geq 0}$ is weakly increasing.*

Proof. Suppose by contradiction there exists a $t_0 \in support(\tau)$ such that $t_0 > 0$ and $l_{t_0} > m_0$. If $U(\pi_{t_0}(l_{t_0}), \pi_{t_0}(l_{t_0})) < L$ for $t \in support(\tau)$. It follows from Lemma 3 in the Print Appendix, which does not require monotonicity of l_t, that it must be the case that $t_0 = 0$.

Consider the remaining case of $U(\pi_{t_0}(l_{t_0}), \pi_{t_0}(l_{t_0})) \geq L$: the agent of ex-ante ability l_{t_0} is willing to open his own firm when churned at time t_0. Denote by $t_1 > t_0$ as the next instance when it is weakly optimal for one of the types to separate at beliefs $l_t \leq m_t$:

$$
t_1 \overset{def}{=} \inf\left\{t \in \mathbb{T}^I(p) : \qquad t > t_0 \quad \textit{and} \quad l_t \leq m_t\right\}.
$$

This time is well-defined since $\bar{t} = \sup support(\tau)$ is finite, as seen in Lemma A.26. Moreover, $t_1 > t_0$ since, otherwise, $l_{t_0+} \leq m_{t_0} < l_{t_0}$ and type p would find it strictly optimal to separate at time t_0. Denote by p the ex-ante type willing to separate from the intermediary at time t_1 conditional on good performance. Consider two cases depending on whether type p opens his firm at time t_1 conditional on good performance.

(i) Suppose that $U(\pi_{t_1}(p), \pi_{t_1}(l_{t_1})) \geq L$. Then the optimality of type p to separate at time t_1 versus t_0 is given by

$$
\begin{aligned}
&\int_{t_0}^{t_1} e^{-r(s-t_0)} \cdot \left(p + (1-p) \cdot e^{-\lambda(s-t_0)}\right) \cdot \left[A(\pi_s(m_s)) - rV - rL\right] ds \\
&+ \int_{t_1}^{\infty} e^{-r(s-t_0)} \cdot \left(p + (1-p) \cdot e^{-\lambda(s-t_0)}\right) \cdot \left[A\left(\pi_s(l_{t_1})\right) - rL\right] ds + L \\
&\geq \int_{t_0}^{\infty} e^{-r(s-t_0)} \cdot \left(p + (1-p) \cdot e^{-\lambda(s-t_0)}\right) \cdot \left[A(\pi_s(l_{t_0})) - rL\right] dt + L.
\end{aligned}
\tag{A.99}
$$

Note that the agent's outside option L factors out of the incentive compatibility condition (A.99). Since for every $t \in [t_0, t_1] \cap support(\tau)$ we have $l_t \geq m_t$, it implies that $l_{t_1} \leq m_{t_1} \leq m_t \leq m_{t_0} \leq l_{t_0}$

for every $t \in [t_0, t_1]$. The expected value of waiting to separate at time t_1 can then be bounded by

$$
\begin{aligned}
&\int_{t_0}^{t_1} e^{-r(s-t_0)} \cdot \left(p + (1-p) \cdot e^{-\lambda(s-t_0)}\right) \cdot [A(\pi_s(m_s)) - rV]\ ds \\
&+ \int_{t_1}^{\infty} e^{-r(s-t_0)} \cdot \left(p + (1-p) \cdot e^{-\lambda(s-t_0)}\right) \cdot A\left(\pi_s(l_{t_1})\right)\ ds \\
&< \int_{t_0}^{t_1} e^{-r(s-t_0)} \cdot \left(p + (1-p) \cdot e^{-\lambda(s-t_0)}\right) \cdot [A(\pi_s(m_{t_0})) - rV]\ ds \\
&+ \int_{t_1}^{\infty} e^{-r(s-t_0)} \left(p + (1-p) \cdot e^{-\lambda(s-t_0)}\right) \cdot A\left(\pi_s(m_{t_1})\right)\ ds \\
&\leq \int_{t_0}^{t_1} e^{-r(s-t_0)} \cdot \left(p + (1-p) \cdot e^{-\lambda(s-t_0)}\right) \cdot [A(\pi_s(m_{t_0})) - rV]\ ds \\
&+ \int_{t_1}^{\infty} e^{-r(s-t_0)} \cdot \left(p + (1-p) \cdot e^{-\lambda(s-t_0)}\right) \cdot A\left(\pi_s(m_{t_0})\right)\ ds \\
&= -\int_{t_0}^{t_1} e^{-r(s-t_0)} \cdot \left(p + (1-p) \cdot e^{-\lambda(s-t_0)}\right) \cdot rV\ ds \\
&+ \int_{t_0}^{\infty} e^{-r(s-t_0)} \cdot \left(p + (1-p) \cdot e^{-\lambda(s-t_0)}\right) \cdot A\left(\pi_s(m_{t_0})\right)\ ds \\
&< -\int_{t_0}^{t_1} e^{-r(s-t_0)} \cdot \left(p + (1-p) \cdot e^{-\lambda(s-t_0)}\right) \cdot rV\ ds \\
&+ \int_{t_0}^{\infty} e^{-r(s-t_0)} \cdot \left(p + (1-p) \cdot e^{-\lambda(s-t_0)}\right) \cdot A\left(\pi_s(l_{t_0})\right)\ ds \\
&\leq \int_{t_0}^{\infty} e^{-r(s-t_0)} \cdot \left(p + (1-p) \cdot e^{-\lambda(s-t_0)}\right) \cdot A\left(\pi_s(l_{t_0})\right)\ ds,
\end{aligned}
$$

which contradicts the optimality of stopping (A.99) at $t_1 > t_0$.

(ii) Suppose that $U(\pi_{t_1}(l_{t_1}), \pi_{t_1}(l_{t_1})) < L$. The incentive compatibility of the agent of ability p to separate from the intermediary at time t_1 versus time t_0 can be written as

$$
\begin{aligned}
&\int_{t_0}^{t_1} e^{-r(s-t_0)} \cdot \left(p + (1-p) \cdot e^{-\lambda(s-t_0)}\right) \cdot [A(\pi_s(m_s)) - rV - rL]\ ds + L \\
&\geq \int_{t_0}^{\infty} e^{-r(s-t_0)} \cdot \left(p + (1-p) \cdot e^{-\lambda(s-t_0)}\right) \cdot [A(\pi_s(l_{t_0})) - rL]\ dt + L,
\end{aligned}
\tag{A.100}
$$

which cannot be satisfied since $l_{t_0} > m_{t_0} \geq m_t$ for all $t \in [t_0, t_1]$ and $rV \geq 0$.

The two cases above argument implies that it cannot be the case that $l_{t_0} > m_{t_0}$. □

Lemma 3 in the Print Appendix implies that for agents that are initially hired, i.e., retained for a positive amount of time, the separation times in (A.96) can be simplified as

$$\mathbb{T}^I(p) = \arg\max_t \Big\{ \mathrm{E}_{\tilde{p}_0=p}\left[\int_0^{t\wedge\eta} e^{-rs}\cdot[A(\pi_s(m_s)) - rV - rL]\,ds + V\right] \\ + \mathrm{E}_{\tilde{p}_0=p}\left[e^{-rt}\cdot(U(\pi_t(p),\pi_t(l_t)) - L)\cdot\mathbb{1}\{t<\eta\}\right] + L\Big\}. \tag{A.101}$$

Note that subtracting L in (A.101) is equivalent to considering an alternative revenue function $\hat{A}(p) \stackrel{def}{=} A(p) - rL$. Consequently, we can, without loss, set $L = 0$ in the analysis below, which was one of the reasons for proving Lemma A.27 and Lemma 3.

Lemma A.28 (Pooling belief properties)**.** *Suppose* $\mathbb{T}^I(p_1)\cap\mathbb{T}^I(p_2)\neq\emptyset$ *for* $p_1<p_2$*. Then*

(i) $\mathbb{T}^I(p') = \mathbb{T}^I(p'')$ *for every* $p', p'' \in (p_1,p_2)$*;*

(ii) $l_{t'} \geq l_{t''}$ *for any* $t' < t'' \in \mathbb{T}^I(p)$ *and* $p\in(p_1,p_2)$*.*

Proof. **Part *(*i*)*.** The expected payoff to the intermediary and the agent pair from stopping at time t is $W_0(p;t)$ defined in (A.98). The maximum expected welfare from choosing the optimal separating time t is $W_0(p) = \sup_t W_0(p;t)$. Since $W_0(p;t)$ is linear in p for each t, it follows that $W_0(p)$ is the lower Envelope of linear functions, implying that it is weakly convex.

By definition $\mathbb{T}^I(p) = \arg\max_t W_0(p,t)$. If $t\in\mathbb{T}^I(p_1)\cap\mathbb{T}^I(p_2)$, then $W_0(p) = W_0(p;t)$ for $p\in\{p_1,p_2\}$. Due to weak convexity of $W_0(p)$, it follows that $W_0(p) = W(p;t)$ for all $p\in[p_1,p_2]$. Since separation at time t achieves the maximum joint welfare for all $p\in[p_1,p_2]$, it follows that $t\in\mathbb{T}^I(p)$. Consequently, $\mathbb{T}^I(p_1)\cap\mathbb{T}^I(p_2)\subseteq\mathbb{T}^I(p)$ for every $p\in[p_1,p_2]$.

Suppose, by contradiction, $\exists\, t\in\mathbb{T}^I(p')\setminus\mathbb{T}^I(p'')$ for $p',p''\in(p_1,p_2)$ and $p'\neq p''$. The optimality of $t\in\mathbb{T}^I(p')$ requires that $W_0(p') = W_0(p',t)$. Since $W_0(p)$ is linear over $p\in[p_1,p_2]$, the slopes must also match, meaning that $W_0'(p) = \partial_1 W_0(p,t)$ – otherwise type $p\neq p'$ could achieve a strict improvement over $W_0(p)$ by choosing to stop at t, which would contradict the definition of $W_0(p)$. Consequently, $t\in\mathbb{T}^I(p'')$ for any $p''\in[p_1,p_2]$ which, in turn, implies that $\mathbb{T}^I(p') = \mathbb{T}^I(p'')$ for every $p',p''\in(p_1,p_2)$.

Part *(ii)*. Suppose by contradiction $l_{t'} < l_{t''}$. For a given belief process l we can extend the definition of the compensation process $C(p) = (C_t(p))_{t\geq 0}$ in (19) to an arbitrary deterministic, bounded, and measurable process l. The martingale argument of Lemma 1 continues to hold for such process l and, consequently, $C(p)$ achieves the lower bound in the retention condition (7). Importantly, the expected compensation costs between times t' and t'' in (7) depend solely on belief levels $l_{t'}$ and $l_{t''}$ as well as the agent's private type and not on the specific path of l_t between t' and t''. Consequently, if $l_{t'} < l_{t''}$, (38) – (41) of Lemma (4) of the Print Appendix show that type $p'' > p'$ would find it strictly profitable to separate at time t'' than at time t', which contradicts the fact that $t'\in\mathbb{T}^I(p'')$. It must then be the case that $l_{t'}\geq l_{t''}$. □

Lemma A.29 (Separating equilibrium)**.** *The equilibrium must be separating in types.*

Proof. Suppose by contradiction $t \in \mathbb{T}^I(p') \cap \mathbb{T}^I(p'')$. It follows from the proof of Lemma A.28 that $W_0(p)$ is linear over $p \in (p', p'')$. Since $W_0(p)$ is convex in p, define by p_1 and p_2 the lowest and highest types who find it optimal to separate at time t:

$$p_1 \overset{def}{=} \inf \left\{ p : \, t \in \mathbb{T}^I(p) \right\}, \qquad p_2 \overset{def}{=} \sup \left\{ p : \, t \in \mathbb{T}^I(p) \right\}. \tag{A.102}$$

It follows from Lemma A.28 that $\mathbb{T}^I(p)$ is the same for all $p \in (p_1, p_2)$. Denote this set of times by $\mathbb{T}^*$.

By construction of the lower belief p_1, for all $\mathbb{T}^I(p) \cap \mathbb{T}^* = \emptyset$ for all $p < p_1$. Additionally, $\mathbb{T}^I(p') \cap \mathbb{T}^I(p'') = \emptyset$ for any $p' < p_1$ and $p'' > p_2$ since, otherwise, there would exist $t \in \mathbb{T}^I(p') \cap \mathbb{T}^I(p'')$ and function $W_0(p)$ would be linear over the wider interval $[p', p'']$, contradicting the definition of p_1 and p_2.

Consider $\hat{t} \overset{def}{=} \sup \left\{ t : \, \cup_{p<p_2} \mathbb{T}^I(p) \right\}$ to be the last time when types $p < p_2$ separate from the intermediary. It follows from part *(ii)* of Lemma (A.28) and upper semi-continuity of l_t that $l_{\hat{t}} < p_2$. At the same time, the Never Dissuaded Once Convinced refinement implies that $l_t \geq p_2$ for $t > \hat{t}$ as all of these types have separated before $\hat{t}$ with certainty. Consequently, there is a jump in beliefs at time $\hat{t}$, which contradicts the definition of $\hat{t}$. This implies that $\mathbb{T}^I(p') \cap \mathbb{T}^I(p'') = \emptyset$ for any $p' \neq p''$ and, consequently, the equilibrium must be separating in types. □

Define $\alpha_t \overset{def}{=} \inf\{\mathit{support}\,(p \,|\, t < \eta, t < \tau)\}$ to be the lowest ex-ante type still in the support of the distribution at time t. In the equilibria constructed in Proposition 1 and 2, the client belief equals the lowest remaining type, i.e., $l_t = \alpha_t$, but this may, generally, not be the case. The Never Dissuaded Once Convinced refinement requires only that $l_t \geq \alpha_t$. By definition, α_t is weakly increasing. Since the equilibrium is separating, as derived in Lemma A.29, it follows that α_t is continuous. Since separations occur in finite time, as shown in Lemma A.26, α_t converges to $\overline{p}$ in finite time.

Lemma A.30 shows that the equilibrium process l must be weakly increasing for all $t \leq \bar{t}$. We do so by showing that there cannot be an interval of types (p_1, p_2) willing to separate over two disjoint intervals.

Lemma A.30 (Weakly increasing l)**.** *The equilibrium process l must be globally weakly increasing.*

Proof. Suppose by contradiction $l_{t'} > l_{t''}$ for some $t' > t''$. Suppose $\mathit{support}\,(\tau)$ is nowhere dense over (t', t''). Then, since l is upper semi-continuous and weakly increasing for $t \notin \mathit{support}\,(\tau)$, it follows that $l_{t'} \leq l_{t''}$. Consequently, there exists (t', t'') such that $\mathit{support}\,(\tau)$ is dense in (t', t'') and $l_{t'} > l_{t''}$. The local optimality condition for $\mathit{support}\,(\tau)$ to be dense in (t', t'') requires that for every $t \in (t', t'')$.

$$A(\pi_t(m_t) - rV + A(\pi(l_t)) + \dot{l}_t \cdot \pi_t'(l_t) \cdot \partial_2 U(\pi_t(l_t), \pi_t(l_t)) = 0, \tag{A.103}$$

In particular, (A.103) and, as we have shown in Lemma A.13 above, require process l to be differentiable for $t \in (t', t'')$. Consequently, since $l_{t'} > l_{t''}$, we can choose (t', t'') such that belief process l is weakly decreasing over (t', t'') and *support*(τ) is dense in (t', t''). Denote $p_2 = l_{t'} > l_{t''} = p_1$.

It must be the case that $\alpha_{t'} = \alpha_{t''}$. Otherwise, since the equilibrium is separating and l_t is decreasing over (t', t''), then $p_1 = l_{t''} \geq \alpha_{t''}$. Consequently, $(p_1, p_2) \cap (0, \alpha_{t''}) = \emptyset$, implying that types $(\alpha_{t'}, \alpha_{t''})$ do not separate during (t', t''). Since $\alpha_{t'} = \alpha_{t''}$, it implies that types (p_1, p_2) who are willing to separate during (t', t'') are also willing to separate after time t''. For each type $p \in (p_1, p_2)$ denote by $t_1(p) \in (t', t'')$ to be the time when type p is willing to separate during (t', t'').

Denote $t_2(p) \stackrel{def}{=} \inf\{t : \alpha_t = p\}$ as the final time when type p separates from the intermediary. The separating belief condition requires that $\alpha_{t_2(p)} = l_{t_2(p)} = p$. By the argument above, $t_1(p) < t_2(p)$ for all $p \in (p_1, p_2)$. It follows from Lemma A.13 that process l_t has to be differentiable at $t_2(p)$. The local optimality condition at $t = t_2(p)$ is equivalent to (A.103) evaluated at $l_t = \alpha_t$.

The local optimality condition (A.103) requires that separating times $t_1(p)$ and $t_2(p)$ are continuous in $p \in (p_1, p_2)$. By differentiating the separating equilibrium identity $l_{t(p)} = p$ with respect to p obtain

$$\dot{l}_{t(p)} \cdot t'(p) = 1 \qquad \Rightarrow \qquad t'(p) = \frac{1}{\dot{l}_{t(p)}} = \frac{\pi'_{t(p)}(p) \cdot \partial_2 U\left(\pi_{t(p)}(p), \pi_{t(p)}(p)\right)}{A\left(\pi_{t(p)}(p)\right) + rV - A\left(\pi_{t(p)}(m_{t(p)})\right)}, \tag{A.104}$$

which holds for both $t(p) = t_1(p)$ and $t(p) = t_2(p)$. Since $t_1(p), t_2(p) \in \mathbb{T}^I(p)$, it follows that for every $p \in (p_1, p_2)$:

$$0 = \int_{t_1(p)}^{t_2(p)} e^{-rt} \cdot \left(p + (1-p) \cdot e^{-\lambda t}\right) \cdot \left[A(\pi_t(m_t)) - rV - A(\pi_t(p))\right] dt. \tag{A.105}$$

Differentiate identity (A.105) with respect to p to obtain

$$\begin{aligned}
0 = & \int_{t_1(p)}^{t_2(p)} e^{-rt} \cdot \left(1 - e^{-\lambda t}\right) \cdot \left[A(\pi_t(m_t)) - rV - A(\pi_t(p))\right] dt \\
& - \int_{t_1(p)}^{t_2(p)} e^{-rt} \cdot \left(p + (1-p) \cdot e^{-\lambda t}\right) \cdot A'(\pi_t(p)) \cdot \pi'_t(p)\, dt \\
& + e^{-rt_2(p)} \cdot \left(p + (1-p) \cdot e^{-\lambda t_2(p)}\right) \cdot \left[A\left(\pi_{t_2(p)}\left(m_{t_2(p)}\right)\right) - rV - A\left(\pi_{t_2(p)}(p)\right)\right] \cdot t_2'(p) \\
& - e^{-rt_1(p)} \cdot \left(p + (1-p) \cdot e^{-\lambda t_1(p)}\right) \cdot \left[A\left(\pi_{t_1(p)}\left(m_{t_1(p)}\right)\right) - rV - A\left(\pi_{t_1(p)}(p)\right)\right] \cdot t_1'(p) \\
= & \int_{t_1(p)}^{t_2(p)} e^{-rt} \cdot \left(1 - e^{-\lambda t}\right) \cdot \left[A(\pi_t(m_t)) - rV - A(\pi_t(p))\right] dt \\
& - \int_{t_1(p)}^{t_2(p)} e^{-rt} \cdot \left(p + (1-p) \cdot e^{-\lambda t}\right) \cdot A'(\pi_t(p)) \cdot \pi'_t(p)\, dt \\
& - e^{-rt_2(p)} \cdot \left(p + (1-p) \cdot e^{-\lambda t_2(p)}\right) \cdot \pi'_{t_2(p)}(p) \cdot \partial_2 U\left(\pi_{t_2(p)}(p), \pi_{t_2(p)}(p)\right) \\
& + e^{-rt_1(p)} \cdot \left(p + (1-p) \cdot e^{-\lambda t_1(p)}\right) \cdot \pi'_{t_1(p)}(p) \cdot \partial_2 U\left(\pi_{t_1(p)}(p), \pi_{t_1(p)}(p)\right),
\end{aligned} \tag{A.106}$$

where the equality in (A.106) follows from (A.104) applied to $t_1(p)$ and $t_2(p)$. Note that

$$\begin{aligned}
&e^{-rt}\cdot\left(p+(1-p)\cdot e^{-\lambda t}\right)\cdot\pi_t'(p)\cdot\partial_2 U(\pi_t(p),\pi_t(p))\\
=&e^{-rt}\cdot\left(p+(1-p)\cdot e^{-\lambda t}\right)\cdot\frac{e^{-\lambda t}}{(p+(1-p)\cdot e^{-\lambda t})^2}\\
\times&\int_0^\infty e^{-rs}\cdot\left(\pi_t(p)+(1-\pi_t(p))\cdot e^{-\lambda s}\right)\cdot A'(\pi_{t+s}(p))\cdot\frac{e^{-\lambda s}}{(\pi_t(p)+(1-\pi_t(p))\cdot e^{-\lambda s})^2}\,ds\\
=&\frac{e^{-\lambda t}}{p+(1-p)e^{-\lambda t}}\cdot\int_0^\infty e^{-r(s+t)}\cdot\frac{p+(1-p)\cdot e^{-\lambda t}\cdot e^{-\lambda s}}{p+(1-p)\cdot e^{-\lambda t}}\cdot\frac{A'(\pi_{t+s}(p))\cdot e^{-\lambda s}}{\left(\frac{p}{p+(1-p)\cdot e^{-\lambda t}}+\frac{(1-p)e^{-\lambda t}}{p+(1-p)\cdot e^{-\lambda t}}\cdot e^{-\lambda s}\right)^2}\,ds\\
=&\int_0^\infty e^{-r(s+t)}\cdot\left(p+(1-p)\cdot e^{-\lambda(s+t)}\right)\cdot\frac{A'(\pi_{t+s}(p))\cdot e^{-\lambda(s+t)}}{(p+(1-p)\cdot e^{-\lambda(s+t)})^2}\,ds\\
=&\int_0^\infty e^{-r(s+t)}\cdot\frac{A'(\pi_{t+s}(p))\cdot e^{-\lambda(s+t)}}{p+(1-p)\cdot e^{-\lambda(s+t)}}\,ds=\int_t^\infty e^{-rs}\cdot\frac{A'(\pi_s(p))\cdot e^{-\lambda s}}{p+(1-p)e^{-\lambda s}}\,ds.
\end{aligned}\tag{A.107}$$

Substituting (A.107) in (A.106) obtain

$$\begin{aligned}
0=&\int_{t_1(p)}^{t_2(p)} e^{-rt}\cdot\left(1-e^{-\lambda t}\right)\cdot\left[A(\pi_t(m_t))-rV-A(\pi_t(p))\right]dt-\int_{t_1(p)}^{t_2(p)} e^{-rt}\frac{A'(\pi_t(p))\cdot e^{-\lambda t}}{p+(1-p)e^{-\lambda t}}\,dt\\
&-\int_{t_2(p)}^{+\infty} e^{-rt}\cdot\frac{A'(\pi_t(p))\cdot e^{-\lambda t}}{p+(1-p)e^{-\lambda t}}\,dt+\int_{t_1(p)}^{+\infty} e^{-rt}\cdot\frac{A'(\pi_t(p))\cdot e^{-\lambda t}}{p+(1-p)e^{-\lambda t}}\,dt\\
=&\int_{t_1(p)}^{t_2(p)} e^{-rt}\cdot\left(1-e^{-\lambda t}\right)\cdot\left[A(\pi_t(m_t))-rV-A(\pi_t(p))\right]dt.
\end{aligned}\tag{A.108}$$

Combining (A.105) and (A.108) obtain

$$\begin{cases}
\displaystyle\int_{t_1(p)}^{t_2(p)} e^{-rt}\cdot\left[A(\pi_t(m_t))-rV-A(\pi_t(p))\right]dt=0,\\
\displaystyle\int_{t_1(p)}^{t_2(p)} e^{-(r+\lambda)t}\cdot\left[A(\pi_t(m_t))-rV-A(\pi_t(p))\right]dt=0.
\end{cases}\tag{A.109}$$

Both equalities (A.109) must hold for all $p\in(p_1,p_2)$. Differentiate the top equality in (A.109) with respect to p to obtain

$$\begin{aligned}
0=&e^{-rt_2(p)}\cdot\left[A\left(\pi_{t_2(p))}(m_{t_2(p)})\right)-rV-A\left(\pi_{t_2(p)}(p)\right)\right]\cdot t_2'(p)\\
&-e^{-rt_1(p)}\cdot\left[A\left(\pi_{t_1(p)}(m_{t_1(p)})\right)-rV-A\left(\pi_{t_1(p)}(p)\right)\right]\cdot t_1'(p)-\int_{t_1(p)}^{t_2(p)} e^{-rt}\cdot A'(\pi_t(p))\cdot\pi_t'(p)\,dt
\end{aligned}$$

$$= -e^{-rt_2(p)} \cdot \pi'_{t_2(p)}(p) \cdot \partial_2 U\left(\pi_{t_2(p)}(p), \pi_{t_2(p)}(p)\right) \tag{A.110}$$

$$+ e^{-rt_1(p)} \cdot \pi'_{t_1(p)}(p) \cdot \partial_2 U\left(\pi_{t_1(p)}(p), \pi_{t_1(p)}(p)\right) - \int_{t_1(p)}^{t_2(p)} e^{-rt} \cdot A'(\pi_t(p)) \cdot \frac{e^{-\lambda t}}{(p+(1-p)e^{-\lambda t})^2}\, dt$$

$$= -\frac{1}{p+(1-p)e^{-\lambda t_2(p)}} \cdot \int_{t_2(p)}^{\infty} e^{-rt} \cdot \frac{A'(\pi_t(p)) \cdot e^{-\lambda t}}{p+(1-p)\cdot e^{-\lambda t}}\, dt \tag{A.111}$$

$$+ \frac{1}{p+(1-p)e^{-\lambda t_1(p)}} \cdot \int_{t_1(p)}^{\infty} e^{-rt} \cdot \frac{A'(\pi_t(p)) \cdot e^{-\lambda t}}{p+(1-p)\cdot e^{-\lambda t}}\, dt - \int_{t_1(p)}^{t_2(p)} e^{-rt} \cdot \frac{A'(\pi_t(p)) \cdot e^{-\lambda t}}{(p+(1-p)\cdot e^{-\lambda t})^2}\, dt$$

$$= \int_{t_1(p)}^{t_2(p)} e^{-(r+\lambda)t} \cdot \frac{A'(\pi_t(p))}{p+(1-p)\cdot e^{-\lambda t}} \cdot \underbrace{\left(\frac{1}{p+(1-p)e^{-\lambda t_1(p)}} - \frac{1}{p+(1-p)e^{-\lambda t}}\right)}_{<0} dt$$

$$+ \underbrace{\left(\frac{1}{p+(1-p)\cdot e^{-\lambda t_1(p)}} - \frac{1}{p+(1-p)e^{-\lambda t_2(p)}}\right)}_{<0} \cdot \int_{t_2(p)}^{\infty} e^{-(r+\lambda)t} \cdot \frac{A'(\pi_t(p))}{p+(1-p)e^{-\lambda t}}\, dt < 0. \tag{A.112}$$

Equality (A.110) follows from (A.104). Equality (A.111) follows from (A.107). Strict inequality (A.112) implies that if (A.109) holds at p, then it cannot hold at $p' \neq p$, thus contradicting the optimality of stopping at times $t_1(p)$, $t_2(p)$ for ex-ante types $p \in (p_1, p_2)$. This leads to a contradiction with $l_{t'} > l_{t''}$ for some $t' < t''$. □

Lemma A.30 establishes that the equilibrium process l must be weakly increasing globally. We can now rely on the proof of Observation 2 to show that the equilibrium constructed in Proposition 2 is unique.

A.2 Extension: Dynamic Contracts between Intermediary and Agent

In what follows, we show that the equilibrium constructed in Proposition 2 remains unchanged if the intermediary and the agent can sign long-term contracts. The intuition is that, so long as the agent is weakly less patient than the intermediary, the latter wishes to keep the agent at his reservation utility $U(p, k)$. This is exactly the same promised utility dynamics that arise in the absence of commitment, and consequently, the equilibrium is unaffected by the incremental ability of the intermediary to commit. This, in a way, can be expected as termination in our model is efficient given client beliefs.

To simplify the derivations below, we assume, without loss, that $L = 0$. We do, however, allow for the possibility of the agent's discount rate to be different from the intermediary, highlighting that the equal discount rate assumption in the main text of the paper is, to a large extent, without loss.

Observation A.2 (Equilibrium with Dynamic Contracts). *Denote by ρ and r the discount rate of the agent and the intermediary, respectively. The equilibrium belief process $(q_t, k_t)_{t\geq 0}$ is the same whether or not the intermediary can offer long-term contracts. In equilibrium, if $\rho > r$, then the agent's continuation* *value under the optimal contract is equal to* $U(\tilde{p}_t, k_t)$ *after every history under the unique optimal contract; if* $\rho = r$*, then the contract in which the agent's continuation value is equal to* $U(\tilde{p}_t, k_t)$ *after every history constitutes an optimal contract.*

Proof. For a given ex-ante type $\tilde{p}_0$ consider a candidate dynamic contract $\hat{\mathcal{C}} = (\hat{C}, \hat{\tau})$ in which the intermediary commits to a long-term cumulative compensation process $(\hat{C}_t)_{t\geq 0}$ and termination date $\hat{\tau}$. The agent's continuation value $\hat{W}_t$ from staying with the intermediary until time $\hat{\tau}$ under contract $\hat{\mathcal{C}}$ is given by

$$\hat{W}_t \stackrel{def}{=} \mathrm{E}_{\tilde{p}_t}\left[\int_t^{\hat{\tau}\wedge\eta} e^{-\rho(s-t)}\, d\hat{C}_s + e^{-\rho(\hat{\tau}\wedge\eta-t)} \cdot U\left(\tilde{p}_{\hat{\tau}}, \hat{k}_{\hat{\tau}}\right)\right],$$

where $\eta \stackrel{def}{=} \inf\{t : X_t = t-1\}$. The agent is retained by the intermediary as long as $\hat{W}_t \geq U(\tilde{p}_t, \hat{k}_t)$ for all $t \leq \hat{\tau}$ with the equality being strict at $t = \hat{\tau} < \eta$. Note that due to the agent's risk neutrality, the intermediary can offer retention incentives conditional on the path of good performance $X_t = t$. Consequently, it is without loss that the agent is let go after generating poor performance and gets paid nothing in that event. It is convenient to define conditional continuation values

$$\hat{W}_t^\theta \stackrel{def}{=} \mathrm{E}_{\tilde{p}_t}\left[\int_t^{\hat{\tau}\wedge\eta} e^{-\rho(s-t)}\, d\hat{C}_s + e^{-\rho(\hat{\tau}\wedge\eta-t)} \cdot U\left(\tilde{p}_{\hat{\tau}}, \hat{k}_{\hat{\tau}}\right) \Big| \theta\right].$$

It follows that $\hat{W}_t = \tilde{p}_t \cdot \hat{W}_t^1 + (1-\tilde{p}_t) \cdot \hat{W}_t^0$. Ito's lemma then implies that

$$d\hat{C}_t - \rho \hat{W}_t^\theta\, dt + \left(L - \hat{W}_t^\theta\right) dN_t^\theta + d\hat{W}_t^\theta = 0.$$

For a given type $\tilde{p}_0$ and termination time $\hat{\tau}$ define a cumulative compensation process $C = (C_t)_{t\geq 0}$ such that the agent's continuation value from working until time $\hat{\tau}$ is exactly equal to his outside option $U(\tilde{p}_t, \hat{k}_t)$. Formally, process C solves

$$U\left(\tilde{p}_t, \hat{k}_t\right) = \mathrm{E}_{\tilde{p}_t}\left[\int_t^{\hat{\tau}\wedge\eta} e^{-\rho(s-t)}\, dC_s + e^{-\rho(\hat{\tau}\wedge\eta-t)} \cdot U\left(\tilde{p}_{\hat{\tau}}, \hat{k}_{\hat{\tau}}\right)\right]. \tag{A.113}$$

Given compensation process C defined in (A.113), it is convenient to define the resulting agent's continuation value conditional on θ as

$$U^\theta\left(\tilde{p}_t, \hat{k}_t\right) = \mathrm{E}_{\tilde{p}_t}\left[\int_t^{\hat{\tau}\wedge\eta} e^{-\rho(s-t)}\, dC_s + e^{-\rho(\hat{\tau}\wedge\eta-t)} \cdot U\left(\tilde{p}_{\hat{\tau}}, \hat{k}_{\hat{\tau}}\right) | \, \theta\right].$$

By definition $U(p_t, \hat{k}_t) = \tilde{p}_t \cdot U^1(\tilde{p}_t, \hat{k}_t) + (1-\tilde{p}_t) \cdot U^0(\tilde{p}_t, \hat{k}_t)$. It follows from Ito's lemma that

$$dC_t - \rho U^\theta\left(\tilde{p}_t, \hat{k}_t\right) dt + \left(L - U\left(\tilde{p}_t, \hat{k}_t\right)\right) dN_t^\theta + dU^\theta\left(\tilde{p}_t, \hat{k}_t\right) = 0.$$

Suppose under the optimal contract $\hat{\mathcal{C}}$ there exists a time t such that the agent's continuation value strictly exceeds his outside option, i.e., $\hat{W}_{\hat{t}} > U\left(\tilde{p}_{\hat{t}}, \hat{k}_{\hat{t}}\right)$. Consider an alternative contract $\tilde{\mathcal{C}} = (\tilde{C}, \hat{\tau})$ in which the intermediary lets the agent go at the same time $\hat{\tau}$ along the path of good performance but switches to a cumulative compensation profile

$$\tilde{C}_s = \begin{cases} \hat{C}_s & \textit{for} \quad s < t, \\ \hat{W}_s - U\left(\tilde{p}_t, \hat{k}_t\right) + C_s - C_t & \textit{for} \quad s \geq t. \end{cases} \tag{A.114}$$

By construction, contract $\tilde{\mathcal{C}}$ satisfies the agent's retention constraint up until time $\hat{\tau}$. The net continuation benefit at time t to the intermediary of the new contract relative to the original contract is

$$\begin{aligned}
&\overbrace{\mathrm{E}\left[\int_t^{\hat{\tau}\wedge\eta} e^{-r(s-t)}\, d\hat{C}_s\right]}^{\textit{costs of contract } \hat{\mathcal{C}}} - \overbrace{\left[\hat{W}_t - U\left(\tilde{p}_t, \hat{k}_t\right) + \mathrm{E}\left[\int_t^{\hat{\tau}\wedge\eta} e^{-r(s-t)}\, dC_s\right]\right]}^{\textit{costs of contract } \tilde{\mathcal{C}}} \\
=&\mathrm{E}\left[\int_t^{\hat{\tau}\wedge\eta} e^{-r(s-t)}\, d\hat{C}_s\right] - \mathrm{E}\left[\int_t^{\hat{\tau}\wedge\eta} e^{-\rho(s-t)}\, d\hat{C}_s\right] + \mathrm{E}\left[\int_t^{\hat{\tau}\wedge\eta} e^{-\rho(s-t)}\, dC_s\right] - \mathrm{E}\left[\int_t^{\hat{\tau}\wedge\eta} e^{-r(s-t)}\, dC_s\right] \\
=&\mathrm{E}\left[\int_t^{\hat{\tau}\wedge\eta} \left(e^{-\rho(s-t)} - e^{-r(s-t)}\right) \cdot \left(dC_s - d\hat{C}_s\right)\right] \\
=&\tilde{p}_t \cdot \mathrm{E}\left[\int_t^{\hat{\tau}\wedge\eta} \left(e^{-\rho(s-t)} - e^{-r(s-t)}\right) \cdot \left(dC_s - d\hat{C}_s\right) \;\middle|\; \theta = 1\right] \\
&+(1-\tilde{p}_t) \cdot \mathrm{E}\left[\int_t^{\hat{\tau}\wedge\eta} \left(e^{-\rho(s-t)} - e^{-r(s-t)}\right) \cdot \left(dC_s - d\hat{C}_s\right) \;\middle|\; \theta = 0\right] \\
=&\tilde{p}_t \cdot \int_t^{\hat{\tau}} \left(e^{-\rho(s-t)} - e^{-r(s-t)}\right) \cdot \left(dC_s - d\hat{C}_s\right) \\
&+(1-\tilde{p}_t) \cdot \int_t^{\hat{\tau}} e^{-\lambda(s-t)} \cdot \left(e^{-\rho(s-t)} - e^{-r(s-t)}\right) \cdot \left(dC_s - d\hat{C}_s\right) \\
=&\tilde{p}_t \cdot \int_t^{\hat{\tau}} \left(e^{-\rho(s-t)} - e^{-r(s-t)}\right) \cdot \left[\rho\left(U^1\left(p_t, \hat{k}_t\right) - \hat{W}_t^1\right) - dU^1\left(p_t, \hat{k}_t\right) + d\hat{W}_t^1\right] dt \\
&+(1-\tilde{p}_t) \cdot \int_t^{\hat{\tau}} e^{-\lambda(s-t)} \cdot \left(e^{-\rho(s-t)} - e^{-r(s-t)}\right) \cdot \left[(\rho+\lambda)\left(U^0\left(p_t, \hat{k}_t\right) - \hat{W}_t^0\right) - dU^0\left(p_t, \hat{k}_t\right) + d\hat{W}_t^0\right] dt \\
=&\tilde{p}_t \cdot \int_t^{\hat{\tau}} \left(1 - e^{(\rho-r)(s-t)}\right) de^{-\rho(s-t)} \left(\hat{W}_s^1 - U^1\left(\tilde{p}_s, \hat{k}_s\right)\right) \\
&+(1-\tilde{p}_t) \cdot \int_t^{\hat{\tau}} \left(1 - e^{(\rho-r)(s-t)}\right) de^{-(\rho+\lambda)(s-t)} \left(\hat{W}_s^0 - U^0\left(\tilde{p}_s, \hat{k}_s\right)\right) \\
=&\tilde{p}_t \cdot \underbrace{\left(1 - e^{(\rho-r)(s-t)}\right) \cdot e^{-\rho(s-t)} \cdot \left(\hat{W}_s^1 - U^1\left(\tilde{p}_s, \hat{k}_s\right)\right)\Big|_{s=t}^{s=\hat{\tau}}}_{=0}
\end{aligned}$$

$$
\begin{aligned}
&-\tilde{p}_t \cdot \int_t^{\hat{\tau}} e^{-\rho(s-t)} \cdot \left(\hat{W}_s^1 - U^1\left(\tilde{p}_s, \hat{k}_s\right)\right) \, d\left(1 - e^{(\rho-r)(s-t)}\right) \\
&+(1-\tilde{p}_t) \cdot \underbrace{\left(1 - e^{(\rho-r)(s-t)}\right) \cdot e^{-(\rho+\lambda)(s-t)} \cdot \left(\hat{W}_s^0 - U^0\left(\tilde{p}_s, \hat{k}_s\right)\right) \Big|_{s=t}^{s=\hat{\tau}}}_{=0} \\
&-(1-\tilde{p}_t) \cdot \int_t^{\hat{\tau}} e^{-(\rho+\lambda)(s-t)} \cdot \left(\hat{W}_s^0 - U^0\left(\tilde{p}_s, \hat{k}_s\right)\right) \, d\left(1 - e^{(\rho-r)(s-t)}\right) \\
&= -\tilde{p}_t \cdot \int_t^{\hat{\tau}} e^{-\rho(s-t)} \cdot \left(\hat{W}_s^1 - U^1\left(\tilde{p}_s, \hat{k}_s\right)\right) \, d\left(1 - e^{(\rho-r)(s-t)}\right) \\
&-(1-\tilde{p}_t) \cdot \int_t^{\hat{\tau}} e^{-(\rho+\lambda)(s-t)} \cdot \left(\hat{W}_s^0 - U^0\left(\tilde{p}_s, \hat{k}_s\right)\right) \, d\left(1 - e^{(\rho-r)(s-t)}\right) \\
&=(\rho - r) \cdot \int_t^{\hat{\tau}} e^{-\rho(s-t)} \cdot \tilde{p}_t \cdot \left(\hat{W}_s^1 - U^1\left(\tilde{p}_s, \hat{k}_s\right)\right) \, de^{(\rho-r)(s-t)} \\
&+(\rho - r) \cdot \int_t^{\hat{\tau}} e^{-\rho(s-t)} \cdot (1-\tilde{p}_t) \cdot e^{-\lambda(s-t)} \cdot \left(\hat{W}_s^0 - U^0\left(\tilde{p}_s, \hat{k}_s\right)\right) \, de^{(\rho-r)(s-t)} \\
&=(\rho - r) \cdot \int_t^{\hat{\tau}} e^{-\rho(s-t)} \cdot \left(\tilde{p}_t + (1-\tilde{p}_t) \cdot e^{-\lambda(s-t)}\right) \cdot \tilde{p}_s \cdot \left(\hat{W}_s^1 - U^1\left(\tilde{p}_s, \hat{k}_s\right)\right) \, de^{(\rho-r)(s-t)} \\
&+(\rho - r) \cdot \int_t^{\hat{\tau}} e^{-\rho(s-t)} \cdot \left(\tilde{p}_t + (1-\tilde{p}_t) \cdot e^{-\lambda(s-t)}\right) \cdot (1-\tilde{p}_s) \cdot \left(\hat{W}_s^0 - U^0\left(\tilde{p}_s, \hat{k}_s\right)\right) \, de^{(\rho-r)(s-t)} \\
&=(\rho - r) \cdot \int_t^{\hat{\tau}} e^{-\rho(s-t)} \cdot \left(\tilde{p}_t + (1-\tilde{p}_t) \cdot e^{-\lambda(s-t)}\right) \cdot \left(\hat{W}_s - U\left(\tilde{p}_s, \hat{k}_s\right)\right) \, de^{(\rho-r)(s-t)} \overset{(i)}{\geq} 0.
\end{aligned}
$$

where inequality (i) is strict for $\rho > r$ unless $\hat{W}_s = U(\tilde{p}_s, \hat{k}_s)$ for every $s \in [t, \hat{\tau}]$. This implies that under the optimal contract $\hat{\mathcal{C}}$ it must be the case that $\hat{W}_t = U(\tilde{p}_t, \hat{k}_t)$. If $\rho = r$, then the transition from contract $\hat{\mathcal{C}}$ to contract $\tilde{\mathcal{C}}$ is costless and hence contract $\tilde{\mathcal{C}}$ is optimal whenever contract $\hat{\mathcal{C}}$ was optimal. This proves the first part of the observation.

Equilibrium equivalence. We have just established that for a given belief process $(\hat{q}_t, \hat{k}_t)_{t \geq 0}$ the optimal long-term contract keeps the agent at his reservation wage $U(\tilde{p}_t, k_t)$ after every history. This implies that when the intermediary chooses time τ when to let the agent go, she faces the same retention costs with or without commitment. For this reason, the optimal termination time τ has to be subgame perfect – if it was not, then the intermediary could extend the agent's employment ex-post without violating the promise-keeping constraint but making a positive profit. Consequently, the equilibrium with commitment has to result in the same churning policy. Under the assumptions of Observation 2, the equilibrium belief process has to be equal to (q_t, k_t) derived in Propositions 1 and 2. □

A.3 Extension: Imperfect Competition for Agents

The intermediary can underpay the agent since his outside option is to either open a firm but be perceived as a lower type or leave the industry altogether. Suppose, if the agent leaves the current intermediary, he can find one new intermediary to work for with probability ζ_1, and finds two or more intermediaries to work for with probability ζ_2.[5] For tractability, we assume the intermediary at the new firm has the same information $\tilde{p}$ about the agent's ability. The possibility of finding multiple intermediaries is beneficial to the agent as they compete for his services by offering a signing bonus.[6] With probability $1-\zeta_1-\zeta_2$, the agent does not find a new intermediary and either starts his own firm or leaves the industry. The possibility of not finding a new intermediary and having to start his firm prematurely poses a risk for the agent if he chooses to leave the incumbent intermediary.

Observation A.3. *The equilibrium quiet periods, churning periods, and client belief processes* $(k_t, q_t)_{t\geq 0}$ *are unchanged. The agent's compensation depends on his endogenous bargaining power* $\zeta \overset{def}{=} \zeta_2/(\zeta_1+\zeta_2)$ *and is given by*

$$w_t^{\zeta}(\tilde{p}_t) = \underbrace{(1-\zeta)\cdot w_R(\tilde{p}_t,k_t) + \zeta\cdot (A(q_t) - rV)}_{reservation\ wage\ with\ bargaining\ power\ \zeta} - \underbrace{(1-\zeta)\cdot \gamma_t \cdot \partial_2 U(\tilde{p}_t,k_t)}_{pay\ for\ reputation}. \tag{A.115}$$

The intermediary's expected value from employing the agent decreases in ζ.

Proof. **Agent's outside option.** Consider the equilibrium wage $\tilde{w}_t^{\zeta}$ received by the agent of ability $\tilde{p}_t$ at time t along the path of good performance. Denote the intermediary's expected value of employing type $\tilde{p}$ while the belief is k as

$$V_t^{\zeta}(\tilde{p}_0) \overset{def}{=} \mathrm{E}_{\tilde{p}_0}\left[\int_t^{\tau\wedge\eta} e^{-r(s-t)}\cdot\left(A(q_s) - \tilde{w}_s^{\zeta}\right)\, ds + e^{-r\tau\wedge\eta}\cdot V \,\middle|\, X_t = t\right]. \tag{A.116}$$

Denote by $U_t^{\zeta}(\tilde{p})$ the expected value of the agent of ex-ante ability $\tilde{p}_0$ at time t given his bargaining power ζ defined as $\zeta = \frac{\zeta_2}{1-\zeta_1}$. If the agent leaves the intermediary, then with probability ζ_0 he does not find an alternative intermediary, in which case he opens his own firm and collects $U_t(\tilde{p}_0, l_t) \overset{def}{=} U^*(\pi_t(\tilde{p}_0), \pi_t(l_t)) = \max\{U(\pi_t(\tilde{p}_0), \pi_t(l_t)),\, L\}$, where $U(\tilde{p}, k)$ is defined in (5). With probability ζ_1 the agent finds only one intermediary, in which case his continuation value remains $U_t^{\zeta}(\tilde{p}_0)$ as the new intermediary replaces the previous one. With probability ζ_2 the agent finds multiple intermediaries who compete for his services a-la Bertrand via a sign-on bonus and, consequently, the agent collects both his outside option $U_t^{\zeta}(\tilde{p}_0)$ as well as the intermediary's value-added $V_t^{\zeta}(\tilde{p}_0) - V$. This implies that, in equilibrium, the agent's expected

[5] If the agent approached only one intermediary at a time, the intermediary would retain all of the bargaining power, as shown in Diamond (1971). The main model corresponds to the case of $\zeta_2 = 0$.

[6] In the absence of commitment to long-term contracts, multiple intermediaries compete a-la Bertrand by offering an up-front payment to the agent but, subsequently, keeping the agent at his outside option.

value in each period must satisfy

$$\begin{aligned}U_t^\zeta(\tilde p_0) &= \zeta_0\cdot U_t(\tilde p_0,l_t)+\zeta_1\cdot U_t^\zeta(\tilde p_0)+\zeta_2\cdot\Big(U_t^\zeta(\tilde p_0)+V_t^\zeta(\tilde p_0)-V\Big),\\ (1-\zeta_1)\cdot U_t^\zeta(\tilde p_0)-\zeta_0\cdot U_t(\tilde p_0,l_t) &= \zeta_2\cdot\Big(U_t^\zeta(\tilde p_0)+V_t^\zeta(\tilde p_0)-V\Big),\\ (1-\zeta_1-\zeta_2)\cdot\Big(U_t^\zeta(\tilde p_0)-U_t(\tilde p_0,l_t)\Big) &= \zeta_2\cdot\Big(V_t^\zeta(\tilde p_0)-V\Big),\\ U_t^\zeta(\tilde p_0)-U_t(\tilde p_0,l_t) &= \frac{\zeta}{1-\zeta}\cdot\Big(V_t^\zeta(\tilde p_0)-V\Big). \end{aligned}\tag{A.117}$$

Agent's equilibrium wage. Given belief processes $(k_t,q_t)_{t\geq 0}$ the value of the intermediary from employing the agent of skill $\tilde p_t$ is

$$rV_t^\zeta(\tilde p_0)=A(q_t)-\tilde w_t^\zeta+\lambda(1-\pi_t(\tilde p_0))\cdot\Big(V-V_t^\zeta(\tilde p_0)\Big)+\dot V_t(\tilde p_0).\tag{A.118}$$

Given wage process $\tilde w_t$ the equilibrium payoff of the agent is given by $U_t^\zeta(\tilde p_0)$

$$rU_t^\zeta(\tilde p_0)=\tilde w_t^\zeta+\lambda(1-\pi_t(\tilde p_0))\cdot\Big(L-U_t^\zeta(\tilde p_0)\Big)+\dot U_t^\zeta(\tilde p_0).\tag{A.119}$$

Denote by γ_t^ζ is the endogenous churning rate expressed in the posterior type k_t given the agent's bargaining power ζ. Churning rate γ_t^ζ in the posterior type space $k=(k_t)_{t\geq0}$ maps to a churning rate $\gamma_t^\zeta/\pi_t'(l_t)$ in the ex-ante type space $l=(l_t)_{t\geq0}$. It follows from (19) that agent's outside option $U_t(\tilde p_0,l_t)$ satisfies

$$rU_t(\tilde p_0,l_t)=w_R(\pi_t(\tilde p_0),\pi_t(l_t))-\frac{\gamma_t^\zeta}{\pi_t'(l_t)}\cdot\partial_2U_t(\tilde p_0,l_t)+\dot U_t(\tilde p_0,l_t).\tag{A.120}$$

Subtracting (A.120) from (A.119) to obtain

$$\begin{aligned}r\cdot\Big(U_t^\zeta(\tilde p_0)-U_t(\tilde p_0,l_t)\Big) &= \tilde w_t^\zeta-w_R(\pi_t(\tilde p_0),\pi_t(l_t))+\frac{\gamma_t^\zeta}{\pi_t'(l_t)}\cdot\partial_2U_t(\tilde p_0,l_t)\\ &\quad+\lambda(1-\pi_t(\tilde p_0))\cdot\Big(U_t(\tilde p_0,l_t)-U_t^\zeta(\tilde p_0)\Big)+\dot U_t^\zeta(\tilde p_0)-\dot U_t(\tilde p_0,l_t).\end{aligned}\tag{A.121}$$

Substitute (A.117) into (A.121) to obtain

$$\begin{aligned}r\cdot\frac{\zeta}{1-\zeta}\cdot\Big(V_t^\zeta(\tilde p_0)-V\Big) &= \tilde w_t^\zeta-w_R(\pi_t(\tilde p_0),\pi_t(l_t))+\frac{\gamma_t^\zeta}{\pi_t'(l_t)}\cdot\partial_2U_t(\tilde p_0,l_t)\\ &\quad+\lambda(1-\pi_t(\tilde p_0))\cdot\frac{\zeta}{1-\zeta}\cdot\Big(V-V_t^\zeta(\tilde p_0)\Big)+\frac{\zeta}{1-\zeta}\cdot\dot V_t^\zeta(\tilde p_0).\end{aligned}\tag{A.122}$$

Multiply (A.118) by $\frac{\zeta}{1-\zeta}$ and subtract from (A.122) to obtain

$$\frac{\zeta}{1-\zeta}\Big[r\Big(V_t^\zeta(\tilde p_0)-V\Big)-rV_t^\zeta(\tilde p_0)\Big]=\tilde w_t^\zeta-w_R(\pi_t(\tilde p_0),\pi_t(l_t))+\frac{\gamma_t^\zeta}{\pi_t'(l_t)}\cdot\partial_2U_t(\tilde p_0,l_t)-\frac{\zeta}{1-\zeta}\Big(A(q_t)-\tilde w_t^\zeta\Big)$$

$$= \frac{\tilde{w}_t^\zeta}{1-\zeta} - w_R(\pi_t(\tilde{p}_0), \pi_t(l_t)) + \frac{\gamma_t^\zeta}{\pi_t'(l_t)} \partial_2 U_t(\tilde{p}_0, l_t) - \frac{\zeta}{1-\zeta} A(q_t).$$

Solving the above equality for $\tilde{w}_t^\zeta$ obtain

$$\tilde{w}_t^\zeta = \zeta \cdot (A(q_t) - rV) + (1-\zeta) \cdot \left(w_R(\pi_t(\tilde{p}_0), \pi_t(l_t)) - \frac{\gamma_t^\zeta}{\pi_t'(l_t)} \partial_2 U_t(\tilde{p}_0, l_t) \right). \tag{A.123}$$

Equality (A.115) follows from (A.123) once we substitute the ex-ante types $(\tilde{p}_0, l_t)$ and time t with corresponding ex-post types $(\tilde{p}_t, k_t)$.

Intermediary equilibrium. The intermediary's expected value from the cutoff agent at the time of turnover is equal to her outside option V, i.e., $V_\tau^\zeta(l_\tau) = V$. The intermediary's profit wedge is equal to her opportunity cost when she lets go of the cutoff agent at time τ, which we can express as

$$A(q_\tau) - \tilde{w}_t^\zeta = rV, \tag{A.124}$$

$$A(q_\tau) - \zeta \cdot (A(q_\tau) - rV) - (1-\zeta) \cdot \left(w_R(k_\tau, k_\tau) - \gamma_t^\zeta \cdot \partial_2 U(k_\tau, k_\tau) \right) = rV,$$

$$A(q_\tau) - rV - w_R(k_\tau, k_\tau) + \gamma_t^\zeta \cdot \partial_2 U\,(k_\tau, k_\tau)) = 0,$$

$$\frac{rV + w_R(k_t, k_t) - A(q_t)}{\partial_2 U(\tilde{p}_t, k_t)} = \gamma_t^\zeta, \tag{A.125}$$

for every $t \in \mathit{support}(\tau)$. For a given equilibrium value of V, the churning rate γ^ζ obtained in (A.125) coincides with the churning rate γ obtained in (12). Belief process $k = (k_t)_{t \geq 0}$ thus does not depend on the agent's bargaining power ζ given the intermediary's expected value V.

By combining the intermediary's objective (6) with the agent's retention constraint (7) we see that similar to the efficient turnover literature, the equilibrium turnover time τ maximizes the joint payoff of the intermediary-agent pair given client belief process (q, k) and is, thus, unaffected by the agent's bargaining power ζ which determines how they share the surplus. If the agent's bargaining power ζ is higher, he receives a greater share of the total surplus, thus reducing the intermediary's expected value. □

A.4 Extension: Paying for Reputation versus Paying for Training

Agents are willing to accept below reservation wages to establish a reputation with clients. While such pay-for-reputation intuition is reminiscent of the pay-for-training logic of Becker (1962), improvements to the agent's reputation lead him to generate greater revenues without affecting his underlying ability. We show that building a reputation via the intermediary's selective retention differs from the intermediary's training, as churning occurs later in the agent's career while training occurs earlier and is accompanied by distinct reputational benefits.

Suppose that in each period, the intermediary can spend a private cost $c(a)$ to transform an unskilled ($\theta = 0$) agent to a skilled ($\theta = 1$) agent with intensity $a \geq 0$. Denote by $a_t(\tilde{p}_t)$ to be the training offered by the intermediary to an agent with skill $\tilde{p}_t$ at time t. The agent's posterior type $\tilde{p}_t$ then evolves based on both his performance and training

$$d\tilde{p}_t = \underbrace{\lambda \tilde{p}_t(1-\tilde{p}_t)\,dt + \tilde{p}_t \cdot (dX_t - dt)}_{\textit{learning from performance}} + \underbrace{a_t(\tilde{p}_t)(1-\tilde{p}_t)\,dt.}_{\textit{training}} \tag{A.126}$$

The agent's private belief $\tilde{p}_t$ increases if he performs well due to learning from performance and further increases if he is trained due to fundamental improvements to his ability. Clients, however, do not directly observe the amount of training provided to the agent, similar to Acemoglu and Pischke (1998), and must form beliefs about it. This affects the agent's reputation as the clients' belief about the worst remaining agent k_t now increases not only due to learning from performance signals and the intermediary's churning as can be seen in (9), but also due to the intermediary's training as captured by the $a_t(\tilde{p}_t) \cdot (1-\tilde{p}_t)\,dt$ term in (A.126) evaluated at the cutoff type $\tilde{p}_t = k_t$. Surprisingly, training not only improves the ability of lower-skilled agents but also the reputation of every retained agent by increasing the ability of the cutoff agent k_t and compressing the residual information asymmetry about retained types $\tilde{p}_t \geq k_t$.

Lemma A.31 (Wages with training)**.** *The agent's equilibrium wage reflects his incentives to pay for his training as well as the reputational gains stemming both from the intermediary's training and churning:*

$$w_t(\tilde{p}_t) = A(k_t) - \underbrace{a_t(\tilde{p}_t)(1-\tilde{p}_t) \cdot \partial_1 U(\tilde{p}_t, k_t)}_{\textit{pay-for-training}} - \underbrace{a_t(k_t)(1-k_t) \cdot \partial_2 U(\tilde{p}_t, k_t))}_{\substack{\textit{pay-for-reputation} \\ \textit{from training}}} - \underbrace{\gamma_t \cdot \partial_2 U(\tilde{p}_t, k_t)}_{\substack{\textit{pay-for-reputation} \\ \textit{from churning}}}. \tag{A.127}$$

Proof. The agent's outside option $U^*(p,k)$ satisfies (18). The agent's private belief process $\tilde{p} = (\tilde{p}_t)_{t \geq 0}$ follows (A.126). The belief process $k = (k_t)_{t \geq 0}$ about the cutoff type follows

$$\dot{k}_t = \lambda k_t(1-k_t) + a_t(k_t) \cdot (1-k_t) + \gamma_t, \tag{A.128}$$

where $\gamma = (\gamma_t)_{t \geq 0}$ is the weakly positive churning rate and $a_t(k_t)$ is the amount of training provided to the cutoff type k_t at time t. Note that given the possibility of training, the intermediary's payoffs differ from that of Section 3.1. Given belief processes $\tilde{p}$ and k given by (A.126) and (A.128), in order to keep the agent at his reservation value $U(\tilde{p}_t, k_t)$ after every history, his wage $w_t^T(\tilde{p})$ must satisfy

$$\begin{aligned} rU(\tilde{p}_t, k_t) &= w_t^T(\tilde{p}_t) + \Big[\lambda \tilde{p}_t(1-\tilde{p}_t) + a(\tilde{p}_t, k_t) \cdot (1-\tilde{p}_t)\Big] \cdot \partial_1 U(\tilde{p}_t, k_t) \\ &\quad + \Big[\lambda k_t(1-k_t) + a_t(k_t) \cdot (1-k_t) + \gamma_t\Big] \cdot \partial_2 U(\tilde{p}_t, k_t) + \lambda(1-\tilde{p}_t) \cdot (L - U(\tilde{p}_t, k_t)). \end{aligned} \tag{A.129}$$

Subtracting (18) from (A.129) we obtain (A.127). ∎

A higher-skilled agent pays less for improvements to his ability as a result of training since for high $\tilde{p}_t$, his ability responds less to training as can be seen from the second term in (A.126). At the same time, a higher-skilled agent puts a higher value on reputational growth, as shown in (11), and is thus willing to pay more for the reputational benefits associated with both churning and training of lower-skilled agents.[7]

The intermediary's incentive to invest in general training, as pointed out by Acemoglu and Pischke (1999), is determined by the extent to which the intermediary can capture the value added to this training which, in turn, depends on the frictions in the agent's labor market. Denote by $V_t(\tilde{p}_t)$ the continuation value of the intermediary from employing an agent $\tilde{p}_t$. The amount of training is, then, pinned down by equating the marginal cost of training to the marginal value of the underlying ability of both the agent and the intermediary

$$c'\big(a_t(\tilde{p}_t)\big) = \underbrace{(1-\tilde{p}_t)\cdot\partial_1 U(\tilde{p}_t,k_t)}_{\textit{agent's gain from training}} \quad + \quad \underbrace{(1-\tilde{p}_t)\cdot\partial_1 V_t(\tilde{p}_t).}_{\textit{intermediary's gain from training}} \tag{A.130}$$

The agent is willing to pay for his training in (A.130) to the extent that he can capture some of the surplus, which is limited by the asymmetric information difference between $\tilde{p}_t$ and k_t. The intermediary values higher-skilled agents more since they are more likely to generate good performance, be retained for longer, and pay for reputation. This implies that for the same degree of asymmetric information, she has a greater incentive in training a higher-skilled agent, captured by $\partial_1 V_t(\tilde{p}_t)$ increasing in $\tilde{p}_t$, however as higher-skilled agents are also less likely to benefit from training, this term is further multiplied in (A.130) by the probability $1-\tilde{p}_t$ that training will be beneficial. The intermediary's incentive to train the agent, however, declines as her information advantage diminishes.

Observation A.4 (Training incentives)**.** *Suppose the public performance signal is not very informative, as captured by $r \geq \lambda$. Then, a higher-skilled agent has a longer career with the intermediary: he pays more for building a reputation but less for training. The intermediary's incentive to train the agent declines as he is about to be let go.*

Proof. The expected value to the intermediary from employing the agent of ability $\tilde{p}_t$ is to choose a combination of training $a_t(\tilde{p}_t)$ and termination time τ specific to this agent. Holding client beliefs $(q,k)_{t\geq 0}$ fixed, the choice of effort provision and termination time τ is a decision problem of the intermediary. Consider the intermediary's value function $\hat{V}_t(\tilde{p}_0)$ that corresponds to offering the same training $a_t(k_t)$ to all retained agents and then choosing turnover time τ optimally, while keeping client belief processes $(q_t,k_t)_{t\geq 0}$ unchanged. Value function $\hat{V}_t(\tilde{p}_0)$ is a lower bound to the intermediary's more flexible effort

[7] Acemoglu and Pischke (1998) consider a model in which higher-skilled agents value training more than lower-skilled agents. Such complementarity can be easily embedded in (A.126) and makes our results even stronger by giving the higher-skilled agents an additional reason to work for the intermediary. While Acemoglu and Pischke (1998) consider a static model, they emphasize in Section III.D the importance of dynamics in future work. Due to dynamic considerations, wage compression is present in our model, even if skill and training are substitutes.

choice problem. We show below that $\hat{V}_t(\tilde{p}_0)$ is strictly increasing in the agent's private type $\tilde{p}_t$. This, in turn, implies that if the intermediary finds it optimal to let go of the cutoff agent $\tilde{p}_t = k_t$, retaining all agents of ability $\tilde{p}_t > k_t$ is strictly profitable.

It is convenient to denote $u_\theta(k) \overset{def}{=} U(\theta, k)$. Then, since $U(p,k)$ is linear in p, it follows that $U(p,k) = p \cdot u_1(k) + (1-p) \cdot u_0(k)$. If the intermediary offers the agent training $a_t(k_t)$, then her flow payoff is equal to

$$\begin{aligned} & a_t(k_t) \cdot (1-\tilde{p}_t) \cdot \partial_1 U(\tilde{p}_t, k_t) + [\gamma_t + a_t(k_t)(1-k_t)] \cdot \partial_2 U(\tilde{p}_t, k_t) - c(a_t(k_t)) \\ =& a_t(k_t) \cdot (1-\tilde{p}_t) \cdot [u_1(k_t) - u_0(k_t)] \\ &+ [\gamma_t + a_t(k_t) \cdot (1-k_t)] \cdot \left[\tilde{p}_t \cdot u_1'(k_t) + (1-\tilde{p}_t) \cdot u_0'(k_t)\right] - c(a_t(k_t)). \end{aligned} \tag{A.131}$$

It follows from the definition of $u_\theta(k) \overset{def}{=} U(\theta, k)$ in (5) and Ito's lemma that

$$ru_\theta(k) = A(k) + \lambda k(1-k) \cdot u'(k) + \lambda \cdot (1-\theta) \cdot (L - u_\theta(k)). \tag{A.132}$$

The intermediary's flow payoff in (A.131) is linear in the private belief $\tilde{p}_t$ with the slope equal to

$$\begin{aligned} & - a_t(k_t) \cdot [u_1(k_t) - u_0(k_t)] + [\gamma_t + a_t(k_t) \cdot (1-k_t)] \cdot \left[u_1'(k_t) - u_0'(k_t)\right] \\ & \geq a_t(k_t) \cdot \left[- (u_1(k_t) - u_0(k_t)) + (1-k_t) \cdot \left(u_1'(k_t) - u_0'(k_t)\right)\right]. \end{aligned} \tag{A.133}$$

It follows from (A.132) that

$$\begin{aligned} \lambda k(1-k) \cdot (u_1'(k) - u_0'(k)) &= r \cdot (u_1(k) - u_0(k)) + \lambda(1-k) \cdot (L - u_0(k)) \\ (1-k) \cdot (u_1'(k) - u_0'(k)) &= \frac{r \cdot (u_1(k) - u_0(k))}{\lambda k} + \frac{1-k}{k} \cdot (L - u_0(k)). \end{aligned} \tag{A.134}$$

Substitute (A.134) into (A.133) to obtain

$$\begin{aligned} & a_t(k_t) \cdot \left[- (u_1(k_t) - u_0(k_t)) + (1-k_t) \cdot \left(u_1'(k_t) - u_0'(k_t)\right)\right] \\ \sim & - (u_1(k_t) - u_0(k_t)) + \frac{r \cdot (u_1(k_t) - u_0(k_t))}{\lambda k_t} + \frac{1-k_t}{k_t} \cdot (L - u_0(k_t)) \\ =& u_1(k_t) \cdot \left[-1 + \frac{r}{\lambda k_t}\right] + u_0(k_t) \cdot \left[1 - \frac{r}{\lambda k_t} - \frac{1-k_t}{k_t}\right] + \frac{1-k_t}{k_t} \cdot L \end{aligned} \tag{A.135}$$

We have

$$u_0(k) = \int_0^\infty e^{-(r+\lambda)t} \cdot \left[A(\pi_t(k)) - rL\right] + L \leq \frac{r}{r+\lambda} \cdot \int_0^\infty e^{-rt} \cdot \left[A(\pi_t(k)) - rL\right] + L$$

$$= \frac{r}{r+\lambda} \cdot [u_1(k) - L] + L = \frac{r}{r+\lambda} \cdot u_1(k) + \frac{\lambda}{r+\lambda} \cdot L. \tag{A.136}$$

Suppose $\lambda \leq r$. Then substituting (A.136) into (A.135) obtain

$$\begin{aligned}
&u_1(k_t) \cdot \left[-1 + \frac{r}{\lambda k_t}\right] + u_0(k_t) \cdot \left[1 - \frac{r}{\lambda k_t} - \frac{1-k_t}{k_t}\right] + \frac{1-k_t}{k_t} \cdot L \\
\geq &u_1(k_t) \cdot \left[-1 + \frac{r}{\lambda k_t}\right] + \left[\frac{r}{r+\lambda} \cdot u_1(k) + \frac{\lambda}{r+\lambda} \cdot L\right] \cdot \left[1 - \frac{r}{\lambda k_t} - \frac{1-k_t}{k_t}\right] + \frac{1-k_t}{k_t} \cdot L \\
= &u_1(k_t) \cdot \left[-1 + \frac{r}{\lambda k_t} + \frac{r}{r+\lambda} - \frac{r}{r+\lambda} \cdot \frac{r}{\lambda k_t} - \frac{r}{r+\lambda} \cdot \frac{1-k_t}{k_t}\right] \\
+ &L \cdot \left[\frac{\lambda}{r+\lambda} - \frac{\lambda}{r+\lambda} \cdot \frac{r}{\lambda k_t} + \frac{r}{r+\lambda} \cdot \frac{1-k_t}{k_t}\right] \\
= &u_1(k_t) \cdot \left[-\frac{\lambda}{r+\lambda} + \frac{\lambda}{r+\lambda} \cdot \frac{r}{\lambda k_t} - \frac{r}{r+\lambda} \cdot \frac{1-k_t}{k_t}\right] + L \cdot \left[\frac{\lambda}{r+\lambda} - \frac{\lambda}{r+\lambda} \cdot \frac{r}{\lambda k_t} + \frac{r}{r+\lambda} \cdot \frac{1-k_t}{k_t}\right] \\
= &\Big[u_1(k_t) - L\Big] \cdot \left[\frac{\lambda}{r+\lambda} \cdot \frac{r}{\lambda k_t} - \frac{\lambda}{r+\lambda} - \frac{r}{r+\lambda} \cdot \frac{1-k_t}{k_t}\right] = \Big[u_1(k_t) - L\Big] \cdot \left[\lambda \cdot \frac{r}{\lambda k_t} - \lambda - r \cdot \frac{1-k_t}{k_t}\right] \\
= &\Big[u_1(k_t) - L\Big] \cdot \left[\frac{r}{k_t} - \lambda - r \cdot \frac{1-k_t}{k_t}\right] = \frac{u_1(k_t) - L}{k_t} \cdot [r - \lambda k_t - r \cdot (1-k_t)] \\
= &\frac{u_1(k_t) - L}{k_t} \cdot [k_t \cdot (r - \lambda)] \geq 0.
\end{aligned} \tag{A.137}$$

Inequality (A.137) shows that higher skilled agents generate a greater flow profit for the intermediary in the upper bound $\hat{V}(\tilde{p}_t, k_t)$. This implies that if an agent of skill $\tilde{p}_t = k_t$ should be retained, then higher-skilled agents $\tilde{p}_t > k_t$ should also be retained. Similarly, if type $\tilde{p}_t = k_t$ is being let go, then strictly higher profitability of higher skilled agents implies that the intermediary prefers to retain them for $\tilde{p}_t > k_t$. □

The sufficient condition that $r \geq \lambda$ ensures that higher-skilled agents value the reputation-building aspects of working for the intermediary more than lower-skilled agents value the training provided to them.[8] The intermediary has no interest in training the agent when he is close to being let go, while this is precisely the time when the agent builds reputation via the intermediary's churning, as can be seen from the churning rate γ illustrated in Figure 1b for the case without training. Churning occurs when the information asymmetry is low, while training, in contrast, is focused on quiet periods when information asymmetry is high and the churning rate γ is absent or low. This distinction is most pronounced for higher-skilled agents towards the end of their careers with the intermediary, when the training benefits are smallest, while the churning rate is greatest.

[8] Condition $r \geq \lambda$ is sufficient for all revenue functions $A(\cdot)$ and distributions $F(\cdot)$ and, thus, can be relaxed for specific functional forms. This condition can be relaxed if we were to introduce skill depreciation into the model, making training more valuable for higher-skilled agents similar to Acemoglu and Pischke (1998).

A.5 Extension: Possibility for Market Breakdown in the Absence of the Intermediary

In this section, we show if the expected value of the agent's services can be negative, i.e., if $A(0) < 0$, then the market for his services can break down – clients are unwilling to procure services as they have a negative value for them. Such under-provision of services may be inefficient as clients are short-lived and do not internalize the experimentation benefits associated with the agent providing services to future clients. Consequently, the agent cannot generate performance signals when working independently and, thus, escape the breakdown region. We show that the intermediary's ex-ante selection, as described by Observation 1, can restore these services even if the intermediary does not subsidize early clients.

In what follows, we refer to the agent as being credit-constrained if he cannot accept a negative payment from clients, i.e., services must be sold for a weakly positive amount. For simplicity of the analysis, we set the agent's outside option $L = 0$.

Observation A.5 (Market breakdown under autarky)**.** *A credit-constrained agent can start his own firm under autarky, i.e., in the absence of the intermediary, if and only if the ex-ante belief distribution* $\tilde{p}_0 \sim F(\cdot)$ *satisfies* $\mathrm{E}[\tilde{p}_0] < A^{-1}(0)$*.*

Proof. Suppose the agent's reputation $k \geq A^{-1}(0)$. Then all agents prefer to start their own firm, rather than taking their outside option $L = 0$. Given the benefits of staying in the industry, all agents stay in the industry as long as they perform well, resulting in a value function $U(\tilde{p}_0, k)$ defined in (5). As this applies to agents of every skill, this implies that all agents enter, and the clients' rational expectations require that $k_0 = \mathrm{E}[\tilde{p}_0]$. Consequently, if $\mathrm{E}[\tilde{p}_0] \geq A^{-1}(0)$ all agents open a firm in equilibrium.

Suppose the agent's reputation $\mathrm{E}[\tilde{p}_0] < A^{-1}(0)$ then no agent can open a firm in equilibrium as, if only a fraction of agents were to enter and $k_0 \geq A^{-1}(0)$, then all agents would then prefer to enter due to the positive option value of generating good performance in the future. Consequently, positive entry cannot be an equilibrium, and the agent is unable to contract directly with clients if $\mathrm{E}[\tilde{p}_0] < A^{-1}(0)$. □

Observation A.5 implies that in the presence of credit constraints, the agent's expected value from having private belief $\tilde{p}$ and reputation k is equal to

$$U^*(\tilde{p}, k) \overset{def}{=} U(\tilde{p}, k) \cdot \mathbb{1}\{k \geq A^{-1}(0)\}, \tag{A.138}$$

where $U(\tilde{p}, k)$ is defined in (5). The credit constraint shows up as a discontinuity of the agent's expected value of the agent at $k = A^{-1}(0)$ given by $U^*(p, A^{-1}(0)-) = 0 < U(p, A^{-1}(0)) = U^*(p, A^{-1}(0))$.

We now show that as long as the intermediary's reservation value V is sufficiently large – the exact value depending on the shape of the revenue function $A(p)$ – then the presence of the intermediary restores a market for the agent's services. This holds even if the intermediary is also credit constrained and is not in a position to subsidize the agent's services, i.e., pay clients to use them.[9]

Observation A.6 (Intermediary equilibrium with a lemons problem)**.** *Suppose $A(\pi_t(\underline{p}))$ is weakly concave in $t \geq 0$. Then there exists a $\underline{V}$ such that for every $V > \underline{V}$ there exists an intermediary equilibrium in which the hiring threshold k_0 is such that $q_0 = \mathrm{E}\left[\tilde{p}_0 | \tilde{p}_0 \geq k_0\right] \geq A^{-1}(0)$ implying that the market for agent's services is restored.*

Proof. Define $\underline{k}^P \overset{def}{=} Q_0^{-1}(A^{-1}(0))$ as the lowest type that is able to open his own firm if he pools with all types above it. Define $\underline{k}^S \overset{def}{=} A^{-1}(0)$ as the lowest type that is able to open his own firm if he is identified correctly. Due to the subsidy associated with pooling, it follows that $\underline{k}^S > \underline{k}^P$.

Consider a candidate hiring threshold k_0. Define $\hat{t}(k_0)$ as the first time when the agent of type k_0 can open his own firm if the outside clients correctly perceive him

$$\hat{t}(k_0) \overset{def}{=} \inf\left\{t \geq 0 : \quad \pi_t(k_0) \geq \underline{k}^S\right\}.$$

If the agent is initially hired at $t = 0$, then his reservation wage is equal to 0 during $[0, \hat{t}(k_0)]$. Since his revenues $A(\pi_t(q_0))$ are increasing in t along the path of good performance, then all initially hired agents, i.e., those that are not let go at $t = 0$, are retained at least until $\hat{t}(k_0)$ along the path of good performance – otherwise it would be unprofitable to retain them beyond $t = 0$. Define time $t^*(k_0)$ as the first time when the agent becomes unprofitable for the intermediary, i.e., the length of the initial quiet period as a function of k_0,

$$t^*(k_0) \overset{def}{=} \inf\{t \geq \hat{t}(k_0) : \quad A(Q_t(k_0)) - A(\pi_t(k_0)) \leq rV\}.$$

If $A\left(Q_{\hat{t}(k_0)}(k_0)\right) - A\left(\pi_{\hat{t}(k_0)}(k_0)\right) \leq rV$, then $t^*(k_0) = \hat{t}(k_0)$ and the churning period starts immediately once the agent is able to open his own firm if he leaves. If, however, $A\left(Q_{\hat{t}(k_0)}(k_0)\right) - A\left(\pi_{\hat{t}(k_0)}(k_0)\right) > rV$, then $t^*(k_0) > \hat{t}(k_0)$ and the quiet period extends beyond the time when the agent can contract directly with clients.

The intermediary's profitability of employing the agent of skill k_0 is positive only as long as clients are willing to purchase the agent's services, i.e., $k_0 > \underline{k}^P$. If the intermediary employs agents $[k_0, 1]$ in equilibrium, then the expected value from employing the worst agent k_0 is

$$\begin{aligned} G(k_0; V) \overset{def}{=} & \int_0^{\hat{t}(k_0)} e^{-rt} \cdot \left(k_0 + (1 - k_0) \cdot e^{-\lambda t}\right) \cdot \overbrace{\left[A\left(Q_t(k_0)\right) - rV\right]}^{\textit{increasing in } t} dt \\ & + \int_{\hat{t}(k_0)}^{t^*(k_0)} e^{-rt} \cdot \left(k_0 + (1 - k_0) e^{-\lambda t}\right) \cdot \underbrace{\left[A\left(Q_t(k_0)\right) - A(\pi_t(k_0)) - rV\right]}_{\textit{decreasing in } t \textit{ due to concavity of } A(\pi_t(\underline{p})) \textit{ in } t} dt + V \end{aligned}$$

[9] If the intermediary can subsidize the agent's services, then, as to be expected, she can easily alleviate the agent's credit constraints.

$$= \int_{0}^{t^*(k_0)} e^{-rt} \cdot \left(k_0 + (1-k_0) \cdot e^{-\lambda t}\right) \cdot \left[A\left(Q_t(k_0)\right) - \max\left\{0, A(\pi_t(k_0))\right\} - rV\right] \, dt + V.$$

By definition of $\underline{k}^S$ it follows that $A(\pi_t(k_0)) \geq 0$ for all $k_0 \geq \underline{k}^S$ and $t \geq 0$. This implies that $G(k_0; V) \geq V$ for every $k_0 \geq \underline{k}^S$. At $V = 0$ we have

$$G(k_0; 0) = \int_{\hat{t}(k_0)}^{t^*(k_0)} e^{-rt} \cdot \left(k_0 + (1-k_0) \cdot e^{-\lambda t}\right) \cdot \left[A\left(Q_t(k_0)\right) - \max\left\{0, A(\pi_t(k_0))\right\}\right] \, dt. \tag{A.139}$$

If $k_0 \geq \underline{k}^P$, then the integrand in (A.139) is positive, implying that $G(k_0; 0) > 0$. This implies that if $V = 0$, then the intermediary finds it optimal to hire all agents who are less skilled than $\underline{k}^P$. Consequently, if $V = 0$ the intermediary does not contribute to positive selection and the equilibrium would feature the same market breakdown as in Observation A.5 as the quality of entering agents would fall below $Q(\underline{k}^P) = A^{-1}(0)$ and the intermediary would be unable to sell the agent's services.

The intermediary's incentive to hire the less skilled agent decreases in V, as can be seen from

$$\frac{\partial}{\partial V} \left[G(k_0; V) - V\right] = - \int_{0}^{t^*(k_0)} e^{-rt} \cdot \left(k_0 + (1-k_0) \cdot e^{-\lambda t}\right) \cdot r \, dt < 0.$$

Define $\underline{V}$ as the minimal reservation value at which the intermediary prefers not to hire any agents below $\underline{k}^P$, formally defined as

$$\underline{V} \stackrel{def}{=} \inf\{V \geq 0 : \quad s.t. \quad G(k_0; V) < V \quad \forall \, k_0 \leq \underline{k}^P\}. \tag{A.140}$$

For a given $V \geq \underline{V}$ denote by k_0 the lowest skilled agent the intermediary is willing to hire, i.e., $G(k_0; V) = V$. Since the intermediary is always willing to hire the interval of agents $\tilde{p}_0 \geq k_0 \geq \underline{k}^S$, such a type always exists. Moreover, by definition of $\underline{V}$ it must be the case that $k_0 \geq \underline{k}^P$, implying that $k_0 \in [\underline{k}^P, \underline{k}^S]$. Define

$$\begin{aligned} G(\tilde{p}_0, k_0; V) \stackrel{def}{=} & \int_{0}^{\hat{t}(k_0)} e^{-rt} \cdot \left(\tilde{p}_0 + (1-\tilde{p}_0) \cdot e^{-\lambda t}\right) \cdot \left[A\left(Q_t(k_0)\right) - rV\right] \, dt \\ & + \int_{\hat{t}(k_0)}^{t^*(k_0)} e^{-rt} \cdot \left(\tilde{p}_0 + (1-\tilde{p}_0) \cdot e^{-\lambda t}\right) \cdot \left[A\left(Q_t(k_0)\right) - A(\pi_t(k_0)) - rV\right] \, dt + V. \end{aligned}$$

It is unprofitable to hire agent $\tilde{p}_0 < k_0$ and retain them until $t^*(k_0)$ if

$$\begin{aligned} 0 \leq \frac{\partial G}{\partial \tilde{p}_0}(\tilde{p}_0, k_0; V) = & \int_{0}^{\hat{t}(k_0)} e^{-rt} \cdot \left(1 - e^{-\lambda t}\right) \cdot \left[A\left(Q_t(k_0)\right) - rV\right] \, dt \\ & + \int_{\hat{t}(k_0)}^{t^*(k_0)} e^{-rt} \cdot \left(1 - e^{-\lambda t}\right) \cdot \left[A\left(Q_t(k_0)\right) - A(\pi_t(k_0)) - rV\right] \, dt \\ = & \int_{0}^{t^*(k_0)} e^{-rt} \cdot \left(1 - e^{-\lambda t}\right) \cdot \left[A\left(Q_t(k_0)\right) - \max\left\{0, A(\pi_t(k_0))\right\} - rV\right] \, dt. \end{aligned}$$

By definition of k_0, we have $G(k_0; V) = G(k_0, k_0; V) = V$. This implies that the agent of type k_0 is exactly indifferent between cash flow stream $C_t \equiv 0$ and cash flow stream $C'_t = A\left(Q_t(k_0)\right) - \max\left\{0, A(\pi_t(k_0))\right\} - rV$. Cash flow $C = (C_t)_{t \geq 0}$ and $C' = (C'_t)_{t \geq 0}$ satisfy the single crossing property with $C_t > C'_t$ for $t < t'$ an d $C_t < C'_t$ for $t > t'$, where t' is defined as $A(\pi_{t'}(Q(k_0))) = rV$. It follows from the single-crossing property established in Lemma A.16 earlier that type $\tilde{p}_0 > k_0$ strictly prefers cash flow C' to C, while type $p < k_0$ strictly prefers C to C'. Consequently

$$\begin{aligned}
\frac{\partial G}{\partial \tilde{p}_0}(\tilde{p}_0, k_0; V) &= \int_0^{t^*(k_0)} e^{-rt} \cdot \left(1 - e^{-\lambda t}\right) \cdot \left[A\left(Q_t(k_0)\right) - \max\left\{0, A(\pi_t(k_0))\right\} - rV\right] dt \\
&= \underbrace{\int_0^{t^*(k_0)} e^{-rt} \cdot \left[A\left(Q_t(k_0)\right) - \max\left\{0, A(\pi_t(k_0))\right\} - rV\right] dt}_{>0} \\
&\quad - \underbrace{\int_0^{t^*(k_0)} e^{-(r+\lambda)t} \cdot \left[A\left(Q_t(k_0)\right) - \max\left\{0, A(\pi_t(k_0))\right\} - rV\right] dt}_{<0} > 0.
\end{aligned}$$

Consequently, if $G(k_0; V) = 0$, then the intermediary does not find it profitable to hire agents of ability $\tilde{p}_0 < k_0$. Consequently, the hiring threshold $k_0 \in [\underline{k}^P, \underline{k}^S]$ constitutes an equilibrium. Since $k_0 \geq \underline{k}^P$, the clients are willing to pay for the agent's services, but only if he starts his career with the intermediary. □

A.6 A More Detailed Literature Overview

Our paper provides a new framework to think about compensation and turnover that builds on dynamic asymmetric information and public learning and, importantly, without relying on any form of dynamic commitment. We accomplish it by combining two influential but previously disconnected literatures on public learning, such as Jovanovic (1979) and Farber and Gibbons (1996), with the literature on asymmetric information, such as Greenwald (1986) and Waldman (1984), in the labor market. We show how quiet and churning periods can manifest as endogenous promotions – distinct times in the agent's career when his compensation discretely increases, and subsequent turnover rate discretely decreases. This is a novel perspective on promotions that does not involve public job assignment and shows how the quiet periods and pay-for-reputation mechanics manifest in the worker's career. In this section,n we provide a more detailed discussion of our contribution to the literature on labor markets under asymmetric information, promotions, and the up-or-out contract literature.

Asymmetric Information and Public Learning

The churning period in our model features some similarities to the gradual turnover in Greenwald (1986), however as we explained above, our paper derives novel equilibrium features on top of what is conveyed in

Greenwald (1986). To further highlight the distinction from Greenwald (1986), we show in Section A.7 of this Online Appendix that introducing public news in the model of Greenwald (1986) leads to a qualitatively different prediction relative to our model. Specifically, when public signals are introduced, they shrink the residual information asymmetry and, interacted with the exogenous separation rate assumed in Greenwald (1986), result in less turnover in Greenwald (1986), not more as we show in our model. It has also become clear to us that the exogenous separations approach in Greenwald (1986) substantially limits the theoretical tractability of incorporating public news in Greenwald (1986), making it much more difficult to characterize the pay-for-reputation dynamics, making our approach, if anything, simpler.

While pay-for-reputation does arise in equilibrium in Greenwald (1986) and Waldman (1984), it is not the focus of these papers. There is a separate literature exploring the role of public learning on worker allocation and turnover as pioneered by Jovanovic (1979) and further explored in papers such as Farber and Gibbons (1996). To our surprise, there has been almost no interactions between these two literatures. Part of our contribution is to bridge this gap by considering an asymmetric information model in the presence of public performance signals. Modeling such interaction is important for professional services firms such as finance, investment banking, accounting, law, and management consulting. This approach allows us to identify new qualitative features of the agent's career, such as quiet periods, but also derive endogenous up-or-out promotions.

(i) **Quiet and churning periods.** Our model generates a quiet period during which all initially hired types are retained as long as they perform well. This quiet period corresponds to a static adverse selection friction on the worker, yet as the asymmetric information declines thanks to public performance signals, we show that the agent's career endogenously transitions to a churning phase in which his ability is gradually revealed via the retention decision. The distinction between the quiet and churning periods offers a testable prediction on both the turnover-tenure relationship as well as the decomposition of tenure into performance- and non-performance-related components, which has been done, for instance, in the context of CEO labor outcomes. The length of the quiet period is driven by the magnitude of the asymmetric information and the speed at which it dissipates through public news. The quiet period and its endogenous transition to the churning period during the agent's career is conceptually novel. Importantly, this cannot arise in Greenwald (1986) as all turnover originates from an inefficient exogenous separation rate which, if set to 0, eliminates all turnover in that model.

(ii) **Endogenous up-or-out promotions.** Section 3.3 considers endogenous promotions characterized by discrete increases in compensation and reductions in subsequent turnover. Such promotions occur in our model when there are multiple quiet periods – a consequence of the dynamic real option problem solved by the intermediary when choosing when to retain or let go of a worker. These results highlight the dynamic interaction between asymmetric information and performance signals and provide a novel perspective on promotions relative to how they have been viewed by the prior literature via either public task assignment, e.g., Waldman (1984), or moral hazard, e.g., Kahn and Huberman (1988). We also relate our findings to the up-or-out contracts resulting from moral hazard considerations. We review our contribution to the promotions and up-or-out contract literatures.

(iii) **Signaling with contracts.** Our framework also contributes to the broader literature on dynamic signaling and asymmetric information. Our model effectively extends education signaling of Spence (1973) to signaling with a career path. The quiet period in our paper is fundamentally different from the no-trade periods in Kremer and Skrzypacz (2007) and Daley and Green (2012) in which the lack of contractability of future signals leads higher types to delay the irreversible trade until the public news reveal their type. In our model, the agent's future wage depends on subsequent performance realizations. Moreover, we explicitly consider in Section 5 the possibility of the agent signing long-term contracts with clients. We provide conditions under which, even if the agent can sign efficient signaling contracts with clients, in equilibrium,m he still prefers to signal his ability via the intermediary. The intuition is that clients are either dispersed and feature turnover, while the agent-client contracts remain private – this is true in most professional settings, including finance, investment banking, accounting, law, and consulting. We show, in a completely revised Section 5 that the agent prefers to signal his ability via the intermediary whenever clients feature a sufficient degree of turnover. The intuition is that the tenure with the intermediary is a simple and credible signal that transcends across clients which private contracts do not. This highlights the value of signaling with a career history, such as a CV, rather than signaling with a contract. They also bridge the gap between signaling models with private offers, e.g., Swinkels (1999) and Noldeke and Van Damme (1990).

Promotions

A promotion in our setting is characterized by a discrete increase in pay and a discrete reduction in turnover probability. The continuous-time dynamics of our model highlight these career discontinuities that we identify with promotions. Unlike the classic work of Waldman (1984), promotions by themselves do not convey additional information – it is the gradual churning that precedes them that does. In addition, unlike the vast majority of the literature on promotions which we survey below, we do not rely on worker allocation to obtain endogenous promotions in the model. In this sense, our model could be seen as combining the classic papers of Greenwald (1986), Waldman (1984), and Farber and Gibbons (1996) to obtain endogenous promotions driven purely by asymmetric information and public learning about performance.

A promotion in our model arises at the transition between a churning and a quiet period – all agents that are not churned by the start of the quiet period receive a promotion. We show that such promotions can arise for arbitrary revenue functions $A(\cdot)$ if the agent is accumulating intermediary-specific human capital while being employed. This is consistent with Gilson and Mnookin (1989) who describes how associates in Law firms are let go gradually prior to their final evaluation to partnership and, hence, the gradual churning period preceding the promotion is when the information about the agent's ability is being revealed. The intermediary's promotion decision, thus, does not rely on active task reallocation of the agent, as in Waldman (1984) – as long as the agent has survived the churning period, he obtains the promotion. Moreover, the promotion is not associated with any discrete change of tasks as the revenue stream of the agent is a continuous function of his tenure. Moreover, since the promotion is characterized by the churning period followed by the quiet period, there must be sufficient residual asymmetric information about promoted agents – this means that promotions take place on a cohort-level timeline rather than when the agent's individual actual skill reaches the "promotion level."

A significant part of the promotion literature has focused on the efficiency of allocation of workers to tasks within a firm. The seminal contribution to this literature is Waldman (1984) who studies the tension between preserving asymmetric information about a worker and the signaling role of a promotion. The intuition is that similar to the quiet period in our paper, the firm can underpay the worker in the presence of asymmetric information. In the presence of multiple, e.g., low-talent and high-talent tasks, then the firm may publicly promote workers to the high-talent task that are sufficiently skilled. The benefit of such a public promotion is improved worker allocation, while the cost stems from the signaling effect such action entails – the outside markets infer that the promoted workers are higher skilled and the incumbent firm cannot underpay them as much in the labor market. While the paper is set in two periods, the intuition of Waldman (1984) is effectively static. Bernhardt (1995) extends Waldman (1984) to a dynamic setting highlighting the role of firm-specific capital accumulation – a worker is promoted once his accumulated capital is sufficient for the high-talent task. An important assumption is that the public promotion is tightly linked to the tasks associated with the specific job title – a firm would benefit from privately reallocating the worker to the more efficient task without promoting him. Our perspective on promotions is very much distinct from this approach – a promotion is purely market driven since it happens when the revenues, driven by the outside markets, reach a level that is high enough to warrant a quiet period that follows a churning period. In this sense, the intermediary in our model does not publicly allocate tasks, i.e., we do not impose firm hierarchy exogenously, and so the model is robust to the intermediary privately reassigning the worker to a different task without publicly promoting him. Put differently, while the promotion in Waldman (1984) and Bernhardt (1995) is defined as the allocation of a worker to the high-talent task, we define a promotion endogenously as a discrete increase in the worker's pay and a decrease reduction in the likelihood that he will be let go. The combination of private and public learning are critical to our results and our findings thus combine the asymmetric information aspects of Waldman

(1984) and Greenwald (1986) with the literature on public learning in the labor market such as Farber and Gibbons (1996), Gibbons and Waldman (1999), and Barlevy and Neal (2019).

An additional insight of our paper is that promotions are up-or-out and should preserve residual asymmetric information about the worker's general ability – this is necessary to generate a quiet period and occurs even if promotion occurs in response to firm-specific capital acquisition. This implies that up-or-out promotions in our model take place on a cohort level, rather than an agent level, even though the model is set in continuous time, which captures what we see in services firms such as law, finance, accounting, consulting, and even academia. This is quite distinct from Bernhardt (1995) who shows that promotions happen on a worker level and that they perfectly reveal the general skill of the worker. Moreover, in Bernhardt (1995) promotions follow a standard practice – not promoted workers are not let go – rather than the up-or-out practice that we obtain in our paper, consistent with careers in professional services firms, which is the focus of our paper.

Up-or-out Contracts

Our approach to obtaining up-or-out promotions in the presence of dynamic asymmetric information and public news is different from existing theories of up-or-out promotions focused on moral hazard frictions. The classic argument follows Kahn and Huberman (1988) in which it is efficient for the worker to invest in non-contractable firm-specific human capital to improve production. The worker faces a hold-up problem, however, as the firm cannot commit to increase the worker's wage in response to his human capital acquisition by claiming that the worker's human capital is still low. This hold-up problem precludes the worker from acquiring firm-specific capital and leads to substantial production inefficiencies. The commitment to an up-or-out contract restores incentives – if the firm commits to pay the worker a higher wage upon retention, then the firm only retains the worker at this high wage if he has acquired firm-specific capital since paying that high wage to a relatively unproductive worker is unprofitable. The up-or-out contract results in the firm paying the agent a high wage if he acquired human capital, and letting him go if he does not. This generates incentives for the worker to acquire human capital in the first place, leading to increases in production efficiencies. Waldman (1990) extends the argument of Kahn and Huberman (1988) by showing that under asymmetric information, an up-or-out contract also induces the worker to invest in general human capital and hence up-or-out contracts can arise in professions in which general skill is relatively more important. In both of these classic papers, it is necessary that the employer commits to the long-term up-or-out contract to manage the hold-up problem.

Ghosh and Waldman (2010) point out that promotion practices do not always go hand-in-hand with up-or-out practices – in many jobs, such as construction work, the worker can stay employed at a particular level of an organization without being promoted. In others, most notably the professional services industry such as finance, law, and accounting, the worker is either promoted or let go. Such differing practices pose a puzzle since it seems inefficient to let go of capable workers, such as legal associates, who are very close to the promotion threshold but still fall below it. The paper shows that when firm-specific capital acquisition is relatively low, then the argument of Waldman (1990) leads up-or-out practices to be optimal to generate incentives for the workers to invest in capital. If, however, firm-specific capital is relatively high, then the firm is better off not letting go of unpromoted workers since they have gained substantial firm-specific capital the value of which can be extracted. Promotions are not too costly to the firm since their signaling effect, as studied in Waldman (1984), is relatively weak as this firm-specific capital cannot be easily ported to a different firm. Ghosh and Waldman (2010) effectively combine the insights of Waldman (1984) and Waldman (1990), explaining why some firms use standard promotion practices, while other firms rely on up-or-out promotion practices, with the latter being optimal when general human capital is relatively more important than firm-specific human capital.

A number of papers have pointed out that up-or-out contracts are especially prevalent in the professional services industry while also highlighting the importance of general human capital in these professions. Axelson and Bond (2015) consider a repeated moral hazard setting and show that the high degree of moral hazard in finance and other professional service professions can result in up-or-out career structures and temporal segmentation of the labor market. Barlevy and Neal (2019) show how up-or-out contracts can help with the allocative efficiency of workers in the professional services industry.[10] The model thus focuses on the role of public learning about the worker's ability. Bar-Isaac and Leaver (2022) study a model of training and information disclosure, helping explain why professional services firms also establish outplacement career services as part of their up-or-out career structures. In the model, they do so by committing to publicly disclose bad matches that will not hurt the agents as much in the secondary job market if they do not receive a promotion. Bar-Isaac and Lévy (2022) study how labor market conditions lead the firm to assign the employee to informative tasks and the implications it carries for outplacement services.

A.7 Implications of Public News in Greenwald (1986)

In this section we formally consider the model of Greenwald (1986). We show that, unlike in our model, it leads to a reduction in agent turnover and does not result in quiet periods.

Model setup. The model of Greenwald (1986) can be embedded in the notation of our model as follows. Assume time is discrete and indexed by $t = 0, 1, 2$. First, we first prove our results analytically in Section A.7 in the context of a two-period model and then show these results numerically in Section A.7 for the

[10]Barlevy and Neal (2019) also contains an excellent literature overview of up-or-out contracts, and being relatively recent ensures that we do not miss important papers in our overview.

three period model. We denote the common discount factor between period as $\beta \in (0,1)$.

It is convenient embed the model of Greenwald (1986) in the notation of Section 2 of our paper. The agent's latent ability is equal to $\theta \in \{0,1\}$. As in Greenwald (1986), we assume the agent's productivity is equal to his type, i.e., $A(\theta) = \theta$. We assume that parties do not directly observe θ and are, instead, endowed with an imperfect private signal $\tilde{p}_0$ about it, distributed via a $F(\cdot)$ over $[\underline{p}, 1]$. As in Greenwald (1986), we assume this private information is observed at $t = 0$, and thus there are no strategic interactions in this first period as no party is informed.

An important assumption in Greenwald (1986) is that agents separate at an exogenous, i.e., independent of ability rate $\delta > 0$ in every period. Importantly, other firms do not observe whether the agent has separated for exogenous reasons, leading to an adverse selection problem.[11]

Three Period Model without a Public Signal

First, we characterize the solution to the model outlined above in the absence of a private signal, i.e., mirror the analysis of Greenwald (1986). Our aim is to illustrate the implications of the exogenous separation rate δ on the equilibrium strategies and reservation values. Just as in the analysis of our model, the equilibrium features cutoff strategies. Consider $(l_t)_{t=0,1,2}$ to be the highest ex-ante type still employed by the intermediary at time t. The initial condition is $l_0 = \underline{p}$.

Due to the exogenous separation rate δ, the average separating type is a conditional average across agents that separate exogenously and agents who are strategically let go by the intermediary. Denote by $(z_t)_{t=1,2}$ the average ex-ante separating type. In period $t = 2$ obtain

$$\begin{aligned}
z_2 &\overset{def}{=} \frac{\delta \cdot \mathrm{E}\left[\tilde{p}_0 \mid \tilde{p}_0 \in [l_1, 1]\right] + (1-\delta) \cdot \mathrm{E}\left[\tilde{p}_0 \cdot \mathbb{1}\{\tilde{p}_0 \in [l_1, l_2]\} \mid \tilde{p}_0 \in [l_1, 1]\right]}{\delta + (1-\delta) \cdot \mathrm{P}(\tilde{p}_0 \in [l_1, l_2] \mid \tilde{p}_0 \in [l_1, 1])} \\
&= \frac{\delta \cdot \frac{\int_{l_1}^{1} x\, dF(x)}{1-F(l_1)} + (1-\delta) \cdot \frac{\int_{l_1}^{l_2} x\, dF(x)}{1-F(l_1)}}{\delta + (1-\delta) \cdot \frac{F(l_2)-F(l_1)}{1-F(l_1)}} = \frac{\delta \cdot \int_{l_1}^{1} x\, dF(x) + (1-\delta) \cdot \int_{l_1}^{l_2} x\, dF(x)}{\delta \cdot (1-F(l_1)) + (1-\delta) \cdot (F(l_2) - F(l_1))} \\
&= \frac{\int_{l_1}^{l_2} x\, dF(x) + \delta \cdot \int_{l_2}^{1} x\, dF(x)}{\delta + (1-\delta) \cdot F(l_2) - F(l_1)}.
\end{aligned} \tag{A.141}$$

Similarly, the average separating type at $t = 1$ is given by

$$z_1 \overset{def}{=} \frac{\int_{l_0}^{l_1} x\, dF(x) + \delta \cdot \int_{l_1}^{1} x\, dF(x)}{\delta + (1-\delta) \cdot F(l_1) - F(l_0)} \overset{F(l_0)=0}{=} \frac{\int_{l_0}^{l_1} x\, dF(x) + \delta \cdot \int_{l_1}^{1} x\, dF(x)}{\delta + (1-\delta) \cdot F(l_1)}. \tag{A.142}$$

Lemma A.32. *The agent's outside option is equal to z_2 if he separates in period $t = 2$ and equal to $(1+\beta) \cdot z_1$ if he separates in period $t = 1$.*

[11] Such exogenous separation rate results in a trembling-hand-perfection like refinement and, in a setting like Spence (1973) or Noldeke and Van Damme (1990), would force a pooling equilibrium at $t = 0$.

Proof. Equation 9 on page 334 of Greenwald (1986). □

Retention at $t = 2$. The intermediary chooses which agent to let go at $t = 2$ based on their net profitability – the necessary wage z_2 needs to be greater than the intermediary's profit l_2 from employing the cutoff agent. As the intermediary has let go of all agents with ex-ante ability $\tilde{p}_0 \in [\underline{p}, l_1]$, the support of agents retained at the start of period $t = 2$ is $[l_1, 1]$. The intermediary's indifference condition can then be written as

$$l_1 = z_2 = \frac{\int_{l_1}^{l_2} x\, dF(x) + \delta \cdot \int_{l_2}^{1} x\, dF(x)}{\delta + (1-\delta)\cdot F(l_2) - F(l_1)}$$

$$l_2 \cdot (\delta + (1-\delta)\cdot F(l_2) - F(l_1)) = \int_{l_1}^{l_2} x\, dF(x) + \delta \cdot \int_{l_2}^{1} x\, dF(x)$$

$$\int_{l_1}^{l_2} (x - l_2)\, dF(x) + \delta \cdot \int_{l_2}^{1} (x - l_2)\, dF(x) = 0. \tag{A.143}$$

The left hand side of (A.143) is decreasing in l_2. Moreover, it is positive if $l_2 = l_1$ and negative if $l_2 = 1$. Consequently, there exists a unique solution $l_2 = L_2(l_1)$ which, importantly, depends on the cutoff type in period $t = 1$. Importantly, if the model has only two periods, then we can take $l_1 = \underline{p}$ in which case (A.143) characterizes the equilibrium cutoff type of the intermediary, which considerably simplifies the analysis.

Retention at $t = 1$. The intermediary must be similarly indifferent to retaining the cutoff agent $t = 1$ in period $t = 1$. The wage necessary to compensate the agent in period $t = 1$ must be such that collecting it and his reservation wage z_2 in period $t = 2$ makes him indifferent to collecting his outside option $(1+\beta)\cdot z_1$, as characterized by Lemma A.32. This pins down wage w_1 as

$$w_1 + \beta \cdot z_2 = (1+\beta)\cdot z_1 \qquad \Rightarrow \qquad w_1 = z_1 - \beta \cdot (z_2 - z_1).$$

The intermediary's indifference condition can then be written as

$$l_1 = z_1 - \beta \cdot (z_2 - z_1) = (1+\beta)\cdot z_1 - \beta \cdot z_2$$

Write this expression out

$$l_1 = (1+\beta) \cdot \frac{\int_{l_0}^{l_1} x\, dF(x) + \delta \cdot \int_{l_1}^{1} x\, dF(x)}{\delta + (1-\delta)\cdot F(l_1)} - \beta \cdot L_2(l_1), \tag{A.144}$$

where we have used the intermediary's period 2 indifference condition written as $z_2 = l_2 = L_2(l_1)$. Indifference condition (A.144) is not as simple as (A.143) due to the ambiguous shape of the second period response function $L_2(l_1)$. For this reason, we first consider a two-period version of Greenwald (1986) with news analytically, and then illustrate the three-period extension numerically.

Two Period Model with a Public Signal

Consider a public signal $\tilde{s}$ about the agent's ability $\tilde{p}_0$ that arrives in period $t = 1$. For comparability with the main text in which the performance process X defined in (1) reveals the low type agent with intensity $\lambda \cdot (1-\theta)$, suppose that $\tilde{s} \in \{L, H\}$ is a perfect bad news signal, i.e., it reveals the low type agent with probability $\lambda \in [0, 1]$, i.e.,

$$\mathrm{P}(s = L \,|\, \theta) = \lambda \cdot (1-\theta), \qquad \mathrm{P}(s = H \,|\, \theta) = 1 - \lambda \cdot (1-\theta) \tag{A.145}$$

The ex-ante probability that the signal realization will be high, i.e., $s = H$, is equal to

$$\mathrm{P}(s = H) = \int_{\underline{p}}^{1} \mathrm{P}(s = H \,|\, \tilde{p} = y)\; dF(y) = \int_{\underline{p}}^{1} (1 - \lambda \cdot (1-y))\, dF(y). \tag{A.146}$$

The posterior belief about an agent with ex-ante ability $\tilde{p}_0$ and signal $s = H$ is given by

$$\begin{aligned} \mathrm{P}_{\tilde{p}_0}(\theta = 1 \,|\, s = H) &= \frac{\mathrm{P}_{\tilde{p}_0}(\theta = 1, s = H)}{\mathrm{P}_{\tilde{p}_0}(s = H)} \\ &= \frac{\mathrm{P}(s = H \,|\, \theta = 1) \cdot \mathrm{P}_{\tilde{p}_0}(\theta = 1)}{\mathrm{P}(s = H \,|\, \theta = 1) \cdot \mathrm{P}_{\tilde{p}_0}(\theta = 1) + \mathrm{P}(s = H \,|\, \theta = 0) \cdot \mathrm{P}_{\tilde{p}_0}(\theta = 0)} \\ &= \frac{\tilde{p}_0}{\tilde{p}_0 + (1-\lambda) \cdot (1-\tilde{p}_0)} = \frac{\tilde{p}_0}{\tilde{p}_0 \cdot \lambda + 1 - \lambda}. \end{aligned} \tag{A.147}$$

One can invert (A.147) to obtain the private ex-ante type $\tilde{p}_0$ from the posterior $\mathrm{P}_{\tilde{p}_0}(\theta = 1 \,|\, s = H)$ as

$$\tilde{p}_0 = \frac{\mathrm{P}_{\tilde{p}_0}(\theta = 1 \,|\, s = H) \cdot (1-\lambda)}{1 - \lambda \cdot \mathrm{P}_{\tilde{p}_0}(\theta = 1 \,|\, s = H)}. \tag{A.148}$$

In the two-period version of the model the intermediary only chooses whether or not to let the agent go at $t = 1$. We denote by $l_1^s \in \{l_1^{S=L},\, l_1^{S=H}\}$ the intermediary's indifference threshold, which is indexed by the signal realization s which affects the posterior type distribution.

Observation A.7 (Decreasing turnover threshold)**.** *The churning threshold conditional on signal realization $s = L$ is given by $l_1^{s=L} = \underline{p}$, meaning that there is no strategic turnover. The churning threshold $l_1^{s=H}$ conditional on signal realization $s = H$ is decreasing in λ, meaning that more precise public news reduce strategic turnover even conditional on the positive signal realization.*

Proof. Conditional on signal $s = L$ the agent is revealed to have type $\theta = 0$ and, consequently, there is no more asymmetric information in the labor market. In this case, the agent's outside option is equal to 0 and the incumbent firm is indifferent between employing the agent and not. In this case we can interpret the agent as remaining employed by this firm unless he separates for exogenous reasons with probability

δ.[12] Consequently, the expected turnover conditional on $s = L$ is equal to δ, and the incumbent firm does not churn lower skilled agents, implying that $l^{s=L} = \underline{p}$.

Conditional on signal $s = H$ the agent is not revealed to be unskilled, which serves as a positive signal. The posterior belief about the agent of ex-ante ability $\tilde{p}_0$ is given by (A.147). Denote by $l = l_1^{s=H}$ as the lowest ex-ante type that is retained by the intermediary at $t = 1$ conditional on the public realization of signal $s = H$.

The agent of ex-ante ability $\tilde{p}_0$ receives signal $s = H$ with probability $1 - \lambda(1 - \tilde{p}_0)$. If $\tilde{p}_0 < l_1^{s=H}$, then the agent leaves the firm with certainty in period $t = 1$. If $\tilde{p}_0 \geq l_1^{s=H}$, then the agent leaves the firm only for exogenous reasons, which occur with probability δ. The expected ex-ante quality of agents leaving the firm at $t = 1$ as a function of the cutoff threshold $l_1^{s=H}$ is then given by

$$z_1^{s=H}(l_1^{s=H}) \stackrel{def}{=} \frac{\int_{\underline{p}}^{l_1^{s=H}} x \cdot (1 - \lambda(1-x))\, dF(x) + \delta \cdot \int_{l_1^{s=H}}^{1} x \cdot (1 - \lambda(1-x))\, dF(x)}{\int_{\underline{p}}^{l_1^{s=H}} (1 - \lambda(1-x))\, dF(x) + \delta \cdot \int_{l_1^{s=H}}^{1} (1 - \lambda(1-x))\, dF(x)}.$$

The firm's retention indifference condition $z_1^{S=H}(l_1^{s=H}) = l_1^{s=H}$ implies that the equilibrium retention threshold $l_1^{s=H}$ is given by

$$z_1^{s=H}(l_1^{s=H}) = \frac{\int_{\underline{p}}^{l_1^{s=H}} x \cdot (1 - \lambda(1-x))\, dF(x) + \delta \cdot \int_{l_1^{s=H}}^{1} x \cdot (1 - \lambda(1-x))\, dF(x)}{\int_{\underline{p}}^{l_1^{s=H}} (1 - \lambda(1-x))\, dF(x) + \delta \cdot \int_{l_1^{s=H}}^{1} (1 - \lambda(1-x))\, dF(x)} = l_1^{s=H}$$

$$\int_{\underline{p}}^{l_1^{s=H}} \left(x - l_1^{s=H}\right) \cdot (1 - \lambda(1-x))\, dF(x) + \delta \cdot \int_{l_1^{s=H}}^{1} \left(x - l_1^{s=H}\right) \cdot (1 - \lambda(1-x))\, dF(x) = 0. \tag{A.149}$$

Define

$$G(l, \lambda) \stackrel{def}{=} \int_{\underline{p}}^{l} (x - l) \cdot (1 - \lambda(1-x))\, dF(x) + \delta \cdot \int_{l}^{1} (x - l) \cdot (1 - \lambda(1-x))\, dF(x). \tag{A.150}$$

The fixed point (A.149) can then be expressed as $G(l_1^{s=H}, \lambda) = 0$. Differentiating $G(l_1^{s=H}, \lambda)$ defined in (A.150) with respect to $l_1^{s=H}$ obtain

$$\frac{\partial}{\partial l} G(l, \lambda) = -\int_{\underline{p}}^{l} (1 - \lambda(1-x))\, dF(x) - \delta \cdot \int_{l}^{1} (1 - \lambda(1-x))\, dF(x) < 0. \tag{A.151}$$

Differentiating $G(l, \lambda)$ defined in defined in (A.149) with respect to λ obtain

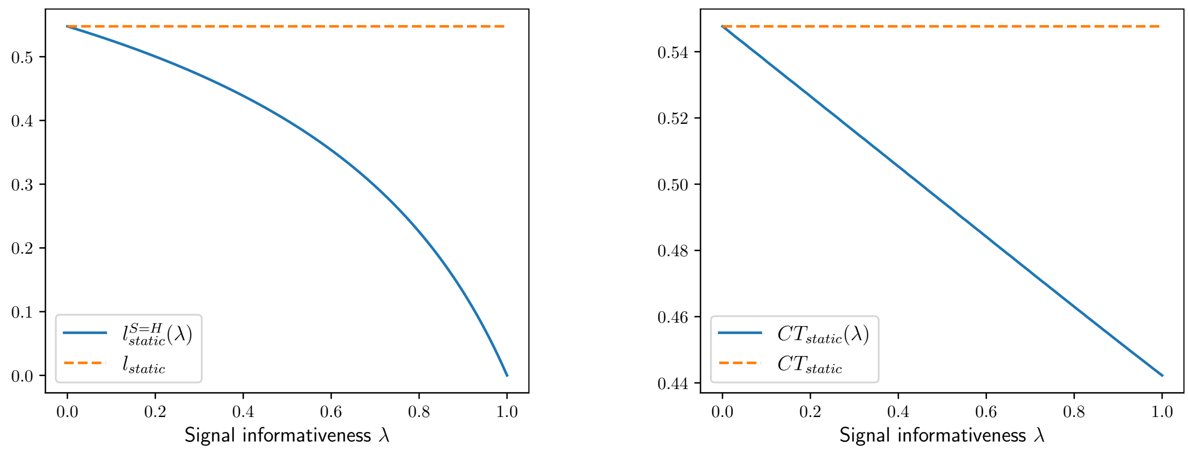
$$\frac{\partial}{\partial \lambda} G(l, \lambda) \stackrel{def}{=} -\int_{\underline{p}}^{l} (x - l) \cdot (1 - x)\, dF(x) - \delta \cdot \int_{l}^{1} (x - l) \cdot (1 - x)\, dF(x) < 0.$$

[12] Such resolution of the incumbent firm's indifference would follow if there are infinitesimal switching costs for the agent.

Differentiating $G(l_1^{s=H}(\lambda), \lambda) = 0$ obtain

$$\frac{d}{d\lambda} l_1^{s=H}(\lambda) \cdot \frac{\partial}{\partial l} G(l, \lambda)\Big|_{l=l_1^{s=H}(\lambda)} + \frac{\partial}{\partial \lambda} G(l, \lambda)\Big|_{l=l_1^{s=H}(\lambda)} = 0 \quad \Rightarrow \quad \frac{d}{d\lambda} l_1^{s=H}(\lambda) = -\frac{\frac{\partial}{\partial \lambda} G(l, \lambda)\Big|_{l=l_1^{s=H}(\lambda)}}{\frac{\partial}{\partial l} G(l, \lambda)\Big|_{l=l_1^{s=H}(\lambda)}} < 0,$$

which proves that $l_1^{s=H}(\lambda)$ is decreasing in λ. □



(a) Churning cutoff $l^{s=H}(\lambda)$ conditional on signal realization $s = H$ is a solution to (A.149). Churning cutoff $l_{static} = L_2(\underline{p})$ in the absence of public news is the solution to (A.143) setting $l_1 = \underline{p}$.

(b) Expected turnover $CT(\lambda)$ as a function of signal informativeness λ. Expected turnover CT_{static} corresponds to the static model without a public signal.

Figure A.2: Comparative statics of churning thresholds and expected churning in the two-period model of Greenwald (1986) with respect to the informativeness parameter λ of the public signal. Parameters: $\beta = 0.8$, $\delta = 0.3$, $\tilde{p} \sim U[0.3, 1]$.

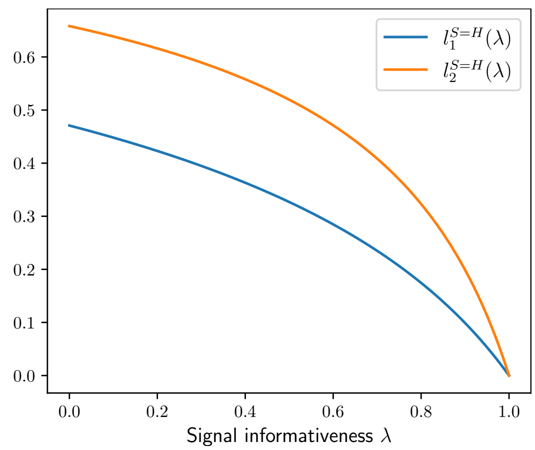
Figure A.2a illustrates the result of Observation A.7 – the churning threshold $l^{s=H}(\lambda)$ is strictly decreasing in λ and equal to the churning threshold in the absence of a public signal if $\lambda = 0$, i.e., the public signal is not at all informative. Figure A.2b shows that the expected cumulative turnover defined as

$$CT(\lambda) = \mathrm{P}\big(\tilde{p}_0 < l^{s=H}(\lambda) \,|\, S = H\big) \cdot \mathrm{P}(S = H)\,,$$

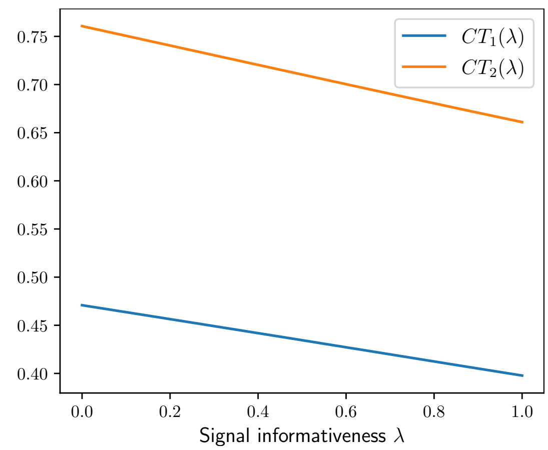
i.e., the probability that the agent leaves the incumbent firm in period $t = 1$, is decreasing in the informativeness λ of the public signal. This is due to both the decrease in the churning threshold $l^{s=H}(\lambda)$ and the reduced likelihood of generating a false positive signal $S = H$ for a higher λ. Figure A.2 and Observation A.7 show that public information, both positive and negative, reduces turnover in the two-period model of Greenwald (1986).

Three Period Model with a Public Signal

The three-period version of the model is essentially the same as described above, but there is an additional period, denoted by $t = 2$, for production and turnover. Figure A.3 illustrates the decrease in churning thresholds and expected turnover in the three-period version of the model of Greenwald (1986). For simplicity, we assume that the public signal only arrives in period $t = 1$.[13] Figure A.3a shows that churning cutoffs in both periods $t = 1$ and $t = 2$ decrease in the precision of asymmetric information λ. Similarly, the expected cumulative turnover $CT_t(\lambda) = \mathrm{P}\big(\tilde{p}_0 < l_t^{S=H}(\lambda)\big) \cdot \mathrm{P}(S = H)$ also decreases with the precision of asymmetric information.



(a) Churning cutoffs $l_{t=1}^{S=H}(\lambda)$ and $l_{t=2}^{S=H}(\lambda)$ conditional on signal realization $S = H$.



(b) Cumulative expected turnover $CT_{t=1}^{S=H}(\lambda)$ and $CT_{t=2}^{S=H}(\lambda)$ in periods $t = 1$ and $t = 2$ respectively.

Figure A.3: Comparative statics of churning thresholds and expected churning in a three-period model of Greenwald (1986) with respect to the informativeness parameter λ of the public signal. Parameters: $\beta = 0.8$, $\delta = 0.3$, $\tilde{p} \sim U[0.3, 1]$.

The above analysis shows that a reduction in asymmetric information results in less turnover in Greenwald (1986) in both the two- and three-period versions of Greenwald (1986). The economics behind this result is that turnover is fundamentally inefficient in Greenwald (1986), as it does not result in productive gains but, instead, mitigates the adverse selection frictions, resulting in agents separating for exogenous and unobservable reasons at rate δ.

[13] The analysis of the two-period model above shows that adding an additional signal in the final period $t = 2$ would further reduce the churning threshold and turnover.

B Equilibrium Selection via the Lowest Continuation Surplus Refinement

B.1 Refinement Idea and Intuition

In the main text, we establish equilibrium uniqueness when the clients' beliefs are weakly increasing in initial types either globally, as shown in Observation 2 or only during quiet periods as shown in Observation A.1. This restriction on the off-path belief process $(k_t)_{t\geq 0}$ is intuitive and clear, so we adopt it in the paper's main text. In this Online Appendix B, we show that this same equilibrium is uniquely selected from a limiting perturbation approach to off-equilibrium beliefs. This approach is related to equilibrium selection in Quantal Response Equilibria (QRE) of McKelvey and Palfrey (1995) and is equivalent to the divinity refinement of Cho and Kreps (1987) whenever the continuation value of the players after separation is independent of the private information. In what follows, we describe our perturbation approach, characterize the resulting equilibrium refinement, and then go on to prove uniqueness under this refinement. We then consider a two-type model and construct the equilibria of the perturbed model explicitly, highlighting the existence of the limiting sequence in the simplified setting.

Challenges of known refinements. The key challenge in pinning down equilibrium beliefs is in understanding the clients' beliefs during quiet periods when there is no separation on-path. Existing signaling models feature a single informed player, which allows the application of known equilibrium refinements such as divinity. Such refinements consider which types of agents are most likely to deviate and assign the most off-path likelihood to those agents. Our model is distinct in that there are two informed players – an agent and an intermediary – with distinct preferences. Application of known refinements requires attribution of which of the players initiates the split. Even then, we show in Section B.8 that if a divine equilibrium exists, either under a reasonable extension to a three-player model or the standard definition of D1 to a two-player version of our model, then it coincides with the quiet-churning equilibrium we construct in Section 3 of the main text. We prove that a divine equilibrium exists if the quiet period is not too long, as proxied by a not too small intermediary outside option V. Surprisingly, however, we show that if the intermediary's outside option V is too low, then a divine equilibrium may not exist at all. This is a surprising finding and we present a detailed analysis, discussion, and illustrations in Section B.8 of this Online Appendix.

These challenges make known refinements inapplicable directly, and for this reason, we approach the model from first principles by introducing disutility shocks akin to Acemoglu and Pischke (1998) into the employment relationship between the intermediary and the agent and studying the belief refinement that arises as the variance of these shocks converges to zero.

Perturbation approach. We assume, similar to Acemoglu and Pischke (1998), that in every instance an agent can receive a discrete disutility shock for working with the intermediary.[14] This disutility shock has an exponential distribution with parameter $\Delta > 0$ and rate of arrival ε.[15] Upon the arrival of this shock, the intermediary and the agent choose whether the value of the continued relationship is sufficient for the intermediary to compensate for the agent's disutility shock.[16] Consequently, the agent is retained only if the continued value of the relationship exceeds the value of the shock. Such stochastic shocks imply that there do not exist quiet periods in the perturbed model as in every instance, there is a positive probability of every agent receiving a high enough shock to trigger a split. Tracking the evolving distribution of separating types becomes highly complex and it is generally infeasible to solve this perturbed model in closed form. We show, however, that as the variance of the shocks in the perturbed model converges to 0, as captured by Δ increasing to infinity, the beliefs during any quiet period converge to the intermediary-agent pair that has the lowest equilibrium continuation surplus relative to the value of separating immediately. The resulting refinement is consistent with a stability intuition in that the players that have the most to lose from deviating would be least likely to do so. The resulting refinement echoes the relationships between divinity and stability, as shown by Cho and Sobel (1990), and coincides with those refinements in signaling games whether the ex-post value of signaling is independent of the agent's private beliefs, such as Spence (1973), Noldeke and Van Damme (1990), Swinkels (1999), Kremer and Skrzypacz (2007) among others. This refinement also presents an illustration how Quantal Response Equilibria of McKelvey and Palfrey (1995) can be used in equilibrium selection in signaling models. This equilibrium selection approach is formulated in Section B.2 and the limiting argument is presented in Section B.4.

Equilibrium uniqueness. We show in Section B.3 that the equilibrium constructed in Proposition 1 of the main text, featuring a quiet period followed by a churning period, is the unique equilibrium satisfying our refinement subject to an additional regularity condition that belief process $(k_t)_{t\geq 0}$ is continuous. First, we show that there must exist a final date $\bar{t}$ by which agents of all types leave the intermediary. If there is pooling prior to $\bar{t}$, then there must be a quiet period prior to $\bar{t}$. Our belief refinement then pins down that belief during this quiet period should equal to the belief of the lowest type separating at time $\bar{t}$. This results in a jump in beliefs at $\bar{t}$, which contradicts belief continuity. We then extend this argument to show that the equilibrium must be separating prior to $\bar{t}$, allowing us to pin down the dynamics of the churning period. Finally, and just like in the equilibrium constructed in Section 3 of the paper, the lowest skilled agents have the least surplus when working for the intermediary, resulting in the clients' beliefs tracking this worst agent during the quiet period. For tractability, we conduct this analysis under the assumption

[14] Due to flexible wages, it is not important as to which party, the intermediary or the agent, are subject to the disutility shock.

[15] The argument extends to any distribution with a decreasing likelihood ratio property.

[16] This ex-post negotiation between the intermediary and the agent alleviates the need for clients to attribute deviations to a particular party since, if a deviation is observed, the clients simply infer that no mutual agreement has been reached.

of Section 3 of the main text that $A(\pi_t(p))$ is concave in t, but this analysis can be extended beyond it with additional work. This first principles approach provides a rigorous micro-foundation for the equilibrium we construct in Section 3.1.

Online Appendix B structure. In Section B.2 we introduce the necessary notation, propose the candidate off-path refinement, discuss its relationship with existing refinements, formulate the main result and provide the intuition for the proofs. We show formally in Section B.3 that the equilibrium we construct in Section 3 of the main text is unique under our proposed refinement and the additional regularity condition that client beliefs are continuous along the path of good performance. In Section B.4 we construct the perturbed version of the model and show that the limiting beliefs of the perturbed model as the variance of the shocks converges to zero has to satisfy the belief refinement outlined in Section B.2 and, consequently, must converge to the unique equilibrium derived in Section B.3. We extend our results to the binary version of the model in Section B.5, i.e., one in which the initial type $\tilde{p}_0$ features only two types $\{\underline{p}, \overline{p}\} = \{p^L, p^H\}$ and explicitly construct an equilibrium of the perturbed version of the model in Section B.6. Finally, in Section B.8, we consider the divinity refinement in the context of the two-type model and show that a divine equilibrium may not exist when the initial quiet period is long.

B.2 Lowest Continuation Surplus Belief Refinement

The model setting mimics the description in Section 3 of the main text. Additionally, we assume that the agent's outside option $L = 0$, which, as we have shown in the proof of Observation A.1, is without loss. Finally, as in Section 3 of the main text, we assume that $A(\pi_t(\underline{p}))$ is weakly concave in t. This is done for simplicity due to the extensive nature of the proofs in this Online Appendix B.

To streamline the notation throughout Online Appendix B, we denote by $p = \tilde{p}_0$ the ex-ante type of the agent as indexed by the value of his private signal at the start of the game. The baseline equilibrium notion follows the Perfect Bayesian Equilibrium definition 1 in Section 2.

As in the main text proofs, define by $l_t \overset{def}{=} \pi_{-t}(k_t)$ to be the ex-ante type that clients assign to an agent leaving the intermediary at time t along the path of good performance. Similarly, define by $m_t \overset{def}{=} \pi_{-t}(q_t)$ to be the ex-ante average type along the path of good performance. We have shown in Lemma A.27 of the Online Appendix A that it must be the case that $l_t \leq m_t$ in every equilibrium.

On-path separations. Denote by $\tau(p) \subset [0, +\infty]$ to be the, possibly random, on-path stopping time when type p leaves the intermediary along the path of good performance in a candidate equilibrium. Denote $\mathbb{T}(p) \overset{def}{=} \mathit{support}(\tau(p))$ to be the support of $\tau(p)$ for each p. Denote the last date when type p separates in equilibrium by $\bar{t}(p) \overset{def}{=} \sup_{p \in [\underline{p}, \overline{p}]} \mathbb{T}(p)$. Denote by $\mathbb{T} \overset{def}{=} \mathit{support}(\tau) = \cup_p \mathbb{T}(p)$ to be the set of times when an agent separates from the intermediary along the equilibrium path. Denote by $\bar{t} \overset{def}{=} \sup \mathbb{T} = \sup \bar{t}(p)$ to be the last time when a type separates in equilibrium. We have shown in Lemma A.26 of the Online

Appendix A that $\bar{t}$ is finite and bounded by (A.97). Denote by $R(t)$ the support of ex-ante types that stay with the intermediary beyond time t with a positive probability along the equilibrium path

$$R(t) \stackrel{def}{=} closure\big\{p: \quad \mathrm{P}_p(\tau(p) > t) > 0\big\}. \tag{B.1}$$

Continuation values. Denote by $W_t(p)$ to be the joint continuation welfare at time t of the intermediary and the agent with ex-ante skill p who has performed well up to time t is given by:

$$\begin{aligned} W_t(p) \stackrel{def}{=} \max_{\hat{\tau} \geq t} \bigg\{ \int_t^{\hat{\tau}} & e^{-r(s-t)} \cdot \Big[\pi_t(p) + (1-\pi_t(p)) \cdot e^{-\lambda(s-t)}\Big] \cdot [A(\pi_s(m_s)) - rV] \; ds \\ & + e^{-r(\hat{\tau}-t)} \cdot \Big[\pi_{\hat{\tau}}(p) \cdot u_1(\pi_{\hat{\tau}}(l_{\hat{\tau}})) + (1-\pi_{\hat{\tau}}(p)) \cdot e^{-\lambda(\hat{\tau}-t)} \cdot u_0(\pi_{\hat{\tau}}(l_{\hat{\tau}}))\Big] + V \bigg\}. \end{aligned} \tag{B.2}$$

If $t \leq \bar{t}(p)$, then $W_t(p)$ is the on-path continuation value, while if $t > \bar{t}(p)$, then $W_t(p)$ is the off-path continuation value for the agent of ex-ante type p. The outside option at time t of the agent of ex-ante type p who has performed well up to time t, given client belief l_t, is given by

$$U_t(p) \stackrel{def}{=} U(\pi_t(p), \pi_t(l_t)) = \pi_t(p) \cdot u_1(\pi_t(l_t)) + (1-\pi_t(p)) \cdot u_0(\pi_t(l_t)). \tag{B.3}$$

Denote by $V_t(p)$ to be the continuation value at time t of the intermediary employing the agent of ex-ante type p who has performed well up until time t:

$$V_t(p) \stackrel{def}{=} W_t(p) - U_t(p). \tag{B.4}$$

The value added by the agent working for the intermediary is equal to the gain $V_t(p) - V$ that the intermediary gains relative to her outside option. This value also captures the joint equilibrium surplus of the intermediary-agent pair relative to the joint value of their outside options.

B.1.1 Lowest Continuation Surplus Belief Refinement Definition

We can now formally introduce the off-path belief refinement in which the clients attribute off-path separations to intermediary-agent types p that have the lowest continuation surplus $W_t(p)$ relative to their joint outside option $U_t(p) + V$. Following definition (B.4), this joint surplus is equal to $V_t(p) - V$, i.e., the intermediary's expected surplus from employing the agent of ex-ante type p from time t onwards. This is motivated by the idea that if the intermediary and the agent were to receive relationship-specific shocks, then it is precisely the pairs with the lowest continuation surplus who would be most likely to terminate their relationship.

Definition 4 (Lowest continuation surplus belief refinement)**.** *An equilibrium satisfies the lowest continuation surplus" refinement if for every $t \in [0, \bar{t}] \backslash \mathbb{T}$ the off-equilibrium path belief l_t lies in the convex hull of the* *support of the residual types $R(t)$, i.e., $l_t \in [\min R(t), \max R(t)]$ and corresponds to the intermediary-agent pair with the lowest continuation surplus*

$$V_t(l_t) = \min_{x \in [R(t)]} V_t(x). \tag{B.5}$$

Definition (4) requires, first, that clients assign off-path deviations to types that remain with the intermediary with positive probability at time t along the path of good performance. This first condition is equivalent to condition (ii) of Observation (ii) and can be equivalently restated that clients place zero weight on agents who have separated from the intermediary with certainty. The intuition for it is that only agents who are still employed by the intermediary would be subjected to separation shocks – a result we derive formally in Section B.4 when considering the perturbed version of the model. Second, among the ex-ante types that remain with the intermediary, clients assign off-path deviations to those who have the lowest continuation value surplus. Requirement (B.5) is also motivated by the perturbed version of the model and we derive its necessity in Section B.4 when considering the effect of relationship-specific shocks. While we require that definition 4 applies only to off-path beliefs, it is clear that on-path separations must also satisfy both of the outlined requirements following Bayesian consistency of client beliefs (3) and separation optimality in (6).

The definition of $V_t(p)$ in (B.4) captures the surplus obtained by the informed parties and, consequently, definition 4 can be extended beyond two informed players. Definition 4 is also less restrictive than requiring that $l_t \in \arg\min_{x \in R(t)} V_t(x)$ as it allows for beliefs l_t to lie in the convex hull of $R(t)$, thus allowing for mixed strategies, and such that $V_t(l_t)$ minimizes (B.5). It is also possible to embed right continuity of the belief process $(l_t)_{t \geq 0}$ in definition 4 by replacing (B.5) with $V_{t+}(l_t) = \min_{x \in R(t)} V_{t+}(x)$.

B.1.2 Relationship to Divinity and the Intuitive Criterion

Definition 4, as we mentioned above, can be extended to a single informed player, in which $V_t(p)$ denotes the equilibrium gain of the informed player. In this section we describe how the resulting refinement compares to established approaches in signaling games, such as the intuitive criterion of Cho and Kreps (1987) and divinity in Cho and Sobel (1990), which is equivalent to universal divinity and independence of never-weak-best responses of Kohlberg and Mertens (1986) in our monotonic signaling game. We also relate Definition 4 to trembling hand perfection of Selten and Bielefeld (1988).

Comparison to trembling hand perfection. We start by highlighting that definition 4 is distinct from the trembling hand perfection of Selten and Bielefeld (1988). The lowest continuation surplus refinement assumes that an off-path deviation is much more likely to arise from a type with the lowest continuation surplus, whereas trembling hand perfection assumes that the probability of a deviation is uniform across private types. The latter assumption implies that off-path beliefs must reflect the average player type that has not separated with positive probability, which is economically restrictive as it immediately rules out separating equilibria in numerous signaling models such as Spence (1973) and Noldeke and Van Damme (1990) among others.

We show in Section B.6 that the lowest continuation surplus refinement follows from a limiting perturbation argument in which the informed party receives a stochastic shock. Such perturbations result in a Quantal Response Equilibrium (QRE) as pioneered by McKelvey and Palfrey (1995) and McKelvey and Palfrey (1998) who also show that the selected equilibria differ from the trembling hand perfection of Selten and Bielefeld (1988) as the probabilities of off-path deviations are not uniformly distributed across informed player types and, instead, decrease in their continuation payoff, just like in definition 4.

Comparison to divinity. The expected value $W_t(p)$ of the intermediary-agent pair, defined in (B.2), is increasing in the clients' belief process l. Cho and Sobel (1990) show that such monotonicity results in an equivalence between universal divinity and independence of never-weak-best-responses. Consequently, the comparison of our refinement to divinity also speaks to these two established refinements.

The divinity refinement attributes off-path separations to a private type p that is willing to separate off-path for the widest range of hypothetical beliefs by investors. In the context of the monotone signaling game like hours, for a given type p we can define a time t indifference belief $d_t(p)$ such that the equilibrium payoff $W_t(p)$ to the informed player, is equal to immediate separation at the client belief $d_t(p)$. Formally, $d_t(p)$ is a solution to

$$W_t(p) = V + U(\pi_t(p), \pi_t(d_t(p))). \tag{B.6}$$

Due to the monotonicity of the signaling game, player p is then willing to separate off-path for any belief $(d_t(p), 1]$ assigned to him by clients. The divinity refinement then attributes an off-path separation to a player type p with the broadest range of client beliefs $(d_t(p), 1]$ for which he is willing to separate, which can be rewritten as the clients' belief l_t to belong to $\arg\min_p d_t(p)$. The divinity refinement, thus, considers informed player indifference in the client belief space. The lowest continuation surplus refinement defined via Definition 4 focuses on expected equilibrium values and explores which types are more likely to deviate in response to match preference shocks. This is conceptually similar, but can and should be thought of as stability-like perturbations to the informed agent's expected payoff space, rather than perturbations to client beliefs.

Definition 4 is equivalent to divinity whenever the expected post-separation value of the informed player is independent from his private belief, e.g., if the right hand side of (4) is independent of the private belief p. This is satisfied in signaling games such as Spence (1973) and Kremer and Skrzypacz (2007) where the agent's expected value post separation is independent of future information flow. In this case, the indifference belief $d_t(p)$ is strictly increasing in the agent's expected value $W_t(p)$ and, consequently,

$\arg\min_p W_t(p) = \arg\min_p d_t(p)$. Consequently, the distinction between divinity and the lowest continuation surplus refinement depends on the dependence of the agent's post-signaling value on his type. We show in Section B.1.3 that this is a meaningful distinction by illustrating the existence of a lowest continuation surplus equilibrium and non-existence of the corresponding divine equilibrium.

Comparison to the intuitive criterion. In a monotone signaling game such as ours, the intuitive criterion of Cho and Kreps (1987) rules out a time t deviation by player of ex-ante private type p if the equilibrium expected payoff $W_0(p)$ by this player is greater than the expected payoff this player would have received if he was perceived as the highest skilled type $\bar{p}$ at the time of his deviation. The intuitive criterion leads to a unique equilibrium in the two-type model of Spence (1973), but results in multiple equilibria when there are more than two types. A similar limitation arises in our setting – the equilibrium we construct in Section 3 of the main text survives the intuitive criterion, however the intuitive criterion does not discipline beliefs during any and every quiet period as every agent would prefer to be separate while being perceived as the highest type. The more formal distinction is that definition 4 considers the types that are most likely to deviate, which we micro-found in Section B.4 via a stochastic perturbation of the model, while the intuitive criterion rules out a rather narrow set of types that would be highly unlikely to deviate which is insufficient to refine the set of equilibria when there are either more than two types, or when there is a quiet period. We can show that when there are just two types, then the intuitive criterion results in the same equilibrium as the lowest continuation surplus refinement in 4.

B.1.3 Unique Equilibrium Satisfying Lowest Continuation Surplus Refinement

We now state the main result of this Online Appendix B – the equilibrium we construct in Section 3 is unique under the lowest continuation surplus refinement 4.

Proposition B.1 (Unique Equilibrium)**.** *Suppose $A(\pi_t(p))$ is concave in time t and $L = 0$. The Perfect Bayesian Equilibrium constructed in Proposition 1 is the unique equilibrium in which client belief process l is continuous along the path of good performance and satisfies the lowest continuation surplus refinement 4.*

Given the extensive nature of the proofs, we provide these results under the simplifying assumption of Section 3 that $A(\pi_t(p))$ is concave in t. As we have shown in the proof of Observation A.1, it is without loss to assume that $L = 0$, and we do not repeat the argument here out of concision. We also assume that the client belief process l is continuous – as we show in the formal proofs of Section B.3 this eliminates pooling in the final period without relying on any additional refinement assumptions. We overview the intuition and steps for the proof of Proposition B.1 written out formally in Section B.3.

First, as shown by Lemma A.26, the equilibrium separations must happen in finite time. We then work backwards. If there is pooling in the final period $\bar{t} = \mathit{support}(\tau)$, we show that there must be a quiet period preceding it. As higher-skilled agents have more to gain from reputation building, the lowest continuation surplus refinement 4 implies that client beliefs during the quiet period track the lowest skilled agent who separates at time $\bar{t}$. This, however, results in a jump in client beliefs at time $\bar{t}$, implying that there cannot be pooling at time $\bar{t}$ – this is the only moment where belief continuity matters and we show that by relaxing this condition, the equilibrium is very similar, but may feature pooling in the final period.[17] Next, we show that separation prior to $\bar{t}$ must be gradual with no pools of agents separating from the intermediary – we show that such jumps would result in positive belief jumps, which would violate incentive compatibility. We then apply the equivalent of Lemma A.30 from Section A.1 of Online Appendix A which shows that gradual separation periods must be increasing in types and unique in the sense that there cannot be two distinct periods during which an agent would be willing to separate. These steps prove that there must be a unique churning period characterized by Proposition 1. Finally, as higher-skilled agents have the most to gain from reputation building, the lowest continuation surplus refinement picks the lowest-skilled agent as the most likely to deviate during the quiet period, thus completing the equilibrium characterization. We now proceed to the formal proofs.

B.3 Proof of Proposition B.1 (unique equilibrium)

Denote the set of times when it is weakly optimal for ex-ante type p to separate

$$\begin{aligned}\mathbb{T}^I(p) \overset{def}{=} \arg\max_{\tau}\Bigg\{\int_0^{\tau} e^{-rt}\cdot\Big[p+(1-p)\cdot e^{-\lambda t}\Big]\cdot\big[A(\pi_t(m_t))-rV\big]\,dt \\ +e^{-r\tau}\cdot\Big[p\cdot u_1(\pi_\tau(l_\tau))+(1-p)\cdot e^{-\lambda\tau}\cdot u_0(\pi_\tau(l_\tau))\Big]+V\Bigg\},\end{aligned} \tag{B.7}$$

where superscript I in $\mathbb{T}^I$ stands for indifference. Denote by $\mathbb{T}^I \overset{def}{=} \cup_p \mathbb{T}^I(p)$ to be the set of times when there is an ex-ante type that finds it weakly optimal to separate.

Denote by $S(t)$ to be the set of ex-ante types which separate at time t with a positive probability, i.e.,

$$S(t) \overset{def}{=} \{p: \quad t\in\mathbb{T}(p)\} \qquad \textit{(on-path separation)} \tag{B.8}$$

Denote by $S^I(t)$ the set of ex-ante types for whom it is weakly optimal to separate at time t, i.e.,

$$S^I(t) \overset{def}{=} \big\{p: \quad t\in\mathbb{T}^I(p)\big\} \qquad \textit{(weak optimality)} \tag{B.9}$$

[17] Such final period separations can also be taken care of by expanding definition 4 to off-path separations that occur after time $\bar{t}$.

Define function $W_0(p;t)$ as the joint value obtained by type p for stopping at time t, given by

$$\begin{aligned} W_0(p;t) \stackrel{def}{=} & \int_0^t e^{-rs} \cdot \left[p + (1-p) \cdot e^{-\lambda s}\right] \cdot \left[A(\pi_s(m_s)) - rV\right] \, ds \\ & + e^{-rt} \left[p \cdot u_1(\pi_t(l_t)) + (1-p) e^{-\lambda t} \cdot u_0(\pi_t(l_t))\right] + V. \end{aligned} \tag{B.10}$$

As can be seen in (B.10), function $W_0(p;t)$ is linear in p.

Lemma B.1 (Separating set properties)**.** *Set $S^I(t)$ is convex for every t. If $\left|S^I(t_1) \cap S^I(t_2)\right| > 1$ then $S^I(t_1) = S^I(t_2)$ and $\mathbb{T}^I(p') = \mathbb{T}^I(p'')$ for every $p', p'' \in S^I(t_1) = S^I(t_2)$.*

Proof. Suppose $p_1, p_2 \in S^I(t)$. This implies that $t \in \mathbb{T}^I(p_1) \cap \mathbb{T}^I(p_2)$. It follows that $W_0(p_1) = W_0(p_1;t)$ and $W_0(p_2) = W_0(p_2;t)$. Due to weak convexity of $W_0(p)$, it implies that $W_0(p) = W_0(p;t)$ for all $p \in [p_1, p_2]$. Consequently, $S^I(t)$ is a convex set.

Following the above argument, $W_0(p)$ is linear over any $(p_1, p_2) \subseteq S^I(t)$ for each t. This implies that for any $p_1 \in S^I(t_1)$ and $p_2 \in S^I(t_2)$, and $\hat{p} \in S^I(t_1) \cap S^I(t_2)$ it follows that $W_0'(p_1;t) = W_0'(\hat{p};t) = W_0'(p_2;t)$. Since $W_0(p_1;t) = W_0(p_1)$ and $W_0(p_2;t) = W_0(p_2)$, then it follows that $W_0(p)$ is linear over $[p_1, p_2]$ and equal to $W_0(p;t)$. □

Lemma B.2 (Stopping time properties)**.** *Suppose $p_1, p_2 \in S^I(t)$. Then $\mathbb{T}^I(p_1) = \mathbb{T}^I(p_2) = \mathbb{T}^I(\hat{p})$ for any $\hat{p} \in [p_1, p_2]$.*

Proof. From Lemma B.1 it follows that $[p_1, p_2] \in S^I(t)$, meaning that $W_0(p)$ is linear over $[p_1, p_2]$ and $W_0(p) = W_0(p;t)$ for every $p \in [p_1, p_2]$. Consider a type $p' \in (p_1, p_2)$ and suppose that $t' \neq t$ and $t' \in \mathbb{T}^I(p')$. This implies that $W_0(p') = W_0(p';t')$. Then it must be the case that

$$W_0'(p') = \frac{\partial}{\partial p} W_0(p;t')\Big|_{p=p'}. \tag{B.11}$$

If (B.11) did not hold, then $W_0(p'-\varepsilon) < W_0(p'-\varepsilon;t')$ or $W_0(p'+\varepsilon) < W_0'(p'+\varepsilon;t')$, which would contradict with the definition of $W_0(p)$. □

Lemma B.3 (Strict convexity of $W_0(p)$)**.** *Value function $W_0(p)$ is strictly convex if and only if $|S^I(t)| \in \{0, 1\}$ for each t.*

Proof. Suppose $W_0(p)$ is strictly convex. Then it implies that there exists at most a single type such that $W_0(p) = W_0(p;t(p))$ given the tangency condition. Conversely, if $|S^I(t)| > 1$ then Lemma B.1 implies that $W_0(p) = W_0(p;t)$ is linear over $S^I(t)$, contradicting strict convexity of $W_0(p)$. □

Lemma B.4 (Best type separates last)**.** *In any equilibrium, it must be the case that $\bar{t} \in \mathbb{T}^I(\bar{p})$, i.e., the highest ex-ante type $\bar{p}$ finds it weakly optimal to separate at the final date $\bar{t}$.*

Proof. Suppose $\bar{t}(\bar{p}) < \bar{t}$. Denote $\hat{l} \overset{def}{=} \inf\left\{S\left(\mathbb{T}^I(\bar{p})\right)\right\}$ to be the lowest type that is willing to separate at the same time as $\bar{p}$. Following Lemma B.1, $\hat{l}$ is well defined and $S^I(\mathbb{T}^I(\bar{p})) = \left[\hat{l}, \bar{p}\right]$. For $t > \bar{t}(\bar{p})$ it follows that the only types that separate are types below $\hat{l}$, implying that $l_{\bar{t}(\bar{p})} \geq m_{\bar{t}(\bar{p})}$, which contradicts Lemma A.27 used in the proof of Observation A.1. Hence, it must be the case that $\bar{t} \in \mathbb{T}^I(\bar{p})$. □

Suppose process $(l_t)_{t\geq 0}$ is differentiable at t. It follows from (A.118) that the intermediary's continuation value $V_t(p)$ is also differentiable at t and satisfies

$$\begin{aligned} rV_t(p) - \dot{V}_t(p) = A(\pi_t(m_t)) - A(\pi_t(l_t)) - rV \\ + \left[\pi_t(p) \cdot u_1'(\pi_t(l_t)) + (1 - \pi_t(p)) \cdot u_0'(\pi_t(l_t))\right] \cdot \pi_t'(l_t) \cdot \dot{l}_t. \end{aligned} \tag{B.12}$$

Equilibria Satisfying the Lowest Surplus Belief Refinement 4

Npw we identify the set of equilibria that satisfy the lowest continuation surplus refinement 4. Define $t_S \overset{def}{=} \sup\{t < \bar{t} : \ t \in \mathbb{T}\}$ to be the last moment in which an agent leaves the intermediary along the equilibrium path prior to the final date $\bar{t}$.

Lemma B.5 (Final pooling properties)**.** *Suppose $|S(\bar{t})| > 1$. Then, if the equilibrium satisfies the lowest continuation surplus refinement 4, then $t_S < \bar{t}$. If, additionally, belief process $(l_t)_{t\geq 0}$ is right-continuous, then $l_t = \min R(t) = \min S(\bar{t})$ for every $t \in [t_S, \bar{t})$.*

Proof. **Quiet period.** Suppose $t_S = \bar{t}$, meaning that there exists a sequence of times $(t_n)_{n\geq 0} \in \mathbb{T}$ and such that $t_n \to \bar{t}$. Since $l_{\bar{t}} = m_{\bar{t}}$, incentive compatibility of the separating types requires that $\lim_{n\to\infty} l_{t_n} = m_{\bar{t}}$.

Suppose $S^I(t_{n_j}) \neq S^I(\bar{t})$ for some sub-sequence $(t_{n_j})_{j\in\mathbb{N}} \subseteq (t_n)_{n\in\mathbb{N}}$. Following Lemmas B.1 and B.4, set $S^I(\bar{t})$ is convex and includes $\bar{p}$, it implies that $l_{t_{n_j}} \in \left[\underline{p}, \min S(\bar{t})\right]$. Since $|S(\bar{t})| > 2$ it follows that $m_{\bar{t}} > \min S(\bar{t})$, implying a contradiction with $\lim_{n\to\infty} l_{t_n} = m_{\bar{t}}$. Hence it must be the case that there exists $N > 0$ such that $S^I(t_n) = S^I(\bar{t})$ for $n \geq N$.

Suppose $S^I(t_n) = S^I(\bar{t})$ for every $n \geq N$ implying that the intermediary letting go of agent of skill $p \in S(t_n)$ at time t_n is weakly indifferent to waiting until time $\bar{t}$ to let that agent go. For each $p_0 \in S^I(\bar{t})$ define by $L_t(p)$ to be a hypothetical indifference belief such that the joint value to the intermediary and the agent at time t is equal to the expected value of waiting until time $\bar{t}$ and be perceived as the posterior type $l_{\bar{t}} = m_{\bar{t}}$. Formally, $L_t(p)$ solves

$$V + U_t(L_t(p)) \overset{def}{=} W_t(p). \tag{B.13}$$

Belief process $L_t(p)$ is specific to the ex-ante type p. However, by continuity, it follows that $L_{\bar{t}}(p) = l_{\bar{t}} = m_{\bar{t}}$ for every $p \in S^I(\bar{t})$. Moreover, by the previous observation that $S^I(t_n) = S^I(\bar{t})$, it follows that $L_{t_n}(p) = l_{t_n}$

for every $p \in S^I(\bar{t})$ and $n \geq N$. Differentiating (B.13) with respect to t obtain that

$$\dot{L}_t(p) = \frac{A(\pi_t(L_t(p))) + rV - A(\pi_t(m_t))}{\pi_t'(L_t(p)) \cdot \partial_2 U(\pi_t(p), \pi_t(L_t(p)))}. \tag{B.14}$$

Since $m_{\bar{t}} = l_{\bar{t}} = L_{\bar{t}}(p)$ for every $p \in S^I(\bar{t})$, it follows that $\frac{\partial}{\partial p}\dot{L}_{\bar{t}}(p) < 0$. This implies that there exists $\varepsilon > 0$ such that $\dot{L}_t(p)$ is strictly decreasing in p for every $t \in (\bar{t} - \varepsilon, \bar{t})$. This, in turn, implies that $L_t(p)$ is strictly increasing in p_0 for every $t \in (\bar{t} - \varepsilon, \bar{t})$, leading to a contradiction with the fact that $L_{t_n}(p) = l_t$ for every $p \in S^I(t_n)$ and $n \geq N$. This proves that if $|S(\bar{t})| > 1$, then there exists a $t_S < \bar{t}$ such that $(t_S, \bar{t}) \cap \mathbb{T} = \emptyset$.

Off-path beliefs. Suppose there exists a $B_\epsilon(t_0) \in (t_S, \bar{t})$ such that $l_t \neq \inf S(\bar{t})$. Following Lemma A.27 it follows that $l_t \in (\inf S(\bar{t}), \sup S(\bar{t}))$. To satisfy the lowest continuation surplus refinement 4, the off-path belief l_t solves

$$V_t(l_t) = \min_{x \in R(t)} V_t(x) = \min_{x \in S(\bar{t})} V_t(x). \tag{B.15}$$

All types $p \in S^I(\bar{t})$ find it weakly optimal to wait until time $\bar{t}$ to separate from the intermediary. This implies that function $V_t(p)$ is linear in $\pi_t(p)$ for $p \in S^I(\bar{t})$. As $\pi_t(p)$ is increasing in p, it implies that the solution to (B.15) is interior if and only if $V_t'(p) = 0$ for each $p \in S^I(\bar{t})$ and $t \in B_\epsilon(t_0)$. Differentiating the definition of $V_t(p)$ in (B.4) obtain that $V_t'(p) = 0$ for $p \in S^I(\bar{t})$ if and only if l_t solves

$$\begin{aligned} u_1(\pi_t(l_t)) - u_0(\pi_t(l_t)) = &\int_t^{\bar{t}} \left(e^{-r(s-t)} - e^{-(r+\lambda)(s-t)}\right) \cdot \left(A(\pi_s(m_{\bar{t}})) - rV\right) ds \\ &+ \int_{\bar{t}}^{\infty} \left(e^{-r(s-t)} - e^{-(r+\lambda)(s-t)}\right) \cdot A(\pi_s(m_{\bar{t}}))\, ds. \end{aligned} \tag{B.16}$$

The solution l_t to (B.16) is differentiable for $t \in B_\epsilon(t_0)$. Note that $V_t'(p) = 0$ for all $t \in B_\epsilon(t_0)$ requires that $\dot{V}_t'(p) = 0$. Differentiating (B.12) with respect to p, obtain

$$0 = \pi_t'(p) \cdot \left[u_1'(\pi_t(l_t)) - u_0'(\pi_t(l_t))\right] \cdot \pi_t'(l_t) \cdot \dot{l}_t \qquad \Leftrightarrow \qquad \dot{l}_t = 0$$

for every $t \in B_\epsilon(t_0)$, implying that $l_t = l_{t_0}$ for every $t \in B_\epsilon(t_0)$. Using this, rewrite (B.16) as

$$rV \cdot \int_t^{\bar{t}} \left(e^{-rs} - e^{-rs-\lambda(s-t)}\right) ds = \int_t^{\infty} \left(e^{-rs} - e^{-rs-\lambda(s-t)}\right) \cdot \left(A(\pi_s(m_{\bar{t}})) - A(\pi_s(l_{t_0}))\right) ds \tag{B.17}$$

Equality (B.17) holds for every $t \in B_\epsilon(t_0)$ implying that the derivatives with respect to t of the left and right-hand sides must be equal:

$$\begin{aligned} -r\lambda V \cdot \int_t^{\bar{t}} e^{-(r+\lambda)(s-t)}\, ds &= -\lambda \int_t^{\infty} e^{-rs-\lambda(s-t)} \cdot \left(A(\pi_s(m_{\bar{t}})) - A(\pi_s(l_{t_0}))\right) ds \\ rV \cdot \int_t^{\bar{t}} e^{-(r+\lambda)(s-t)}\, ds &= \int_t^{\infty} e^{-rs-\lambda(s-t)} \cdot \left(A(\pi_s(m_{\bar{t}})) - A(\pi_s(l_{t_0}))\right) ds. \end{aligned}$$

$$rV \cdot \int_t^{\bar{t}} e^{-(r+\lambda)s}\, ds = \int_t^{\infty} e^{-(r+\lambda)s} \cdot \left(A(\pi_s(m_{\bar{t}})) - A(\pi_s(l_{t_0}))\right)\, ds. \tag{B.18}$$

Conditions (B.17) and (B.18) are equivalent to

$$\begin{cases} rV \cdot \int_t^{\bar{t}} e^{-rs}\, ds = \int_t^{\infty} e^{-rs} \cdot \left(A(\pi_s(m_{\bar{t}})) - A(\pi_s(l_{t_0}))\right)\, ds, \\ rV \cdot \int_t^{\bar{t}} e^{-(r+\lambda)s}\, ds = \int_t^{\infty} e^{-(r+\lambda)s} \cdot \left(A(\pi_s(m_{\bar{t}})) - A(\pi_s(l_{t_0}))\right)\, ds \end{cases} \tag{B.19}$$

Conditions (B.19) must be satisfied for all $t \in B_\epsilon(t_0)$, which requires that

$$A(\pi_t(m_{\bar{t}})) - A(\pi_t(l_{t_0})) - rV = 0 \qquad \forall t \in B_\epsilon(t_0). \tag{B.20}$$

If (B.20) is satisfied, then the indifference condition (B.16) requires that

$$\int_{\bar{t}}^{\infty} \left(e^{-r(s-t)} - e^{-(r+\lambda)(s-t)}\right) \cdot \left(A(\pi_s(m_{\bar{t}})) - A(\pi_s(l_{t_0}))\right) = 0,$$

which contradicts $l_{t_0} < m_{\bar{t}}$ obtained in Lemma A.27. This leads to a contradiction with the existence of a $B_\epsilon(t_0) \subset (t_S, \bar{t})$ such that the indifference condition (B.16) is satisfied. By right continuity of the belief process l_t, it implies that $l_t = \min S(t)$ for all $t \in [t_S, \bar{t})$. □

Lemma B.6 (Beliefs at t_S)**.** *Suppose belief process $(l_t)_{t \geq 0}$ is right continuous and satisfies the lowest continuation surplus refinement 4. Then $S(t_S) = \min S(\bar{t})$ and $A(\pi_t(m_t)) - A(\pi_t(l_t)) \leq rV$ for all $t \in [t_S, \bar{t}]$, with the inequality being strict for some $t \in [t_S, \bar{t})$.*

Proof. Following Lemma B.5, the off-path belief during the quiet period $(t_S, \bar{t})$ is $l_t = \min S(\bar{t})$. The restriction on beliefs to be right continuous then requires that $l_{t_S} = \min\{S(\bar{t})\}$. The optimality of stopping at time t_S requires that

$$A(q_{t_S}) - A(\pi_{t_S}(l_{t_S})) \leq rV.$$

Since $(t_S, \bar{t}) \cap \mathbb{T} = \emptyset$ and $A(\pi_t(p))$ is weakly concave in t, it follows that

$$A(\pi_t(m_t)) - A(\pi_t(l_{t_S})) \leq rV \qquad \forall t \in [t_S, \bar{t}]. \tag{B.21}$$

Suppose that (B.21) holds with equality for all $t \in [t_S, \bar{t})$. This, however, implies a contradiction with the optimality of separation at time t_S as by waiting until $\bar{t}$ the agents can be perceived as $l_{\bar{t}} = m_{\bar{t}} > l_{t_S}$ following Lemma A.27. Consequently, (B.21) must be strict for some $t \in [t_S, \bar{t})$.

Suppose that $|S(t_S)| > 1$. It follows from Bayesian consistency that $\mathrm{E}\left[\tilde{p} \,|\, \tau = t_S, X_{t_S} = \mu t_S\right] = l_{t_S} = \min S(\bar{t})$. This implies that there exists a $p' > l_{t_S} \in S^I(t_S)$, implying that $|S^I(t_S) \cap S^I(\bar{t})| > 1$. It then follows from Lemma B.1 that $S^I(t_S) = S^I(\bar{t})$. This implies that the indifference belief, defined in (B.13) satisfies $L_{t_S}(p) = l_{t_S}$ and $L_{\bar{t}}(p) = l_{\bar{t}}$ for all $p \in S^I(t_S) = S^I(\bar{t})$. The dynamics of the indifference belief process $L_t(p)$ in (B.14) combined with the necessary condition (B.21) implies that $L_{t_S}(p)$ is strictly increasing in p for all $s \in [t_S, \bar{t})$. This leads to a contradiction with $S^I(t_S) = S^I(\bar{t})$. □

Lemma B.7 (Kink at t_S)**.** *The joint welfare function $W_t(p)$ has a positive kink at $p = l_{t_S}$, i.e., $W_t'(l_{t_S}-) < W_t'(l_{t_S}+)$.*

Proof. The joint welfare function $W_t(p)$ is linear for $p \in [l_{t_S}, \overline{p}]$. If it were the case that $W_t'(l_{t_S}-) = W_t'(l_{t_S}+)$, then it would imply that types $t \in (l_{t_S}, \overline{p}]$ are indifferent between separating at time t_S and waiting until $\bar{t}$. This, however, contradicts the second part of Lemma B.6 that waiting between t_S and $\bar{t}$ is costly. □

Denote by t_Q the last instance prior to t_S such that there does not exist any separations prior to t_Q, i.e.,

$$t_Q \overset{def}{=} \sup\left\{t \le t_S : \quad \exists \varepsilon > 0 \quad \text{s.t.} \quad (t-\varepsilon, t) \notin \mathbb{T}^I\right\}. \tag{B.22}$$

Lemma B.8 (Separation during (t_Q, t_S))**.** *Suppose $(l_t)_{t\ge 0}$ is right continuous and the equilibrium satisfies the lowest continuation surplus refinement 4. Then there is no pooling for any $t \in [t_Q, t_S]$, i.e., $|S(t)| = 1$.*

Proof. If $t_Q = t_S$, then the statement of Lemma B.8 follows from Lemma B.6. Suppose $t_Q < t_S$. By definition of t_Q, the set of separating times $\mathbb{T}^I$ is dense in (t_Q, t_S). The optimality of separations over the dense period (t_Q, t_S) requires that l_t is continuous for $t \in (t_Q, t_S)$. The first order optimality condition for type $p \in S(t)$ is

$$A(\pi_t(m_t)) - rV - A(\pi_t(l_t)) + \dot{l}_t \cdot \pi_t'(l_t) \cdot \partial_2 U(\pi_t(p), \pi_t(l_t)) = 0. \tag{B.23}$$

Optimality condition (B.23) then also requires for l_t to be differentiable for all $t \in (t_Q, t_S)$. Solving (B.23) for $\dot{l}_t$ obtain

$$\dot{l}_t = \frac{A(\pi_t(l_t)) + rV - A(\pi_t(m_t))}{\pi_t'(l_t) \cdot \partial_2 U(\pi_t(p), \pi_t(l_t))}. \tag{B.24}$$

Moreover, the linearity of $U(\pi_t(p), \pi_t(l))$ in $\pi_t(p)$ implies that process l_t satisfies

$$\dot{l}_t = \frac{A(\pi_t(l_t)) + rV - A(\pi_t(m_t))}{\pi_t'(l_t) \cdot \partial_2 U(\pi_t(l_t), \pi_t(l_t))}. \tag{B.25}$$

The local indifference condition for the separating type $p = l_t$ requires that

$$\dot{l}_t = \frac{A(\pi_t(l_t)) + rV - A(\pi_t(m_t))}{\pi_t'(l_t) \cdot \partial_2 U(\pi_t(l_t), \pi_t(l_t))} \qquad \forall t \in [t_Q, t_S]. \tag{B.26}$$

(i) Suppose $A(\pi_t(l_t)) + rV - A(\pi_t(m_t)) \neq 0$ for a given t. In this case, local optimality condition (B.23) cannot be satisfied by multiple types $p \in S^I(t)$ due to strict monotonicity of $\partial_2 U(\pi_t(p), \pi_t(l_t))$ in

p. The strict optimality of separation times during $[t_Q, t_S]$ implies that the joint welfare $W_0(p)$ is strictly convex at $p = l_t$, implying that $l_t \notin S^I(\bar{t})$ for $t < t_S$. From here, it follows that $\dot{l}_{t_S} > 0$ since, otherwise, it would contradict the linearity of $W_0(p)$ for $p \in S^I(\bar{t})$.

(ii) Consider now the set of times

$$\hat{\mathbb{T}} \stackrel{def}{=} \{t \in [t_Q, t_S] : \quad A(\pi_t(l_t)) + rV - A(\pi_t(m_t)) = 0\}.$$

Given the continuity of l_t for $t \in [t_Q, t_S]$ it follows set $\hat{\mathbb{T}}$ is closed. It follows from (B.25) that $\dot{l}_t = 0$ for every $t \in \hat{\mathbb{T}}$. Consider two cases.

(i) Suppose set $\hat{\mathbb{T}}$ is nowhere dense. Then for every $\hat{t} \in \hat{\mathbb{T}}$ there exist a $\hat{t}_\varepsilon \notin \hat{\mathbb{T}}$ such that $|\hat{t} - \hat{t}_\varepsilon| < \varepsilon$ for every $\varepsilon > 0$. This implies that $W_0(p)$ is strictly convex at every $p = l_{\hat{t}_\varepsilon}$, implying that there cannot be pooling at $\hat{t}_\varepsilon$.

(ii) Suppose segment $[t_1, t_2] \subseteq [t_Q, t_S]$ is dense in $\hat{\mathbb{T}}$. By continuity of l_t over $[t_Q, t_S]$ it follows that $[t_1, t_2] \in \hat{\mathbb{T}}$. It follows from (B.25) that $\dot{l}_t = 0$ for all $t \in [t_1, t_2]$, implying that $l_t = l_{t_1}$ for every $t \in [t_1, t_2]$. This implies that $S^I(t) = S^I(t_1) = S^I(t_2)$ for any $t \in (t_1, t_2)$. Suppose that $|S^I(t_2)| > 1$ meaning that pooling is incentive compatible during the period $[t_1, t_2]$. Since Lemma B.6 proves that $|S^I(t_S)| = 1$, it follows that $t_2 < t_S$. This implies that the joint welfare function $W_0(p)$ is strictly convex over $[l_{t_2}, l_{t_S}]$. This implies that $S^I(t_1) = S^I(t_2) < l_{t_2}$. Consequently, if there is any pooling during $[t_1, t_2]$, it would lead to a positive jump in beliefs at t_2, which contradicts continuity of l_t during $[t_Q, t_S]$. Consequently, there cannot be pooling during the period $[t_1, t_2]$.

□

Lemma B.9 (Uniqueness of smooth separation times)**.** *There does not exist a $B_\varepsilon(p_0) \subset \left[l_{t_Q}, l_{t_S}\right]$ and times $B_{\varepsilon_1}(t_1)$ and $B_{\varepsilon_2}(t_2)$, such that $B_{\varepsilon_1}(t_1) \cap B_{\varepsilon_2}(t_2) = \emptyset$, for which*

(i) the set of separation times $\cup_{p \in B_\varepsilon(p_0)} \mathbb{T}^I(p)$ is dense in $B_{\varepsilon_i}(t_i)$ for $i \in \{1, 2\}$;

(ii) each $B_{\varepsilon_i}(t_i)$ contains a weakly optimal separation time, i.e., $\mathbb{T}^I(p) \cap B_{\varepsilon_i}(t_i) \neq \emptyset$ for each $p \in B_\varepsilon(p_0)$;

(iii) the first order condition holds at each $t \in \left\{\cup_{p \in B_\varepsilon(p_0)} \mathbb{T}^I(p)\right\} \cap \{B_{\varepsilon_1}(t_1) \cup B_{\varepsilon_2}(t_2)\}$, i.e.,

$$A(\pi_t(m_t)) - rV - A(\pi_t(l_t)) + \dot{l}_t \cdot \pi_t'(l_t) \cdot \partial_2 U(\pi_t(l_t), \pi_t(l_t)) = 0. \tag{B.27}$$

Proof. The optimal separation time solves $\tau(p)$ for an (ex-ante) type p solves

$$W_0(p) = \sup_\tau \bigg\{ \int_0^\tau e^{-rt} \cdot \Big(p + (1-p) \cdot e^{-\lambda t}\Big) \cdot (A(\pi_t(m_t)) - rV)\, dt$$

$$
\begin{aligned}
&+e^{-r\tau} \cdot \Big[p \cdot u_1(\pi_\tau(l_\tau)) + (1-p)e^{-\lambda\tau} \cdot u_0(\pi_\tau(l_\tau))\Big] + V\Big\} \\
&\overset{(i)}{=} \left\{\int_0^\tau e^{-rt} \cdot \Big(p + (1-p)e^{-\lambda t}\Big) \cdot (A(\pi_t(m_t)) - rV - A(\pi_t(p)))\, dt\right\} + U(p,p) + V,
\end{aligned} \tag{B.28}
$$

where equality (i) holds because, following Lemma B.8, $l_{t(p)} = p$ for all $t(p) \in [t_Q, t_S]$.

Suppose from the contrary that there exist $B_\varepsilon(p_0)$ and $B_{\varepsilon_1}(t_1)$ and $B_{\varepsilon_2}(t_2)$ satisfying the conditions (i), (ii), and (iii) of Lemma B.9. If $p \in B_\varepsilon(p_0)$, condition (ii) implies that there exist two solutions $t_1(p) < t_2(p)$ to (B.28) such that $t_1(p) \in B_{\varepsilon_1}(p)$ and $t_2(p) \in B_{\varepsilon_2}(p)$. As both stopping times are optimal, and lead to the same expected joint welfare, it implies that

$$
0 = \int_{t_1(p)}^{t_2(p)} e^{-rt} \cdot \Big(p + (1-p) \cdot e^{-\lambda t}\Big) \cdot [A(\pi_t(m_t)) - rV - A(\pi_t(p))]\, dt. \tag{B.29}
$$

Following the argument of Lemma B.8, conditions (i) and (iii) require that $t_1(p)$ and $t_2(p)$ are continuous in p for $p \in B_\varepsilon(p_0)$. Differentiate identity (B.29) with respect to p to obtain

$$
\begin{aligned}
0 =& \int_{t_1(p)}^{t_2(p)} e^{-rt} \cdot \Big(1 - e^{-\lambda t}\Big) \cdot [A(\pi_t(m_t)) - rV - A(\pi_t(p))]\, dt \\
&- \int_{t_1(p)}^{t_2(p)} e^{-rt} \cdot \Big(p + (1-p) \cdot e^{-\lambda t}\Big) \cdot A'(\pi_t(p)) \cdot \pi_t'(p)\, dt \\
&+ e^{-rt_2(p)} \cdot \Big(p + (1-p)e^{-\lambda t_2(p)}\Big) \cdot \Big[A\left(\pi_{t_2(p)}(m_{t_2(p)})\right) - rV - A(\pi_{t_2(p)}(p))\Big] \cdot t_2'(p) \\
&- e^{-rt_1(p)} \cdot \Big(p + (1-p)e^{-\lambda t_1(p)}\Big) \cdot \Big[A\left(\pi_{t_1(p)}(m_{t_1(p)})\right) - rV - A(\pi_{t_1(p)}(p))\Big] \cdot t_1'(p) \\
\overset{(i)}{=}& \int_{t_1(p)}^{t_2(p)} e^{-rt} \cdot \Big(1 - e^{-\lambda t}\Big) \cdot [A(\pi_t(m_t)) - rV - A(\pi_t(p))]\, dt \\
&- \int_{t_1(p)}^{t_2(p)} e^{-rt} \cdot \Big(p + (1-p) \cdot e^{-\lambda t}\Big) \cdot A'(\pi_t(p)) \cdot \pi_t'(p)\, dt \\
&- e^{-rt_2(p)} \cdot \Big(p + (1-p)e^{-\lambda t_2(p)}\Big) \cdot \pi_{t_2(p)}'(p) \cdot \partial_2 U(\pi_{t_2(p)}(p), \pi_{t_2(p)}(p)) \\
&+ e^{-rt_1(p)} \cdot \Big(p + (1-p)e^{-\lambda t_1(p)}\Big) \cdot \pi_{t_1(p)}'(p) \cdot \partial_2 U(\pi_{t_1(p)}(p), \pi_{t_1(p)}(p)),
\end{aligned}
$$

where equality (i) holds by condition (iii) of Lemma B.9. Note that

$$
\begin{aligned}
& e^{-rt} \cdot \Big(p + (1-p)e^{-\lambda t}\Big) \cdot \pi_t'(p) \cdot \partial_2 U(\pi_t(p), \pi_t(p)) \\
=& e^{-rt} \cdot \Big(p + (1-p)e^{-\lambda t}\Big) \cdot \frac{e^{-\lambda t}}{(p + (1-p)e^{-\lambda t})^2} \\
\times& \int_0^\infty e^{-rs} \Big(\pi_t(p) + (1-\pi_t(p))e^{-\lambda s}\Big) \cdot A'(\pi_{t+s}(p)) \cdot \frac{e^{-\lambda s}}{(\pi_t(p) + (1-\pi_t(p))e^{-\lambda s})^2}\, ds
\end{aligned}
$$

$$
\begin{aligned}
=&\frac{e^{-\lambda t}}{p+(1-p)e^{-\lambda t}} \cdot \int_0^{\infty} e^{-r(s+t)} \cdot \frac{p+(1-p)e^{-\lambda t}\cdot e^{-\lambda s}}{p+(1-p)e^{-\lambda t}} \cdot \frac{A'(\pi_{t+s}(p))\cdot e^{-\lambda s}}{\left(\frac{p}{p+(1-p)e^{-\lambda t}}+\frac{(1-p)e^{-\lambda t}}{p+(1-p)e^{-\lambda t}}e^{-\lambda s}\right)^2}\, ds \\
=&\int_0^{\infty} e^{-r(s+t)} \left(p+(1-p)e^{-\lambda(s+t)}\right) \cdot \frac{A'(\pi_{t+s}(p))\cdot e^{-\lambda(s+t)}}{(p+(1-p)e^{-\lambda(s+t)})^2}\, ds \\
=&\int_0^{\infty} e^{-r(s+t)} \cdot \frac{A'(\pi_{t+s}(p))\cdot e^{-\lambda(s+t)}}{p+(1-p)e^{-\lambda(s+t)}}\, ds = \int_t^{\infty} e^{-rs} \cdot \frac{A'(\pi_s(p))\cdot e^{-\lambda s}}{p+(1-p)e^{-\lambda s}}\, ds.
\end{aligned}
$$

The above equality implies that

$$
\begin{aligned}
0 =& \int_{t_1(p)}^{t_2(p)} e^{-rt} \cdot \left(1-e^{-\lambda t}\right) \cdot \left[A(\pi_t(m_t)) - rV - A(\pi_t(p))\right] dt \\
&- \int_{t_1(p)}^{t_2(p)} e^{-rt} \cdot \left(p+(1-p)\cdot e^{-\lambda t}\right) \cdot A'(\pi_t(p)) \cdot \pi_t'(p)\, dt \\
&- e^{-rt_2(p)} \cdot \left(p+(1-p)e^{-\lambda t_2(p)}\right) \cdot \pi'_{t_2(p)}(p) \cdot \partial_2 U(\pi_{t_2(p)}(p), \pi_{t_2(p)}(p)) \\
&+ e^{-rt_1(p)} \cdot \left(p+(1-p)e^{-\lambda t_1(p)}\right) \cdot \pi'_{t_1(p)}(p) \cdot \partial_2 U(\pi_{t_1(p)}(p), \pi_{t_1(p)}(p)) \\
0 =& \int_{t_1(p)}^{t_2(p)} e^{-rt} \cdot \left(1-e^{-\lambda t}\right) \cdot \left[A(\pi_t(m_t)) - rV - A(\pi_t(p))\right] dt \\
&- \int_{t_1(p)}^{t_2(p)} e^{-rt} \cdot A'(\pi_t(p)) \cdot \frac{e^{-\lambda t}}{p+(1-p)e^{-\lambda t}}\, dt \\
&- \int_{t_2(p)}^{+\infty} e^{-rt} \cdot A'(\pi_t(p)) \cdot \frac{e^{-\lambda t}}{p+(1-p)e^{-\lambda t}}\, dt + \int_{t_1(p)}^{+\infty} e^{-rt} \cdot A'(\pi_t(p)) \cdot \frac{e^{-\lambda t}}{p+(1-p)e^{-\lambda t}}\, dt \\
0 =& \int_{t_1(p)}^{t_2(p)} e^{-rt} \cdot \left(1-e^{-\lambda t}\right) \cdot \left[A(\pi_t(m_t)) - rV - A(\pi_t(p))\right] dt. \qquad \text{(B.30)}
\end{aligned}
$$

Combine (B.29) and (B.30) to obtain

$$
\begin{cases}
\int_{t_1(p)}^{t_2(p)} e^{-rt} \cdot \left[A(\pi_t(m_t)) - rV - A(\pi_t(p))\right] dt = 0, \\
\int_{t_1(p)}^{t_2(p)} e^{-(r+\lambda)t} \cdot \left[A(\pi_t(m_t)) - rV - A(\pi_t(p))\right] dt = 0.
\end{cases} \qquad \text{(B.31)}
$$

which holds for all $p \in B_\varepsilon(p_0)$. Differentiate the top equality in (B.31) with respect to p to obtain

$$
\begin{aligned}
0 =& e^{-rt_2(p)} \cdot \left[A\left(\pi_{t_2(p)}(m_{t_2(p)})\right) - rV - A(\pi_{t_2(p)}(p))\right] \cdot t_2'(p) \\
&- e^{-rt_1(p)} \cdot \left[A\left(\pi_{t_1(p)}(m_{t_1(p)})\right) - rV - A(\pi_{t_1(p)}(p))\right] \cdot t_1'(p) \\
&- \int_{t_1(p)}^{t_2(p)} e^{-rt} \cdot A'(\pi_t(p)) \cdot \pi_t'(p)\, dt \\
0 =& -e^{-rt_2(p)} \cdot \pi'_{t_2(p)}(p) \cdot \partial_2 U(\pi_{t_2(p)}(p), \pi_{t_2(p)}(p))
\end{aligned}
$$

$$
\begin{aligned}
&+ e^{-rt_1(p)} \cdot \pi'_{t_1(p)}(p) \cdot \partial_2 U(\pi_{t_1(p)}(p), \pi_{t_1(p)}(p)) \\
&- \int_{t_1(p)}^{t_2(p)} e^{-rt} \cdot A'(\pi_t(p)) \cdot \frac{e^{-\lambda t}}{(p+(1-p)e^{-\lambda t})^2}\, dt \\
0 = &- \frac{1}{p+(1-p)e^{-\lambda t_2(p)}} \cdot \int_{t_2(p)}^{\infty} e^{-rt} \cdot \frac{A'(\pi_t(p)) \cdot e^{-\lambda t}}{p+(1-p)e^{-\lambda t}}\, dt \\
&+ \frac{1}{p+(1-p)e^{-\lambda t_1(p)}} \cdot \int_{t_1(p)}^{\infty} e^{-rt} \cdot \frac{A'(\pi_t(p)) \cdot e^{-\lambda t}}{p+(1-p)e^{-\lambda t}}\, dt \\
&- \int_{t_1(p)}^{t_2(p)} e^{-rt} \cdot A'(\pi_t(p)) \cdot \frac{e^{-\lambda t}}{(p+(1-p)e^{-\lambda t})^2}\, dt \\
0 = &\int_{t_1(p)}^{t_2(p)} e^{-(r+\lambda)t} \cdot \frac{A'(\pi_t(p))}{p+(1-p)e^{-\lambda t}} \cdot \underbrace{\left(\frac{1}{p+(1-p)e^{-\lambda t_1(p)}} - \frac{1}{p+(1-p)e^{-\lambda t}} \right)}_{<0} dt \\
&+ \underbrace{\left(\frac{1}{p+(1-p)e^{-\lambda t_1(p)}} - \frac{1}{p+(1-p)e^{-\lambda t_2(p)}} \right)}_{<0} \cdot \int_{t_2(p)}^{\infty} e^{-(r+\lambda)t} \cdot \frac{A'(\pi_t(p))}{p+(1-p)e^{-\lambda t}}\, dt.
\end{aligned}
$$

The above inequality cannot be satisfied since both terms on the right hand side are negative. This leads to a contradiction with the existence of $t_1(p)$ and $t_2(p)$ which are solutions (B.28), all the while the first order optimal stopping conditions (iii) in the Lemma hold.

□

Lemma B.10 (Increasing separations between t_Q and t_S)**.** *Gradual separations during $[t_Q, t_S]$ are increasing in types, as captured by $\dot{l}_t \geq 0$ for all $t \in (t_Q, t_S)$.*

Proof. Following Lemma B.8, process l_t is differentiable for $t \in [t_Q, t_S]$. Suppose that l_t is non-monotone, meaning that there exists a local minimum or maximum. Without loss, suppose l_t has a strict local minimum at $t_0 \in (t_Q, t_S)$. This implies that there exists a pair $t_1 < t_0 < t_2$ such that

$$
l_{t_1} > l_{t_0} > l_{t_2}.
$$

Consider the type

$$
\hat{p} = \frac{l_{t_0} + \min\{l_{t_1}, l_{t_2}\}}{2}.
$$

Then there exists a $\hat{t}_1 < t_0 < \hat{t}_2$ such that $l_{\hat{t}_1} = l_{\hat{t}_2} = \hat{p}$. This, however, implies a contradiction with Lemma B.9 as there exists a neighborhood $B_\varepsilon(\hat{p})$ that find it optimal to separate both during $B_{\epsilon_1}(\hat{t}_1)$ and $B_{\epsilon_2}(\hat{t}_2)$ with $B_{\epsilon_1}(\hat{t}_1) \cap B_{\epsilon_2}(\hat{t}_2) = \emptyset$ and the first order optimality condition (B.27) holding due to Lemma B.8. This is a contradiction with t_0 being a local maximum of l_t during $[t_Q, t_S]$.

Suppose belief process l_t is decreasing during $[t_Q, t_S]$, implying that $l_{t_Q} > l_{t_S}$. Due to the first order separating condition (B.23) is stricly increasing in type p, it implies that the joint welfare function $W_0(p)$

is strictly convex for $p \in [l_{t_Q}, l_{t_S}]$. Following Lemma B.6, however, this welfare function $W_0(p)$ is linear for $p \in [l_{t_S}, \bar{p}]$, due to the indifference of type l_{t_S} in waiting until time $\bar{t}$ to separate. If $l_{t_Q} > l_{t_S}$, then $[l_{t_Q}, l_{t_S}] \subseteq [l_{t_S}, \bar{p}]$, which leads to a contradiction with l_t being decreasing over $[t_Q, t_S]$. $\square$

Define t^* to be the last time before t_Q when a type is willing to separate

$$t^* \overset{def}{=} \sup\left\{t < t_Q : \quad t \in \mathbb{T}^I\right\} \in [0, t_Q). \tag{B.32}$$

By definition of t_Q, it follows that $t^* < t_Q$ if such a t^* exists.

Lemma B.11 (Quiet period beliefs)**.** *Suppose the equilibrium satisfies the lowest continuation surplus refinement 4 and belief process $l = (l_t)_{t\geq 0}$ is right continuous. If there exists $t^* \geq 0$ then $l_{t^*} \in (l_{t_Q}, l_{t_S}]$ and process $l = (l_t)_{t\geq 0}$ satisfies*

$$\int_t^{t(l_t)} \left(e^{-r(s-t)} - e^{-(r+\lambda)(s-t)}\right) \cdot (A(\pi_s(m_s)) - rV - A(\pi_s(l_t)))\, ds = 0 \tag{B.33}$$

and

$$\operatorname{sign}\left(\dot{l}_t\right) = \operatorname{sign}\left\{\int_t^{t(l_t)} e^{-r(s-t)} \cdot (A(\pi_s(m_s)) - rV - A(\pi_s(l_t)))\, ds\right\}. \tag{B.34}$$

for all $t \in (t^*, t^* + \varepsilon)$ for some $\varepsilon > 0$.

Proof. The joint continuation welfare (B.2) is given by

$$\begin{aligned} W_t(p) = \sup_{\tau}\Bigg\{&\int_t^{\tau} e^{-r(s-t)} \cdot \left(\pi_t(p) + (1 - \pi_t(p)) \cdot e^{-\lambda(s-t)}\right) \cdot (A(\pi_s(m_s)) - rV)\, ds \\ &+ e^{-r(\tau-t)} \cdot \left[\pi_t(p) \cdot u_1(\pi_\tau(l_\tau)) + (1 - \pi_t(p)) \cdot e^{-\lambda(\tau-t)} \cdot u_0(\pi_\tau(l_\tau))\right]\Bigg\} + V. \end{aligned}$$

Following Lemma B.8, this joint welfare $W_t(p)$ is strictly convex in $\pi_t(p)$ for $p \in (l_{t_Q}, l_{t_S})$ as each type has a uniquely optimal date over when to separate. The off-path belief l_t solves

$$V_t(l_t) = W_t(l_t) - U_t(l_t) = \min_{x\in R(t)} V_t(x) = \min_{x\in R(t)} (W_t(x) - U_t(x)) = \min_{x\in[l_{t_Q}, \bar{p}]} (W_t(x) - U_t(x)). \tag{B.35}$$

Suppose that $l_{t^*} = l_{t_Q}$. Following Lemmas B.5, A.27, and B.10 it implies that the set of remaining types $R(t^*) \subseteq [l_{t_Q}, \bar{p}]$. From the lowest continuation surplus refinement, it implies that $l_t \geq l_{t_Q}$ for $t \geq t^*$. In order for type $p = l_{t_Q}$ to find it incentive compatible to separate at time t^* rather than at time $t^* + \epsilon$ it must be the case that

$$A(\pi_{t^*}(m_{t^*}) - rV - A(\pi_{t^*}(l_{t^*})) \leq 0. \tag{B.36}$$

For this same type $p = l_{t_Q}$ to be willing to wait until time t_Q, it then is necessary that (B.36) is binding for all $t \in [t^*, t_Q]$ due to concavity of $A(\pi_t(\underline{p}))$ in t. This, however, contradicts the definition of (B.32) as it means that type $p = l_{t_Q}$ finds it weakly optimal to separate anytime during $[t^*, t_Q]$. This implies a contradiction with $l_{t^*} = l_{t_Q}$.

Suppose that $l_{t^*} > l_{t_S}$. By right continuity there exists an $\varepsilon > 0$ such that $l_t > l_{t_S}$ for every $t \in [t^*, t^*+\varepsilon)$. This implies that $V_t'(p) = 0$ for all $p \in (l_{t_S}, \bar{p}]$ and $t \in [t^*, t^* + \varepsilon)$. This, however, contradicts the argument of Lemma B.5 which showed that there does not existn $l_t \in (l_{t_S}, \bar{p}]$ such that $V_t'(l_t) = 0$ for t in a positive interval.

Suppose $l_{t^*} = l_{t_S}$. Suppose that there exists a sequence $\{t_n\}_{n=1}^{\infty}$ such that $t_n > t^*$, $l_{t_n} \to l_{t^*}$, and $l_{t_n} \geq l_{t_S}$. In order for $t^* = t_S \in \mathbb{T}^I$ it is then necessary that

$$A(\pi_{t^*}(m_{t^*})) - rV - A(\pi_{t^*}(l_{t^*})) \leq 0 \tag{B.37}$$

since, otherwise, at type t_S would find is strictly preferable to separate at time t_n for n high enough. If, however, (B.37) is satisfied, then type l_{t_Q} is strictly better of separating at time t^* and be perceived as $l_{t_S} > l_{t_Q}$, rather than wait until t_Q as can be seen from

$$\begin{aligned}
W_{t^*}(l_{t_Q}) = & \int_{t^*}^{t_Q} \left(\pi_{t^*}(l_{t_Q}) \cdot e^{-r(t-t^*)} + (1 - \pi_{t^*}(l_{t_Q})) \cdot e^{-(r+\lambda)(t-t^*)}\right) \cdot (A(\pi_t(m_t)) - rV)\; dt \\
& + \int_{t_Q}^{\infty} \left(\pi_{t^*}(l_{t_Q}) \cdot e^{-r(t-t^*)} + (1 - \pi_{t^*}(l_{t_Q})) \cdot e^{-(r+\lambda)(t-t^*)}\right) \cdot A(\pi_t(l_{t_Q}))\, dt + V \\
& \overset{(i)}{=} \int_{t^*}^{t_Q} \left(\pi_{t^*}(l_{t_Q}) e^{-r(t-t^*)} + (1 - \pi_{t^*}(l_{t_Q})) e^{-(r+\lambda)(t-t^*)}\right) \underbrace{(A(\pi_t(m_{t^*})) - rV - A(\pi_t(l_{t^*})))}_{\leq 0}\, dt \\
& + \int_{t_Q}^{\infty} \left(\pi_{t^*}(l_{t_Q}) \cdot e^{-r(t-t^*)} + (1 - \pi_{t^*}(l_{t_Q})) \cdot e^{-(r+\lambda)(t-t^*)}\right) \underbrace{(A(\pi_t(l_{t_Q})) - A(\pi_t(l_{t^*})))}_{<0}\, dt \\
& + \int_{t^*}^{\infty} \left(\pi_{t^*}(l_{t_Q}) \cdot e^{-r(t-t^*)} + (1 - \pi_{t^*}(l_{t_Q})) \cdot e^{-(r+\lambda)(t-t^*)}\right) \cdot A(\pi_t(l_{t^*}))\, dt + V \\
& \overset{(ii)}{<} \int_{t^*}^{\infty} \left(\pi_{t^*}(l_{t_Q}) \cdot e^{-r(t-t^*)} + (1 - \pi_{t^*}(l_{t_Q})) \cdot e^{-(r+\lambda)(t-t^*)}\right) \cdot A(\pi_t(l_{t^*}))\, dt + V.
\end{aligned}$$

Equality (i) follows from the fact that $(t^*, t_Q) \not\subset \mathbb{T}$ and hence $m_t = m_{t^*}$ during that period. Strict inequality (ii) follows from weak inequality (B.37), combined with weak concavity $A(\pi_t(\underline{p}))$ in t and strict inequality $l_{t_S} > l_{t_Q}$ which follows from Lemma B.10.

Remaining cases. The above arguments imply that $l_{t^*} \leq l_{t_S}$ and $l_t \in (l_{t_Q}, l_{t_S})$ for all $t \in [t^*, t^* + \varepsilon)$ for some $\epsilon > 0$. This implies that for each $t \in (t^*, t^* + \varepsilon)$ the off-equilibrium path belief l_t must be an interior solution to (B.35) due to strict convexity of $W_t(p)$ for $p \in (l_{t_Q}, l_{t_S})$ as derived in Lemma B.8. Such an interior condition must satisfy

$$W_t'(l_t) = U_t(l_t) \qquad \Leftrightarrow \qquad W_t'(l_t) = [u_1(\pi_t(l_t)) - u_0(\pi_t(l_t))] \cdot \pi_t'(l_t).$$

Use the Envelope theorem to compute the derivative of $W_t(p)$

$$\frac{W_t'(p)}{\pi_t'(p)} = \int_t^{t(p)} e^{-r(s-t)} \left(1 - e^{-(r+\lambda)(s-t)}\right) \cdot (A(\pi_s(m_s)) - rV)\, ds$$
$$+ e^{-r(t(p)-t)} \cdot u_1(\pi_{t(p)}(l_{t(p)})) - e^{-(r+\lambda)(t(p)-t)} \cdot u_0(\pi_{t(p)}(l_{t(p)})).$$

This implies that l_t satisfies $W_t'(p) = U_t'(p)$ if

$$\int_t^{t(l_t)} \left(e^{-r(s-t)} - e^{-(r+\lambda)(s-t)}\right) \cdot (A(\pi_s(m_s)) - rV - A(\pi_s(l_t)))\, ds = 0, \tag{B.38}$$

which yields (B.33). The right continuity of the belief process l_t implies that (B.33) must hold in some neighborhood $B_\varepsilon(t_0) \not\subset \mathbb{T}$. The derivative of (B.33) with respect to t is

$$\begin{aligned} 0 = &\left(e^{-r(t(l_t)-t)} - e^{-(r+\lambda)(t(l_t)-t)}\right) \cdot \overbrace{\left(A(\pi_{t(l_t)}(m_{t(l_t)})) - rV - A(\pi_{T(l_t)}(l_t))\right) \cdot t'(l_t)}^{\leq 0} \cdot \dot{l}_t \\ &\underbrace{- \int_t^{t(l_t)} \left(e^{-r(s-t)} - e^{-(r+\lambda)(s-t)}\right) \cdot A'(\pi_s(l_t)) \cdot \pi_s'(l_t)\, ds \cdot \dot{l}_t}_{\geq 0} \\ &+ \int_t^{t(l_t)} \left(-re^{-r(s-t)} + (r+\lambda)e^{-(r+\lambda)(s-t)}\right) \cdot (A(\pi_s(m_s)) - rV - A(\pi_s(l_t)))\, ds. \end{aligned} \tag{B.39}$$

Use the first order condition $t'(l_t)$ and (B.33), to simplify (B.39) to

$$\begin{aligned} 0 = &-\left(e^{-r(t(l_t)-t)} - e^{-(r+\lambda)(t(l_t)-t)}\right) \cdot \partial_2 U(\pi_{T(l_t)}(l_t), \pi_{T(l_t)}(l_t)) \cdot \pi_{T(l_t)}'(l_t) \cdot \dot{l}_t \\ &- \int_t^{t(l_t)} \left(e^{-r(s-t)} - e^{-(r+\lambda)(s-t)}\right) \cdot A'(\pi_s(l_t)) \cdot \pi_s'(l_t)\, ds \cdot \dot{l}_t \\ &+ \int_t^{t(l_t)} \lambda e^{-(r+\lambda)(s-t)} \cdot (A(\pi_s(m_s)) - rV - A(\pi_s(l_t)))\, ds. \end{aligned} \tag{B.40}$$

The terms multiplying $\dot{l}_t$ in (B.40) have the same sign (negative), implying that

$$\begin{aligned} \text{sign}\left(\dot{l}_t\right) &= \text{sign}\left\{\int_t^{t(l_t)} e^{-(r+\lambda)(s-t)} \cdot (A(\pi_s(m_s)) - rV - A(\pi_s(l_t)))\, ds\right\} \\ &\overset{(i)}{=} \text{sign}\left\{\int_t^{t(l_t)} e^{-r(s-t)} \cdot (A(\pi_s(m_s)) - rV - A(\pi_s(l_t)))\, ds\right\}, \end{aligned}$$

where (i) holds due to (B.33). This yields (B.34). □

Lemma B.12 (Single separating period)**.** *Suppose belief process $(l_t)_{t\geq 0}$ is right continuous and satisfies the lowest continuation surplus refinement 4. Then there does not exist a t^* satisfying definition* (B.32).

Proof. Following Lemma B.11, it follows that $l_{t^*} \in (l_{t_Q}, l_{t_S}]$. Moreover, the requirement of the lowest continuation surplus refinement 4 that $l_t \in R(t)$ implies that $l_t \geq l_{t_Q}$ for all $t \in (t^*, t_Q)$ due to Lemma B.10. For type l_{t_Q} to find it incentive compatible to wait until time t_Q to separate, it must be the case that $A(\pi_t(m_t)) - A(\pi_t(l_{t_Q})) - rV \geq 0$ in the vicinity of t_Q. Moreover, weak concavity of $A(\pi_t(p))$ in t and the fact that there are no separations during $[t^*, t_Q]$ by definition of t^* implies that

$$A(\pi_t(m_t)) - A(\pi_t(l_{t_Q})) \geq rV \qquad \textit{for all} \qquad t \in [t^*, t_Q]. \tag{B.41}$$

(i) Suppose $A\left(\pi_{t^*}(m_{t^*})\right) - A\left(\pi_{t^*}(l_{t_Q})\right) = rV$. From weak concavity of $A(\pi_t(p))$ in t, it implies that $A(\pi_t(m_t)) - A(\pi_t(l_{t_Q})) - rV = 0$ for all $t \in [t^*, t_Q]$. This implies that $l_{t^*} = l_{t_Q}$ and, moreover, that this type is indifferent in separating any time between t^* and t_Q, contradicting the definition of t_Q.

(ii) Suppose $A\left(\pi_{t^*}(m_{t^*})\right) - A\left(\pi_{t^*}(l_{t_Q})\right) > rV$. Following Lemma B.11, it must be the case that $l_{t^*} \in (l_{t_Q}, t_S]$. By right continuity, following Lemma B.11 it implies that

$$\int_{t^*}^{t(l_{t^*})} \left(e^{-r(s-t^*)} - e^{-(r+\lambda)(s-t^*)}\right) \cdot \left(A(\pi_s(m_s)) - rV - A(\pi_s(l_{t^*}))\right) ds = 0. \tag{B.42}$$

(a) First, consider the case

$$\begin{aligned} 0 &< \int_{t^*}^{t(l_{t^*})} e^{-r(s-t^*)} \cdot \left(A(\pi_s(m_s)) - rV - A(\pi_s(l_{t'_S}))\right) ds \\ &\overset{(i)}{=} \int_{t^*}^{t(l_{t^*})} e^{-(r+\lambda)(s-t^*)} \cdot \left(A(\pi_s(m_s)) - rV - A(\pi_s(l_{t^*}))\right) ds, \end{aligned}$$

where equality (i) follows from (B.42). This implies that

$$\int_{t^*}^{t(l_{t^*})} \left(l_{t^*} \cdot e^{-r(s-t^*)} + (1 - l_{t^*}) \cdot e^{-(r+\lambda)(s-t^*)}\right) \cdot \left(A(\pi_s(m_s)) - rV - A(\pi_s(l_{t^*}))\right) ds > 0,$$

which implies that type l_{t^*} finds it strictly optimal to wait until time $t(l_{t^*}) \in (t_Q, t_S)$ to separate, rather than separate at time t^*. This is a contradiction.

(b) Second, consider the case

$$\begin{aligned} 0 &> \int_{t^*}^{t(l_{t^*})} e^{-r(s-t)} \cdot \left(A(\pi_s(m_s)) - rV - A(\pi_s(l_{t^*}))\right) ds \\ &\overset{(i)}{=} \int_{t'_S}^{t(l_{t^*})} e^{-(r+\lambda)(s-t)} \cdot \left(A(\pi_s(m_s)) - rV - A(\pi_s(l_{t^*}))\right) ds, \end{aligned}$$

where equality (i) follows from (B.42). Similar to the previous case, type l_{t^*} then finds it strictly optimal to separate at time t^*, rather than wait until $t(l_{t^*}) \in [t_Q, t_S]$, which is a contradiction.

(c) Third, and final, consider the case

$$0 = \int_{t^*}^{t(l_{t^*})} e^{-r(s-t)} \cdot (A(\pi_s(m_s)) - rV - A(\pi_s(l_{t^*}))) \, ds. \tag{B.43}$$

Following (B.34) of Lemma B.11, it implies that $\dot{l}_{t^*} = 0$. This implies that for separation to be optimal for type l_{t^*} at time t^*, it requires that $A(q_{t^*}) - A(\pi_{t^*}(l_{t^*})) < rV$. Rewrite (B.43) as

$$\begin{aligned} 0 = &\overbrace{\int_{t^*}^{t_Q} e^{-r(s-t)} \cdot (A(\pi_s(m_s)) - rV - A(\pi_s(l_{t^*}))) \, ds}^{\overset{(i)}{<}0} \\ &+ \underbrace{\int_{t_Q}^{t(l_{t^*})} e^{-r(s-t)} \cdot (A(\pi_s(m_s)) - rV - A(\pi_s(l_{t^*}))) \, ds}_{\overset{(ii)}{<}0} < 0, \end{aligned} \tag{B.44}$$

Inequality (i) in (B.44) for the first term follows from $A(q_{t^*}) - A(\pi_{t^*}(l_{t^*})) < rV$ and weak concavity of $A(\pi_t(\underline{p}))$ in t. Inequality (ii) in (B.44) for the second term follows from Lemma B.10 and the fact that $l_{t^*} \geq l_t$ for $t < t(l_{t^*})$. Inequality (B.44) is then a contradiction with (B.43), implying that such a case is also inconsistent with equilibrium.

□

Lemma B.13 now concludes the proof of Proposition B.1 by showing that client beliefs track the lowest ex-ante type during the quiet period.

Lemma B.13 (Unique continuous equilibrium)**.** *Suppose belief process $(l_t)_{t\geq 0}$ is continuous and satisfies the lowest continuation surplus refinement 4. Then $t_S = \bar{t}$ and $l_t = \underline{p}$ for $t \leq t_Q$.*

Proof. Suppose $t_S < \bar{t}$. It is necessarily the case that $|S(\bar{t})| > 1$ since, otherwise, waiting between t_S and $\bar{t}$ would not be incentive compatible for type l_{t_S}. Following Lemma B.5 it implies that $l_t = \min S(\bar{t})$ for every $t \in [t_S, \bar{t})$. Consequently, it implies that belief process l_t would have to have a jump at $\bar{t}$ which is a contradiction with continuity. This implies that it must be the case that $t_S = \bar{t}$.

Now, consider the off-equilibrium path beliefs for $t \leq t_Q$. The claim is that if l_t is continuous, then it must be the case that $l_t = \underline{p}$. Consider $t < t_Q$ and suppose that l_t is an interior solution to (B.15), rewritten here as

$$V_t(l_t) = \min_{p\in[\underline{p},\overline{p}]} V_t(p). \tag{B.45}$$

Since $V_t(p) = W_t(p) - U_t(p)$ due to strict convexity of $W_t(p)$ for $p \in [\underline{p}, \overline{p}]$ and the linearity of $U_t(p)$ in p it follows that (B.45) is satisfied if and only if $W_t'(l_t) = U_t'(l_t)$. Following the proof of Lemma B.11, and equation (B.38) specifically, obtain that $W_t'(l_t) = U_t'(l_t)$ if and only if

$$\int_t^{t(l_t)} \left(e^{-r(s-t)} - e^{-(r+\lambda)(s-t)}\right) \cdot \left(A(\pi_s(m_s)) - rV - A(\pi_s(l_t))\right) ds = 0. \tag{B.46}$$

Note, that $l_t \to \underline{p}$ as $t \to t_Q$. This implies that there exists a $t_0 < t_Q$ such that $\dot{l}_{t_0} < 0$ as, otherwise, it would have to be that $l_t \equiv \underline{p}$ for all $t \in [0, t_Q]$. From (B.34), it implies that $\dot{l}_{t_0} < 0$ if and only if

$$\int_{t_0}^{t(l_{t_0})} e^{-r(s-t_0)} \cdot \left(A(\pi_s(m_s)) - rV - A(\pi_s(l_{t_0}))\right) ds < 0. \tag{B.47}$$

Inequality (B.47) combined with (B.46) then implies that

$$\int_{t_0}^{t(l_{t_0})} \left(l_{t_0} \cdot e^{-r(s-t)} + (1 - l_{t_0}) \cdot e^{-(r+\lambda)(s-t)}\right) \cdot \left(A(\pi_s(m_s)) - rV - A(\pi_s(l_{t_0}))\right) ds > 0,$$

implying that type l_{t_0} would find it strictly optimal to separate at time t_0, rather than wait until time $t(l_{t_0})$. Consequently, there cannot be a t_0 such that $l_{t_0} > \underline{p}$ is an interior solution to (B.15) and $\dot{l}_{t_0} < 0$. This implies that $l_t \equiv \underline{p}$ for all $t \in [0, t_Q]$. □

B.4 Perturbation Approach and Limiting Beliefs

Now, we consider a perturbation of the model in which each intermediary-agent pair receive a private relationship-specific disutility shock $\tilde{s}$ distributed according $\tilde{s} \sim \text{Exp}(\Delta)$. We show that as $\Delta \to \infty$ the equilibrium beliefs must converge to the lowest continuation surplus beliefs in definition 4. We assume each intermediary-agent pair receives a shock in period t with intensity $\varepsilon > 0$. Similar to Acemoglu and Pischke (1998), we assume these shocks are discrete – once a shock $\tilde{s}$ arrives at time t, the intermediary and the agent can separate and avoid it, or continue employment, but suffer a joint disutility equal to $-\tilde{s}$.[18] Shocks $\tilde{s}$ imply that there is always a positive probability that an agent will separate from the intermediary, implying that all beliefs can be determined by Bayes rule. Intermediary-agent pairs that have a high continuation value relative to their outside options will be more resilient to relationship-specific shocks and will, thus, be less likely to separate. The exponential distribution of the shocks corresponds to the Logit specification of a Quantal Response Equilibrium introduced in McKelvey and Palfrey (1995) and McKelvey and Palfrey (1998) and can be substantially relaxed.

The magnitude of Δ is inversely related to the magnitude and dispersion of the shock and it converges to 0 as $\Delta \to +\infty$. In what follows, we index the equilibrium parameters and values by the magnitude of the shock Δ. In addition, we refer to $e_t(p; \Delta)$ as the probability of letting go of the agent of ability p at time t following the history of good performance. In addition, we denote by $f_t(p; \Delta)$ the conditional distribution

[18] Due to the wages acting as a transferable utility, it does not matter whether it is the intermediary, the agent, or a combination of the two that receives the disutility shock.

of ex-ante types p that continue to be employed by the intermediary at time t following the path of good performance.

Definition 5 (Perturbed limiting equilibrium). *A limiting equilibrium is a sequence of value functions $V_t(p;\Delta_n)$, probability density processes $f_t(p;\Delta_n)$, and beliefs $l_t(\Delta_n)$ and $m_t(\Delta_n)$ such that*

(i) ***Equilibrium limit:*** *there exists a uniform limit in time t and belief p:*

$$\lim_{n\to\infty}\left\{V_t(p;\Delta_n),\dot{V}_t(p;\Delta_n),f_t(p;\Delta_n),m_t(\Delta_n),l_t(\Delta_n),\gamma_t(\Delta_n)\right\}=\left\{V_t(p),\dot{V}_t(p),f_t(p),m_t,l_t,\gamma_t\right\} \quad \text{(B.48)}$$

(ii) ***Belief consistency:*** *the limiting beliefs $(l_t,m_t,\gamma_t)_{t\geq 0}$, the implied churning strategies $(e_t(p))_{t\geq 0}$, and value functions $(V_t(p))_{t\geq 0}$ for each ex-ante private type p are an equilibrium of a game in which $\varepsilon=0$ and $\Delta=+\infty$.*

(iii) ***Belief regularity:*** *for every $t<\bar{t}$ the support of remaining types $support(p\,|\tau>t,X_t=t)=support(f_t)$ is equal to its derived set, i.e., equal to its limit points.*

Consider the limiting equilibrium $(l_t,m_t,V_t(\cdot),e_t(\cdot))_{t\geq 0}$. For this limiting equilibrium, define by $R(t)$ the set of types that remain with the intermediary with positive probability by time t in the limiting equilibrium:

$$R(t)\overset{def}{=}support(f_t)=cl\,\{p:\quad f_t(p)>0\}\,. \quad \text{(B.49)}$$

Lemma B.14 (Limiting equilibrium beliefs). *The limiting equilibrium belief process satisfies the lowest continuation surplus refinement 4, i.e., $l_t\in R(t)$ and $V_t(l_t)=\min_{p\in R(t)}V_t(p)$.*

Proof. Consider a time $t\notin\mathbb{T}^I$, implying that $V_t(p)>V$ for all $p\in R(t)$. Since $R(t)$ is a closed set, it implies that $\inf_{p\in R(t)}V_t(p)>V$. The value functions $V_t(p;\Delta_n)$ and $V_t(p)$ are weakly convex in p as they are solutions to their respective optimal stopping problems. Point-wise convergence of convex functions on the interval $[0,1]$ is uniform. Consequently, there exists an N sufficiently large such that $\inf_{p\in R(t)}V_t(p;\Delta_n)>V$ for all $n>N$. For $t\notin\mathbb{T}^I$ it follows that Belief $l_t(\Delta_n)$ is given by

$$l_t(\Delta_n)=\frac{\int p\cdot f_t(p;\Delta_n)\cdot e^{-\Delta_n\cdot[V_t(p;\Delta_n)-V]}\,dp}{\int f_t(p;\Delta_n)\cdot e^{-\Delta_n\cdot[V_t(p;\Delta_n)-V]}\,dp}.$$

By definition of $R(t)$ it follows that

$$\lim_{n\to\infty}l_t(\Delta_n)=\frac{\int p\cdot f_t(p;\Delta_n)\cdot e^{-\Delta_n\cdot[V_t(p;\Delta_n)-V]}\,dp}{\int f_t(p;\Delta_n)\cdot e^{-\Delta_n\cdot[V_t(p;\Delta_n)-V]}\,dp}\overset{(i)}{\in}[\min R(t),\,\max R(t)]\,.$$

Consider now the limit $V_t(l_t)$:

$$\begin{aligned}V_t(l_t) &= V\left(\lim_{n\to\infty} l_t(\Delta_n)\right) = \lim_{n\to\infty} V_t(l_t(\Delta_n)) = \lim_{n\to\infty} V_t\left(\frac{\int p\cdot f_t(p;\Delta_n)\cdot e^{-\Delta_n\cdot[V_t(p;\Delta_n)-V]}\,dp}{\int f_t(p;\Delta_n)\cdot e^{-\Delta_n\cdot[V_t(p;\Delta_n)-V]}\,dp}\right)\\ &\overset{(i)}{\leq} \lim_{n\to\infty}\left[\int V_t(x)\cdot\frac{f_t(x;\Delta_n)\cdot e^{-\Delta_n\cdot[V_t(x;\Delta_n)-V]}}{\int f_t(y;\Delta_n)\cdot e^{-\Delta_n\cdot[V_t(y;\Delta_n)-V]}\,dy}\,dx\right]\\ &= \lim_{n\to\infty}\left[\int V_t(x)\cdot\frac{1}{\int\frac{f_t(y;\Delta_n)}{f_t(x;\Delta_n)}\cdot e^{\Delta_n\cdot(V_t(x;\Delta_n)-V_t(y;\Delta_n))}\,dy}\,dx\right]\end{aligned} \tag{B.50}$$

where inequality (i) holds by Jensen's inequality due to the convexity of $V_t(p)$ in p.

Choose a $z_t \in \arg\min_{p\in R(t)} V_t(p)$. There exist $\underline{z}_{\varepsilon,t} \leq z \leq \overline{z}_{\varepsilon,t}$ with at least one of the inequalities being strict such that

(i) $(\underline{z}_{\varepsilon,t}, \overline{z}_{\varepsilon,t}) \in R(t)$, which follows from the regularity assumption that the support of $f_t(p)$ is equal to its derived set;

(ii) $V_t(z;\Delta_n) - V_t(z_t;\Delta_n) \leq \varepsilon$ for all $z \in (\underline{z}_{\varepsilon,t}, \overline{z}_{\varepsilon,t})$ and $n \geq N$, which follows from continuity and convexity of $V_t(p)$ over its domain, and uniform convergence of $V_t(p;\Delta_n)$ in p.

Define

$$Q_\varepsilon(t) \overset{def}{=} \{x \in R(t): \quad V_t(x) > V_t(z_t) + 3\cdot\varepsilon\}. \tag{B.51}$$

By uniform convergence of $V_t(x;\Delta_n)$ to $V_t(x)$ there exists N such that for all $n > N$:

$$V_t(x;\Delta_n) > V_t(z_t;\Delta_n) + 2\cdot\varepsilon \qquad \textit{for all} \quad x \in Q_\varepsilon(t). \tag{B.52}$$

Suppose $x \in Q_\varepsilon(t)$. Then

$$\begin{aligned}&\lim_{n\to\infty}\left[\int\frac{f_t(y;\Delta_n)}{f_t(x;\Delta_n)}\cdot e^{\Delta_n\cdot(V_t(x;\Delta_n)-V_t(y;\Delta_n))}\,dy\right]\\ &\geq \lim_{n\to\infty}\left[\int_{\underline{z}_{\varepsilon,t}}^{\overline{z}_{\varepsilon,t}}\frac{f_t(y;\Delta_n)}{f_t(x;\Delta_n)}\cdot e^{\Delta_n\cdot(V_t(x;\Delta_n)-V_t(y;\Delta_n))}\,dy\right]\\ &= \lim_{n\to\infty}\left[\int_{\underline{z}_{\varepsilon,t}}^{\overline{z}_{\varepsilon,t}}\frac{f_t(y;\Delta_n)}{f_t(x;\Delta_n)}\cdot e^{\Delta_n\cdot(V_t(x;\Delta_n)-V_t(z_t;\Delta_n)+V_t(z_t;\Delta_n)-V_t(y;\Delta_n))}\,dy\right]\\ &\geq \lim_{n\to\infty}\left[\int_{\underline{z}_{\varepsilon,t}}^{\overline{z}_{\varepsilon,t}}\frac{f_t(y;\Delta_n)}{f_t(x;\Delta_n)}\cdot e^{\Delta_n\cdot(|V_t(x;\Delta_n)-V_t(z_t;\Delta_n)|-|V_t(z_t;\Delta_n)-V_t(y;\Delta_n)|)}\,dy\right]\\ &\overset{(i)}{\geq} \lim_{n\to\infty}\left[\int_{\underline{z}_{\varepsilon,t}}^{\overline{z}_{\varepsilon,t}}\frac{f_t(y;\Delta_n)}{f_t(x;\Delta_n)}\cdot e^{\Delta_n\cdot(2\cdot\varepsilon-\varepsilon)}\,dy\right] = \lim_{n\to\infty}\left[e^{\Delta_n\cdot\varepsilon}\cdot\int_{\underline{z}_{\varepsilon,t}}^{\overline{z}_{\varepsilon,t}}\frac{f_t(y;\Delta_n)}{f_t(x;\Delta_n)}\,dy\right] = +\infty,\\ \Rightarrow \quad &\lim_{n\to\infty}\left[\int\frac{f_t(y;\Delta_n)}{f_t(x;\Delta_n)}\cdot e^{\Delta_n\cdot(V_t(x;\Delta_n)-V_t(y;\Delta_n))}\,dy\right] = +\infty \qquad \forall\, x \in Q_\varepsilon(t).\end{aligned} \tag{B.53}$$

where inequality (i) follows from $y \in (\underline{z}_{\varepsilon,t}, \overline{z}_{\varepsilon,t})$ and $x \in Q_\varepsilon(t)$, which implies (B.52). Use (B.53) to simplify (B.50) as

$$
\begin{aligned}
V(l_t) &\leq \lim_{n\to\infty}\left[\int V_t(x)\cdot \frac{1}{\int \frac{f_t(y;\Delta_n)}{f_t(x;\Delta_n)}\cdot e^{\Delta_n\cdot(V_t(x;\Delta_n)-V_t(y;\Delta_n))}\,dy}\,dx\right]\\
&= \lim_{n\to\infty}\left[\int_{R(t)\setminus Q_\varepsilon(t)} V_t(x)\cdot \frac{1}{\int \frac{f_t(y;\Delta_n)}{f_t(x;\Delta_n)}\cdot e^{\Delta_n\cdot(V_t(x;\Delta_n)-V_t(y;\Delta_n))}\,dy}\,dx\right]\\
&+ \lim_{n\to\infty}\left[\int_{Q_\varepsilon(t)} V_t(x)\cdot \frac{1}{\int \frac{f_t(y;\Delta_n)}{f_t(x;\Delta_n)}\cdot e^{\Delta_n\cdot(V_t(x;\Delta_n)-V_t(y;\Delta_n))}\,dy}\,dx\right]\\
&\overset{(i)}{=} \lim_{n\to\infty}\left[\int_{R(t)\setminus Q_\varepsilon(t)} x\cdot \frac{1}{\int \frac{f_t(y;\Delta_n)}{f_t(x;\Delta_n)}\cdot e^{\Delta_n\cdot(V_t(x;\Delta_n)-V_t(y;\Delta_n))}\,dy}\,dx\right]\\
\Rightarrow \qquad V(l_t) &\leq \lim_{n\to\infty}\left[\int_{R(t)\setminus Q_\varepsilon(t)} x\cdot \frac{1}{\int \frac{f_t(y;\Delta_n)}{f_t(x;\Delta_n)}\cdot e^{\Delta_n\cdot(V_t(x;\Delta_n)-V_t(y;\Delta_n))}\,dy}\,dx\right]. \qquad \text{(B.54)}
\end{aligned}
$$

where equality (i) in (B.54) follows from equality (B.53).

Now, consider the denominator in (B.54). It needs to be evaluated only for $x \in R(t)\setminus Q_\varepsilon(t)$. Consider an arbitrary $\hat{\varepsilon} > \varepsilon$. By definition of $Q_\varepsilon(t)$ in (B.51) it follows that

$$V_t(y) > V_t(z_t) + 3\cdot\hat{\varepsilon} \qquad \forall y \in Q_{\hat{\varepsilon}}(t).$$

Then

$$
\begin{aligned}
&\lim_{n\to\infty}\left[\int_{Q_{\hat{\varepsilon}}(t)} \frac{f_t(y;\Delta_n)}{f_t(x;\Delta_n)}\cdot e^{\Delta_n\cdot(V_t(x;\Delta_n)-V_t(y;\Delta_n))}\,dy\right]\\
&= \lim_{n\to\infty}\left[\int_{Q_{\hat{\varepsilon}}(t)} \frac{f_t(y;\Delta_n)}{f_t(x;\Delta_n)}\cdot e^{\Delta_n\cdot(V_t(x;\Delta_n)-V_t(x)+V_t(x)-V_t(y)+V_t(y)-V_t(y;\Delta_n))}\,dy\right]\\
&\leq \lim_{n\to\infty}\left[\int_{Q_{\hat{\varepsilon}}(t)} \frac{f_t(y;\Delta_n)}{f_t(x;\Delta_n)}\cdot e^{\Delta_n\cdot(2\cdot\max_x\{V_t(x;\Delta_n)-V_t(x)\}+V_t(x)-V_t(y))\,dy}\,dy\right] \qquad \text{(B.55)}\\
&\leq \lim_{n\to\infty}\left[\int_{Q_{\hat{\varepsilon}}(t)} \frac{f_t(y;\Delta_n)}{f_t(x;\Delta_n)}\cdot e^{\Delta_n\cdot(2\cdot\max_x\{V_t(x;\Delta_n)-V_t(x)\}+|V_t(x)-V_t(z)|-|V_t(y)-V_t(z)|)\,dy}\,dy\right]\\
&= \lim_{n\to\infty}\left[\int_{Q_{\hat{\varepsilon}}(t)} \frac{f_t(y;\Delta_n)}{f_t(x;\Delta_n)}\cdot e^{\Delta_n\cdot(2\cdot\max_x\{V_t(x;\Delta_n)-V_t(x)\}+\varepsilon-\hat{\varepsilon})\,dy}\,dy\right] \overset{(i)}{=} 0,
\end{aligned}
$$

where equality (i) follows the fact that $\exists\, N_{\hat{\varepsilon}-\varepsilon}$ such that $\forall\, n \geq N_{\hat{\varepsilon}-\varepsilon}$ such that

$$2\cdot\max_x\{V_t(x;\Delta_n)-V_t(x)\}+V_t(x)-V_t(y)) < 2\cdot\max_x\{V_t(x;\Delta_n)-V_t(x)\}+\varepsilon-\hat{\varepsilon} < 0$$

due to uniform convergence of $V_t(x;\Delta_n) \to V_t(x)$ in x. Moreover, $\frac{f_t(y;\Delta_n)}{f_t(x;\Delta_n)}$ is bounded due to the fact that $x \in R(t)$ and, consequently $f_t(x;\Delta_n) > f_t(x)/2 > 0$ for n sufficiently large. Since (B.55) holds for every $\hat{\varepsilon} > \varepsilon$, it implies that

$$\begin{aligned} &\lim_{n\to\infty} \left[\int \frac{f_t(y;\Delta_n)}{f_t(x;\Delta_n)} \cdot e^{\Delta_n \cdot (V_t(x;\Delta_n) - V_t(y;\Delta_n))}\, dy \right] \\ &= \lim_{n\to\infty} \left[\int_{R(t)\setminus Q_\varepsilon(t)} \frac{f_t(y;\Delta_n)}{f_t(x;\Delta_n)} \cdot e^{\Delta_n \cdot (V_t(x;\Delta_n) - V_t(y;\Delta_n))}\, dy \right]. \end{aligned} \tag{B.56}$$

Substitute (B.56) into (B.54) to obtain

$$\begin{aligned} V_t(l_t) &\leq \lim_{n\to\infty} \left[\int_{R(t)\setminus Q_\varepsilon(t)} V_t(x) \cdot \frac{f_t(x;\Delta_n)}{\int f_t(y;\Delta_n) \cdot e^{\Delta_n \cdot (V_t(x;\Delta_n) - V_t(y;\Delta_n))}\, dy}\, dx \right] \\ &\overset{(i)}{\leq} \lim_{n\to\infty} \left[\int_{R(t)\setminus Q_\varepsilon(t)} V_t(x) \cdot \frac{f_t(x;\Delta_n)}{\int_{R(t)\setminus Q_{\hat{\varepsilon}}(t)} f_t(y;\Delta_n) \cdot e^{\Delta_n \cdot (V_t(x;\Delta_n) - V_t(y;\Delta_n))}\, dy}\, dx \right]. \end{aligned} \tag{B.57}$$

Inequality (B.57) holds for any $\hat{\varepsilon}$, implying that

$$V_t(l_t) \leq \lim_{n\to\infty} \left[\int_{R(t)\setminus Q_\varepsilon(t)} V_t(x) \cdot \frac{f_t(x;\Delta_n)}{\int_{R(t)\setminus Q_\varepsilon(t)} f_t(y;\Delta_n) \cdot e^{\Delta_n \cdot (V_t(x;\Delta_n) - V_t(y;\Delta_n))}\, dy}\, dx \right]. \tag{B.58}$$

Inequality (B.58) holds for any $\varepsilon > 0$, implying that

$$\begin{aligned} V_t(l_t) &\leq \lim_{\varepsilon\to 0} \lim_{n\to\infty} \left[\int_{R(t)\setminus Q_\varepsilon(t)} \frac{V_t(x) \cdot f_t(x;\Delta_n)}{\int_{R(t)\setminus Q_\varepsilon(t)} f_t(y;\Delta_n) \cdot e^{\Delta_n \cdot (V_t(x;\Delta_n) - V_t(y;\Delta_n))}\, dy}\, dx \right] \\ &\overset{(i)}{=} \lim_{n\to\infty} \left[\int_{R(t)\cap \arg\min V_t(x)} \frac{V_t(x) \cdot f_t(x;\Delta_n)}{\int_{R(t)\cap \arg\min V_t(x)} f_t(y;\Delta_n) \cdot e^{\Delta_n \cdot (V_t(x;\Delta_n) - V_t(y;\Delta_n))}\, dy}\, dx \right] \\ &\overset{(ii)}{=} \min V_t(x). \end{aligned} \tag{B.59}$$

where change of limits in (i) is possible since $Q_\varepsilon(t)$ does not depend on n and equality (ii) holds because $V_t(x)$ is convex, and so is the set $\arg\min V_t(x)$. Inequality (B.59) proves that $V_t(l_t) \leq \min_{l_t \in R(t)} V_t(l_t)$, which, combined with the requirement that $l_t \in R(t)$ implies that $V_t(l_t) = \min_{l_t \in R(t)} V_t(l_t)$. □

B.5 Lowest Continuation Surplus Equilibrium with Binary Types

In this section we consider a simplified version of the model in which there are only two types, $\tilde{p}_0 \in \{\underline{p}, \overline{p}\} = \{p^L, p^H\}$ and show that there is a unique equilibrium satisfying the lowest continuation surplus refinement 4. As there are only two types, denote by $p_t^i = \pi_t(p^i)$ the posterior belief and $V_t^i = V_t(p^i) - V$ the intermediary's continuation value net of her opportunity cost V for an agent of ex-ante type $i \in \{L, H\}$. As in Section B.3, define t_S to be the last moment before time $\bar{t} = \sup\{\mathit{support}(\tau)\}$ when a type separates:

$t_S \stackrel{def}{=} \sup\{t < \bar{t} : \quad t \in \mathbb{T}\}$. The following Lemma is the analogue of Lemma B.5 for a binary ex-ante type distribution.

Lemma B.15 (Final period separation characterization)**.** *Suppose belief process $l = (l_t)_{t\geq 0}$ is right-continuous and satisfies the lowest continuation surplus refinement 4. Then it must be the case that*

(i) both types find it weakly optimal to separate at $\bar{t}$, i.e., $\bar{t}(p^L) = \bar{t}(p^H) = \bar{t}$;

(ii) there exists a quiet period prior to $\bar{t}$, i.e., $t_S < \bar{t}$;

(iii) clients attribute separations during $[t_S, \bar{t}]$ to low skilled agents, i.e., $l_t = p^L$ for every $t \in (t_S, \bar{t})$.

Proof. Suppose $\bar{t}(p^L) < \bar{t}(p^H)$. This implies that $R(t) = \{p^H\}$ for $t \in (\bar{t}(p^L), \bar{t}(p^H)]$. Following the lowest continuation surplus refinement 4, it follows that $l_t = p^H$ for $t \in (\bar{t}(p^L), \bar{t}(p^H)]$, implying a contradiction with the optimality of separating at time $\bar{t}(p^L)$. A symmetric argument is in effect if $\bar{t}(p^L) > \bar{t}(p^H)$ implying that it has to be the case that $\bar{t}(p^L) = \bar{t}(p^H)$.

Results (ii) and (iii) follow from the already proven Lemma B.5 of Section B.3. □

We now show that there exists a unique equilibrium satisfying the lowest continuation surplus refinement 4. As the model features only two types, we weaken the clients' belief continuity assumption with the right continuity.

Lemma B.16 (Unique equilibrium)**.** *Suppose belief process $l = (l_t)_{t\geq 0}$ is right continuous and satisfies the lowest continuation surplus refinement 4. Then $l_t = p^L$ for every $t < \bar{t}$.*

Proof. Following Lemma B.15 it follows that $l_t = p^L$ for $t \in (t_S, \bar{t})$. Right continuity of beliefs then implies that $l_{t_S} = p^L$, meaning that it is the L type separating at time t_S. There exists an $\varepsilon > 0$ such that

$$A(\pi_t(m_{t_S})) - rV - A(p_t^L) < 0 \qquad \forall\, t \in (t_S, t_S + \varepsilon)$$

as, otherwise, type p^L would find it strictly preferable to wait past t_S to separate, which would contradict the definition of t_S. Define

$$t_Q \stackrel{def}{=} \sup\{t < t_S : \qquad \exists \varepsilon > 0 \quad s.t. \quad (t - \varepsilon, t) \notin \mathbb{T}^I\}.$$

Following the same argument and derivations of Lemma B.8 for the continuum of types, it follows that there is no pooling during $[t_Q, t_S]$, implying that $l_t = p^L$ for all $t \in [t_Q, t_S]$. Since it's the same type p^L that is separating during the period $[t_Q, t_S]$, it implies that

$$A(\pi_t(m_t)) - rV - A(p_t^L) = 0 \qquad \forall\, t \in [t_Q, t_S].$$

and that there exists an $\varepsilon > 0$ such that

$$A(\pi_t(m_t)) - rV - A(p_t^L) > 0 \qquad \forall\, t \in (t_Q - \varepsilon, t_Q). \tag{B.60}$$

The next step is to show that type p^H finds it strictly preferable to wait until time $\bar{t}$ to separate than to separate at time t_Q and belief p^L. Denote by $L_t(p)$ the belief at which type p is willing to separate at time t. It satisfies (B.13) which we rewrite here as

$$V + U_t(L_t(p)) = W_t(p).$$

for $p \in \{p^L, p^H\}$. At $t = \bar{t}$ both types are willing to separate at the pooling belief, implying that $L_{\bar{t}}(p^L) = L_{\bar{t}}(p^H) = l_{\bar{t}}$. It follows that

$$\dot{L}_t(p) = \frac{A(\pi_t(L_t(p))) + rV - A(\pi_t(m_t))}{\partial_1 \pi_t(L_t(p)) \cdot \partial_2 U(\pi_t(p), \pi_t(L_t(p)))}.$$

Types p^L and p^H are willing to pool only if there indifference beliefs are equal, i.e., types p^L and p^H find pooling at time $\hat{t}$ incentive compatible if and only if $L_{\hat{t}}(p^L) = L_{\hat{t}}(p^H) = l_{\hat{t}}$. For $t \in [t_S, \bar{t}]$ we have $\dot{L}_t(p^H) < \dot{L}_t(p^L)$, while for $t \in [t_Q, t_S]$ we have $\dot{L}_t(p^H) < \dot{L}_t(p^L) = 0$. Consequently, $L_{t_Q}(p^H) > L_{t_Q}(p^L)$, implying that $V_{t_Q}(p^H) > V_{t_Q}(p^L)$.

Suppose there exists a time $t_* \in [0, t^*)$ which corresponds to the last time when any type is willing to separate preceding time t_Q, i.e.,

$$t_* = \sup\{t < t_Q : \quad t \in \mathbb{T}^I\}.$$

By definition of t_Q, it must be the case that $t_* < t_Q$. Due to the linearity of the value function $V_t(p)$ in p, the local arguments in Lemma B.15, inherited from Lemma B.5, extend to the quiet period (t^*, t_Q), implying that $l_t = p^L$ for $t \in (t^*, t_Q)$. Right continuity of beliefs then requires that $l_{t^*} = p^L$. Since period (t^*, t_Q) is a quiet period, it follows from (B.60) that

$$A(\pi_{t^*}(m_{t^*})) - rV - A(\pi_{t^*}(p^L)) > 0.$$

This contradicts the existence of t^* as the L type manager would strictly benefit from waiting past t^* to separate. Consequently, there are no separations in equilibrium during $[0, t_Q)$.

Finally, we show that that $l_t = p^L$ for all $t < \bar{t}$. The comparison $V_t^H > V_t^L$ follows from

$$V_t^H = \int_t^{t_Q} e^{-r(s-t)} \left(p_t^H + (1 - p_t^H) e^{-\lambda(s-t)}\right) \underbrace{\left(A(\pi_s(m_s)) - rV - A(p_s^L)\right)}_{>0} ds$$

$$
\begin{aligned}
&+\int_{t_Q}^{t_S} e^{-r(s-t)}\left(p_t^H+(1-p_t^H)e^{-\lambda(s-t)}\right)\underbrace{\left(A(\pi_s(m_s))-rV-A(p_s^L)\right)}_{=0}ds\\
&+\int_{t_Q}^{\infty} e^{-r(s-t)}\left(p_t^H+(1-p_t^H)e^{-\lambda(s-t)}\right)\left(A(\pi_s(m_{\bar{t}}))-A(p_s^L)\right)ds\\
&\overset{(i)}{>}\int_{t}^{t_Q} e^{-r(s-t)}\left(p_t^H+(1-p_t^H)e^{-\lambda(s-t)}\right)\underbrace{\left(A(\pi_s(m_s))-rV-A(p_s^L)\right)}_{\geq 0}ds\\
&>\int_{t}^{t_Q} e^{-r(s-t)}\left(p_t^L+(1-p_t^L)e^{-\lambda(s-t)}\right)\left(A(\pi_s(m_s))-rV-A(p_s^L)\right)ds=V_t^L,
\end{aligned}
$$

where the strict inequality (i) follows from $V_{t_Q}(p^H)>V_{t_Q}(p^L)$. □

B.6 Perturbed Model Limiting Equilibrium with a Binary Type

We now consider the perturbation of the model, as described in Section B.4 of this Online Appendix B, in which the intermediary-agent pair receives a relationship-specific disutility shock distributed exponentially with parameter Δ. We show in Section B.1.1 that a limiting equilibrium must satisfy the lowest continuation surplus refinement 4 and, consequently, converge to the equilibrium characterized in Section B.5. As the shocks follow an exponential distribution, the probability that a shock exceeds a value x equals $G(x)=e^{-\Delta\cdot x}$.

B.7 Basic Properties of the Model with Shocks and Binary Types

Denote by e_t^L and e_t^H to be the rate at which the intermediary lets go of the agent of ex-ante ability p^L and p^H respectively at time t. If $V_t^i>0$, then the intermediary prefers to retain the agent absent any shocks, consequently $e_t^i=\varepsilon\cdot G(V_t^i)$. If, however, $V_t^i=0$, then the intermediary lets the agent go with intensity ε if any shock arrives, but then may further churn the agent for strategic considerations. Consequently, $e_t^i\geq\varepsilon$ if $V_t^i=0$.

Lemma B.17. *The probability* $\alpha_t=\mathrm{P}_t\big(p=\tilde{p}^H\,|\,X_t=t,\tau>t\big)$ *that the remaining type is a high type conditional on good performance and continued employment by the intermediary satisfies*

$$
\dot{\alpha}_t=\alpha_t(1-\alpha_t)\cdot\left(e_t^L-e_t^H+\lambda\cdot\left(p_t^H-p_t^L\right)\right). \tag{B.61}
$$

Proof. Denote by α_t the fraction of low types that survive. Then

$$
\begin{aligned}
\alpha_{t+dt}=\mathrm{P}\big(\tilde{p}_t=p_t^H\,|\,remain\big)&=\frac{\mathrm{P}\big(\tilde{p}_t=p_t^H,\,remain\big)}{\mathrm{P}\big(\tilde{p}_t=p_t^H,\,remain\big)+\mathrm{P}\big(\tilde{p}_t=p_t^L,\,remain\big)}\\
&=\frac{\alpha_t\cdot(1-e_t^H\,dt)\cdot(1-(1-p_t^H)\lambda\,dt)}{\alpha_t\cdot(1-e_t^H\,dt)\cdot(1-(1-p_t^H)\lambda\,dt)+(1-\alpha_t)\cdot(1-e_t^L\,dt)\cdot(1-(1-p_t^L)\lambda\,dt)}.
\end{aligned}
$$

Then

$$\alpha_{t+dt} - \alpha_t = \frac{\alpha_t \cdot (1 - e_t^H\, dt) \cdot (1 - (1 - p_t^H)\lambda\, dt)}{\alpha_t \cdot (1 - e_t^H\, dt) \cdot (1 - (1 - p_t^H)\lambda\, dt) + (1 - \alpha_t) \cdot (1 - e_t^L\, dt) \cdot (1 - (1 - p_t^L)\lambda\, dt)} - \alpha_t$$
$$= \frac{\alpha_t(1 - \alpha_t) \cdot \left(e_t^L - \lambda \cdot p_t^L - e_t^H + \lambda \cdot p_t^H\right) \cdot dt}{\left(\alpha_t \cdot (1 - e_t^H\, dt - (1 - p_t^H)\lambda\, dt) + (1 - \alpha_t) \cdot (1 - e_t^L\, dt - (1 - p_t^L)\lambda\, dt)\right)^2}.$$

Dividing both sides by dt and taking $dt = 0$ obtain (B.61). □

In what follows, we construct an equilibrium in which prior to the final date $\bar{t}$ both types separate continuously for $t < \bar{t}$ and may feature an atom of separations in the last period $t = \bar{t}$. In other words, we construct an equilibrium in which separation rates e_t^i are finite for all $t < \bar{t}$. In this case, the average separating type k_t along the path of good performance is given by

$$k_t = \frac{p_t^L \cdot (1 - \alpha_t) \cdot e_t^L + p_t^H \cdot \alpha_t \cdot e_t^H}{(1 - \alpha_t) \cdot e_t^L + \alpha_t \cdot e_t^H}. \tag{B.62}$$

Lemma B.18 (Monotone equilibrium)**.** *Suppose $\varepsilon < \bar{\varepsilon}$. Then, in equilibrium, it must be the case that $k_t \leq q_t$ with the inequality being strict for any $t < \bar{t}$. Consequently, $V_t^H \geq V_t^L$ and $\mathbb{T}^H \subseteq \mathbb{T}(p^L)$.*

Proof. Without loss of generality, suppose $\bar{t} > 0$ and $k_0 \geq q_0$. Define $\mathbb{T}\left(()\, p\right)$ as the set of times when the intermediary chooses to let go of the agent in the absence of a shock. Define

$$\hat{t} \stackrel{def}{=} \inf\left\{t > 0 : \ k_t \in \mathbb{T}\left(p^L\right) \cup \mathbb{T}\left(p^H\right) \text{ and } k_t \leq q_t\right\}.$$

The time $\hat{t}$ is well defined as $k_{\bar{t}} = q_{\bar{t}}$. There exists an ε sufficiently low so that $q_t - \frac{rV}{2} < \pi_t(k_0)$ for every $t \in (0, \bar{t})$ since the types that are departing voluntarily prior to $\hat{t}$ are better than q_t, thus lowering the average, while the rate of exogenous departures ε can be made sufficiently small relative to $r \cdot V$. Define $\hat{\alpha}_t$ as the fraction of high types if (i) there are no departures in the absence of a shock, (ii) all low types leave following arrival of a shock, and (iii) no high type leaves following a shock. Then, $\hat{\alpha}_t \geq \alpha_t$ and from Lemma B.17 it follows that

$$\dot{\hat{\alpha}}_t = \hat{\alpha}_t\,(1 - \hat{\alpha}_t) \cdot \left[\varepsilon + \lambda \cdot (p_t^H - p_t^L)\right]$$
$$\frac{d}{dt} \ln\left(\frac{\hat{\alpha}_t}{1 - \hat{\alpha}_t}\right) = \varepsilon + \lambda \cdot (p_t^H - p_t^L)$$
$$\Rightarrow \qquad \hat{\alpha}_t = \frac{\alpha_0 \cdot e^{\varepsilon \cdot t + \lambda \int_0^t (p_s^H - p_s^L)\, ds}}{1 - \alpha_0 + \alpha_0 \cdot e^{\varepsilon \cdot t + \lambda \int_0^t (p_s^H - p_s^L)\, ds}}.$$

This, then implies that

$$q_t = (1 - \alpha_t) \cdot p_t^L + \alpha_t \cdot p_t^H \leq (1 - \hat{\alpha}_t) \cdot p_t^L + \hat{\alpha}_t \cdot p_t^H. \tag{B.63}$$

The difference between q_t and $\pi_t(q_0)$ is given by

$$
\begin{aligned}
q_t - \pi_t(q_0)) &\leq (1-\hat{\alpha}_t)\cdot p_t^L + \hat{\alpha}_t \cdot p_t^H - \pi_t(q_0) \\
&\leq \left(\frac{\alpha_0 \cdot e^{\varepsilon\cdot t + \lambda \int_0^t (p_s^H - p_s^L)\, ds}}{1-\alpha_0+\alpha_0\cdot e^{\varepsilon\cdot t+\lambda\int_0^t (p_s^H-p_s^L)\, ds}} - \frac{\alpha_0 \cdot e^{\lambda \int_0^t (p_s^H - p_s^L)\, ds}}{1-\alpha_0+\alpha_0\cdot e^{\lambda\int_0^t (p_s^H-p_s^L)\, ds}} \right) \cdot (p_t^H - p_t^L) \\
&= \frac{\alpha_0(1-\alpha_0)\cdot e^{\lambda\int_0^t (p_s^H-p_s^L)\, ds}\cdot \left(e^{\varepsilon t}-1\right)}{\left(1-\alpha_0+\alpha_0\cdot e^{\varepsilon\cdot t+\lambda\int_0^t (p_s^H-p_s^L)\, ds}\right)\left(1-\alpha_0+\alpha_0\cdot e^{\lambda\int_0^t (p_s^H-p_s^L)\, ds}\right)} \cdot (p_t^H - p_t^L) \\
&< (e^{\varepsilon t}-1)\cdot (p_t^H - p_t^L).
\end{aligned} \tag{B.64}
$$

For separation at date $\hat{t}$ to be incentive compatible it must be the case that

$$
\begin{aligned}
&\int_0^{\hat{t}} e^{-rt}\cdot \left(p+(1-p)\cdot e^{-\lambda t}\right)\cdot (A(q_t)-rV)\, dt \\
&+ \int_{\hat{t}}^{\infty} e^{-rt}\cdot \left(p+(1-p)\cdot e^{-\lambda t}\right)\cdot A\left(\pi_{t-\hat{t}}(k_{\hat{t}})\right)\, dt \\
&\geq \int_0^{\infty} e^{-rt}\cdot \left(p+(1-p)\cdot e^{-\lambda t}\right)\cdot A(\pi_t(k_0))\, dt.
\end{aligned} \tag{B.65}
$$

A necessary condition for (B.65) to hold is if it holds at $k_{\hat{t}} = q_{\hat{t}}$ and $k_0 = q_0$. Substituting these values into (B.65) obtain

$$
\begin{aligned}
&\int_{\hat{t}}^{\infty} e^{-rt}\cdot \left(p+(1-p)\cdot e^{-\lambda t}\right)\cdot \left(A\left(\pi_{t-\hat{t}}(q_{\hat{t}})\right) - A(\pi_t(q_0))\right)\, dt \\
&\geq \int_0^{\hat{t}} e^{-rt}\cdot \left(p+(1-p)\cdot e^{-\lambda t}\right)\cdot \underbrace{\left(rV + A(\pi_t(q_0)) - A(q_t)\right)}_{\geq r\cdot V/2}\, dt \\
&\geq \int_0^{\hat{t}} e^{-rt}\cdot \left(p+(1-p)\cdot e^{-\lambda t}\right)\cdot \frac{rV}{2}\, dt.
\end{aligned} \tag{B.66}
$$

The upper bound obtained in (B.64) ensures that inequality (B.66) cannot be satisfied for $\varepsilon \leq \bar{\varepsilon}$ uniformly across Δ. □

Lemma B.19. *Suppose separation rates $(e_t^L, e_t^H)_{t\geq 0}$ are differentiable almost everywhere. Then the churning rate γ_t, as defined in (12) is given by*

$$
\gamma_t \overset{def}{=} \dot{k}_t - \lambda k_t(1-k_t) = \frac{\alpha_t(1-\alpha_t)\cdot \left(p_t^H - p_t^L\right)}{\left((1-\alpha_t)\cdot e_t^L + \alpha_t\cdot e_t^H\right)^2}\cdot \left(\dot{e}_t^H\cdot e_t^L - \dot{e}_t^L\cdot e_t^H\right). \tag{B.67}
$$

Proof. If e_t^L and e_t^H are kept constant, i.e., the fractions of types separating are constant, then it follows that $\dot{k}_t = \lambda k_t(1-k_t)$ and $\gamma_t = 0$. Consequently,

$$\begin{aligned}
\gamma_t =& \frac{p_t^L \cdot (1-\alpha_t) \cdot \dot{e}_t^L + p_t^H \cdot \alpha_t \cdot \dot{e}_t^H}{(1-\alpha_t) \cdot e_t^L + \alpha_t \cdot e_t^H} - \frac{p_t^L \cdot (1-\alpha_t) \cdot e_t^L + p_t^H \cdot \alpha_t \cdot e_t^H}{\left((1-\alpha_t) \cdot e_t^L + \alpha_t \cdot e_t^H\right)^2} \cdot \left((1-\alpha_t) \cdot \dot{e}_t^L + \alpha_t \cdot \dot{e}_t^H\right) \\
=& \frac{\dot{e}_t^L \cdot (1-\alpha_t) \cdot \left[p_t^L \cdot (1-\alpha_t) \cdot e_t^L + \alpha_t \cdot e_t^H - p_t^L \cdot (1-\alpha_t) \cdot e_t^L - p_t^H \cdot \alpha_t \cdot e_t^H\right]}{\left((1-\alpha_t) \cdot e_t^L + \alpha_t \cdot e_t^H\right)^2} \\
+& \frac{\dot{e}_t^H \cdot \alpha_t \cdot \left[p_t^H \cdot \left((1-\alpha_t) \cdot e_t^L + \alpha_t \cdot e_t^H\right) - p_t^L \cdot (1-\alpha_t) \cdot e_t^L - p_t^H \cdot \alpha_t \cdot e_t^H\right]}{\left((1-\alpha_t) \cdot e_t^L + \alpha_t \cdot e_t^H\right)^2} \\
=& \frac{\dot{e}_t^L \cdot (1-\alpha_t) \cdot \left[p_t^L \cdot \alpha_t \cdot e_t^H - p_t^H \cdot \alpha_t \cdot e_t^H\right]}{\left((1-\alpha_t) \cdot e_t^L + \alpha_t \cdot e_t^H\right)^2} + \frac{\dot{e}_t^H \cdot \alpha_t \cdot \left[p_t^H \cdot (1-\alpha_t) \cdot e_t^L - p_t^L \cdot (1-\alpha_t) \cdot e_t^L\right]}{\left((1-\alpha_t) \cdot e_t^L + \alpha_t \cdot e_t^H\right)^2} \\
=& \frac{\dot{e}_t^H \cdot \alpha_t(1-\alpha_t) \cdot (p_t^H - p_t^L) \cdot e_t^L - \dot{e}_t^L \cdot \alpha_t(1-\alpha_t) \cdot (p_t^H - p_t^L) \cdot e_t^H}{\left((1-\alpha_t) \cdot e_t^L + \alpha_t \cdot e_t^H\right)^2} \\
=& \frac{\alpha_t(1-\alpha_t) \cdot (p_t^H - p_t^L) \cdot \left[\dot{e}_t^H \cdot e_t^L - \dot{e}_t^L \cdot e_t^H\right]}{\left((1-\alpha_t) \cdot e_t^L + \alpha_t \cdot e_t^H\right)^2}.
\end{aligned}$$

□

If both $V_t^L > 0$ and $V_t^H > 0$ then (B.67) becomes

$$\begin{aligned}
\gamma_t &= \frac{\alpha_t(1-\alpha_t) \cdot (p_t^H - p_t^L) \cdot \left[-\varepsilon^2 \cdot g(V_t^H) \cdot G(V_t^L) \cdot \dot{V}_t^H + \varepsilon^2 \cdot g(V_t^L) \cdot G(V_t^H) \cdot \dot{V}_t^L\right]}{\left((1-\alpha_t) \cdot \varepsilon \cdot G(V_t^L) + \alpha_t \cdot \varepsilon \cdot G(V_t^H)\right)^2} \\
&= \frac{\alpha_t(1-\alpha_t) \cdot (p_t^H - p_t^L)}{\left((1-\alpha_t) \cdot G(V_t^L) + \alpha_t \cdot G(V_t^H)\right)^2} \cdot \left[g(V_t^L) \cdot G(V_t^H) \cdot \dot{V}_t^L - g(V_t^H) \cdot G(V_t^L) \cdot \dot{V}_t^H\right],
\end{aligned}$$

where $g(\cdot) = \Delta \cdot e^{-\Delta x}$ is the pdf of the shock distribution. If $V_t^H > 0$ while $V_t^L = 0$ then (B.67) becomes

$$\gamma_t = \frac{\alpha_t(1-\alpha_t) \cdot (p_t^H - p_t^L)}{\left((1-\alpha_t) \cdot (1+e_t^L) + \alpha_t \cdot G(V_t^H)\right)^2} \cdot \left[G(V_t^H) \cdot (g(0) \cdot \dot{V}_t^L - \dot{e}_t^L) - g(V_t^H) \cdot (1+e_t^L) \cdot \dot{V}_t^H\right].$$

Under continuous separations, the value functions of type i satisfies

$$\begin{aligned}
rV_t^i &= A(q_t) - A(k_t) - rV + \gamma_t \cdot \partial_2 U(p_t^i, k_t) - \lambda \cdot (1-p_t^i) \cdot V_t^i + \dot{V}_t^i \\
&\quad - \varepsilon \cdot G(-V_t^i) \cdot V_t^i + \varepsilon \cdot \mathrm{E}\left[\tilde{s} \cdot \mathbb{1}\{\tilde{s} \geq -V_t^i\}\right] + \max_{e_t^i \geq 0}\{-e_t^i \cdot V_t^i\},
\end{aligned} \tag{B.68}$$

B.1.1 Limiting Equilibrium Beliefs with Stochastic Preference Shocks

In this section, we show that any well-behaved equilibrium limit as $\Delta \to \infty$ of the perturbed model converges to an equilibrium that satisfies the lowest continuation surplus refinement 4.

Definition 6. *A limiting equilibrium is a sequence of value functions $V_t^i(\Delta_n)$ and beliefs $k_t(\Delta_n)$ and $q_t(\Delta_n)$*

such that

(i) *Limit: there exists a uniform limit in t of*[19]

$$\begin{cases} \lim_{n\to\infty} \left\{ V^L(\Delta_n), \dot{V}^L(\Delta_n), V^H(\Delta_n), \dot{V}^H(\Delta_n) \right\} = \left\{ V^L, \dot{V}^L, V^H, \dot{V}^H \right\}, \\ \lim_{n\to\infty} \left\{ q(\Delta_n), k(\Delta_n), \gamma(\Delta_n) \right\} = \left\{ q, k, \gamma \right\}. \end{cases} \tag{B.69}$$

(ii) *The limiting beliefs (q_t, k_t, γ_t), value functions $(V_t^L, V_t^H)_{t\geq 0}$, and churning strategies (e_t^L, e_t^H) are an equilibrium of a game in which $\varepsilon = 0$ and $\Delta = +\infty$.*

As before, define $\mathbb{T} \overset{def}{=} \{t: \ V_t^L = 0\} \cup \{t: \ V_t^H = 0\}$.

Lemma B.20. *Suppose the belief process $(q_t, k_t)_{t\in[0,\bar{t}]}$ is right continuous. For any limiting equilibrium in which $\alpha_{\bar{t}} < 1$, i.e., such that $q_{\bar{t}} < p_{\bar{t}}^H$, it must be the case that $k_t = p_t^L$ for every $t \in \mathbb{T}$.*

Proof. Consider date t such that $V_t^H > 0$ and $V_t^L > 0$. This implies that $V_t^H(\Delta_n) > 0$ and $V_t^L(\Delta_n) > 0$ for a sufficiently high n. This implies that

$$\begin{aligned}
\lim_{n\to\infty} k_t(\Delta_n) &= \lim_{n\to\infty} \left[\frac{p_t^L \cdot (1-\alpha_t(\Delta_n)) \cdot G(V_t^L(\Delta_n)) + p_t^H \cdot \alpha_t(\Delta_n) \cdot G(V_t^H(\Delta_n))}{(1-\alpha_t(\Delta_n)) \cdot G(V_t^L(\Delta_n)) + \alpha_t(\Delta_n) \cdot G(V_t^H(\Delta_n))} \right] \\
&= \lim_{n\to\infty} \left[\frac{p_t^L \cdot (1-\alpha_t(\Delta_n)) \cdot e^{\Delta_n \cdot \left(V_t^H(\Delta_n) - V_t^L(\Delta_n)\right)} + p_t^H \cdot \alpha_t(\Delta_n)}{(1-\alpha_t(\Delta_n)) \cdot e^{\Delta_n \cdot \left(V_t^H(\Delta_n) - V_t^L(\Delta_n)\right)} + \alpha_t(\Delta_n)} \right]
\end{aligned}$$

This limit is equal to p_t^L if $V_t^L < V_t^H$ and p_t^H if $V_t^H > V_t^L$, with the latter leading to a contradiction with the equilibrium for Δ_n sufficiently high.

Suppose at time t we have $V_t^H = V_t^L > 0$. Define times t_1 and t_2 as

$$t_1 \overset{def}{=} \sup\{s \leq t: \quad V_s^L < V_s^H\}, \qquad t_2 \overset{def}{=} \inf\{s \geq t: \quad V_s^L < V_s^H\}.$$

This implies that $V_t^L = V_t^H$ for $t \in [t_1, t_2]$ and, in addition, requires that $\gamma_t = 0$ for $t \in [t_1, t_2]$.

(i) Case $t_1 > 0$. In this case there exists an $\epsilon > 0$ such that $V_s^H > V_s^L$ for every $s \in (t_1 - \epsilon, t_1)$. By the above argument, it implies that $k_s = p_s^L$ and, since $\gamma_s = 0$ for $s \in [t_1, t_2]$, it implies that $k_t = p_t^L$.

(ii) Case $t_2 < \bar{t}$. In this case there exists an $\epsilon > 0$ such that $V_s^H > V_s^L$ for every $s \in (t_2, t_2 + \epsilon)$. By the above argument, it implies that $k_s = p_s^L$ and, since $\gamma_s = 0$ for $s \in [t_1, t_2]$, it implies that $k_t = p_t^L$.

(iii) Case $t_1 = 0$ and $t_2 = \bar{t}$. This implies that $\pi_{-t}(k_t)$ is constant for all $t \in [0, \bar{t}]$. We know that by optimality of optimal stopping, it requires that $k_t < q_t$. However, it implies a jump in beliefs at $\bar{t}$,

[19] The convergence of $\gamma_t(\Delta_n) \to \gamma_t$ is equivalent to $k_t(\Delta_n)$ converging in C^1 to k_t.

which would contradict $V^H_{\bar{t}-} = V^L_{\bar{t}-}$ as the types leaving in the final instance will be better. This leads to a contradiction.

□

B.8 Candidate D1 Definition, Construction, and Possibility for Equilibrium Nonexistence

So far, we have considered the lowest continuation surplus refinement 4 that we are micro-found in Sections B.4 and B.6 with the perturbation approach to the model. A natural question, which we discussed already in Section B.1.2, is why we do not pursue other refinements, such as Divinity. In this section, we offer an extension of the definition of divinity to our dynamic two-player informed game by considering the joint expected payoff of the intermediary-agent pair. This is the focus of Section B.1.1. We formally prove in Section B.1.2 that the equilibrium we construct in the main text, as well as in Section B.5 of this Online Appendix is the unique equilibrium that may survive the divinity refinement 7.– we do so in the context of a binary model for tractability, but the logic also extends to a continuum of types. These results show that if a divine equilibrium does exist, then it coincides with the one we construct in the main text of the paper and this Online Appendix B. We show in Lemma B.26 of Section B.1.2 that such equilibrium survives the divinity refinement if the quiet period is not too long, as proxied by a sufficiently large intermediary outside option V, while there may be no divine equilibrium if the outside option V is very low and the resulting quiet period is too long. We further illustrate this logic in Section B.1.3 graphically in a we provide a semi-analytic example. The reason for such non-existence is that the initial quiet period offers substantial rents to lower-skilled intermediary-agent pairs, which leads higher types to be more willing to deviate at the start of the game. Given the uniqueness of the candidate equilibrium, this leads to the potential nonexistence of an equilibrium satisfying the divinity definition 7 for some parameter values.

B.1.1 Definition of Divinity in our Setting

Signaling games have a special feature that the agent only acts once. Dynamic signaling is not that the agent cannot act sooner, but it is that the buyers can make interim offers. As long as clients are dispersed, an individual client cannot make such a deviation.

Lemma B.21. *The set of Perfect Bayesian Equilibria of the dynamic game coincides with that of a static game in which the intermediary and the agent jointly commit at $t = 0$ to contingent separation date $\tau(p)$, but this time is not observed by clients until it occurs.*[20]

The only distinction between a standard signaling model and this signaling model is the belief process Q.

[20] Every equilibrium of the dynamic game is an equilibrium of the strategic-form game.

This is not present in the classic models in which the flow value prior to signaling is independent of beliefs. However, the claim is that if it were present, then the result would be the same.

Remark 1. *If a Perfect Bayesian Equilibrium can be supported in the initial dynamic game, then it can be supported in a strategic form game. Divinity is not defined for dynamic games, but it is well defined for the static game in which at $t = 0$ the intermediary and the agent commit to a separation time.*

The ex-ante set of types is $[\underline{p}, \overline{p}]$. Define by $(q_t, k_t)_{t \geq 0}$ to be the equilibrium belief processes. Define $W_t(p_t)$ to be the on-path expected continuation value to the intermediary-agent pair at time t given ex-ante type p along the path of good performance

$$W_t(p) \overset{def}{=} \sup_{\tau} \mathrm{E}_p \left[\int_t^{\tau} e^{-r(s-t)} \cdot (A(q_s) - rV)\ ds + e^{-r(\tau - t)} \cdot U(\pi_\tau(p), k_\tau) \mid X_t = t \right] + V. \tag{B.70}$$

Define $V(p,t)$ as the expected ex-ante value of stopping at time t, i.e.,

$$\begin{aligned} W(p,t) &\overset{def}{=} \mathrm{E}_p \left[\int_0^{t \wedge \eta} e^{-rs} \cdot (A(q_s) - rV)\ ds + e^{-rt\wedge\eta} \cdot U(p_{t\wedge\eta}, k_{t\wedge\eta}) \right] + V \qquad \text{(B.71)} \\ &= \int_0^t e^{-rs} \cdot \left(p + (1-p) \cdot e^{-\lambda s}\right) \cdot (A(q_s) - rV)\ ds + e^{-rt} \cdot \left(p + (1-p) \cdot e^{-\lambda t}\right) \cdot U(\pi_t(p), k_t) + V. \end{aligned}$$

Suppose the intermediary and the agent of ex-ante type p separate and the agent is perceived as having a private ex-ante type l by clients. The joint ex-ante surplus of the agent and intermediary is

$$\begin{aligned} W(p,t,l) \overset{def}{=} & \int_0^t e^{-rs} \cdot \left(p + (1-p) \cdot e^{-\lambda s}\right) \cdot (A(q_s) - rV)\ ds \\ & + e^{-rt} \cdot \left(p + (1-p) \cdot e^{-\lambda t}\right) \cdot U(\pi_t(p), \pi_t(l)) + V. \end{aligned} \tag{B.72}$$

It is valuable for the intermediary and the agent to separate at time t only if the if the client assigns the agent an ex-ante belief l such that $W_0(p) < W(p,t,l)$.

(i) Denote by $D^0(p|t)$ the set of ex-ante beliefs l for which the agent is indifferent between staying on path or choosing action t:

$$D^0(p|t) \overset{def}{=} \{l : \quad W_0(p) = W(p,t,l)\} = \{d_t(p)\}, \tag{B.73}$$

where $d_t(p)$ is the unique solution to

$$W_0(p) = V(p,t,d_t(p)), \tag{B.74}$$

which is well-defined as $W(p,t,l)$ is strictly increasing in l.

(ii) Denote by $D(p|t)$ the set of ex-ante beliefs l for which it is preferable for the intermediary-agent pair to break their relationship at time t given good performance:

$$D(p|t) \stackrel{def}{=} \{l : \quad W_0(p) < W(p,t,l)\} = \{l > d_t(p)\}. \tag{B.75}$$

Definition 7. *Consider three types of refinements.*

(i) ***Divinity (D1)*** *criterion eliminates off-path action t by type p if there exists a type p' such that $D(p|t) \cup D^0(p|t) \subset D(p'|t)$.*

(ii) ***Universal Divinity*** *criterion eliminates off-path action t by type p if $D(p|t) \cup D^0(p|t) \subset \cup_{p' \neq p} D(p'|t)$.*

(iii) ***The Never Weak Best Response (NWBR)*** *criterion eliminates off-path action t by type p if $D^0(p|t) \subset \cup_{p' \neq p} D(p'|t)$.*

The following result applies Cho and Sobel (1990) to payoff structure specified by (B.3).

Lemma B.22. *In the model considered in Section 2 and summarized by the payoff structure* (B.72)*, D1 is equivalent to both Universal Divinity and NWBR.*

Proof. Suppose action t by type p is eliminated by D1, i.e., there exists a type p' such that

$$D^0(p|t) \stackrel{(i)}{\subseteq} D(p|t) \cup D^0(p|t) \subset D(p'|t) \stackrel{(i)}{\subseteq} \bigcup_{\hat{p} \neq p} D(\hat{p}|t). \tag{B.76}$$

Inclusion (ii) in (B.76) implies that type p is eliminated by Universal Divinity. Inclusions (B.76) (i) and (ii) imply that it is also eliminated by NWBR. The monotonicity of the preferences in beliefs, it follows that $D(p|t) = (d_t(p), 1]$ and $D^0(p|t) = \{d_t(p)\}$. The D1 criterion eliminates type p if there exists a p' such that $d_t(p) > d_t(p')$. Moreover, by choosing $p' = \arg\min d_t(p)$, D1 is equivalent to Universal Divinity. NWBR eliminates type p if $d_t(p) > \min_{p'} d_t(p')$, which is equivalent to D1. □

Following Lemma B.22 we refer to the D1 refinement keeping in mind its equivalence to both Universal Divinity and NWBR.

Corollary B.1 (D1 off-path beliefs)**.** *The equilibrium survives the D1 criterion if*

$$k_t \in \pi_t \left(\arg\min_p d_t(p) \right) = \arg\min_p \pi_t(d_t(p)) \qquad \textit{for every} \qquad t \notin \mathbb{T}. \tag{B.77}$$

Relationship to a Single Informed Agent Model as Motivation for D1

Our model features two informed players (the agent and the intermediary) who are interested in maximizing overall surplus, but compete on wages. Our definition of the D1 belief refinement in 7 applies on the level of agent and intermediary, meaning that they agree on a deviation jointly, and then split the benefits/costs via wages. Natural questions may arise in why this is a natural interpretation of the D1 refinement, which is normally applied to a single-informed agent games, to our setting. In what follows we show that we can formulate a single informed player model in which the payoff of this player coincides with the joint payoff of the intermediary-agent pair. The definition of D1 in our setting 7 will then coincide with the classic definition of D1 following Cho and Sobel (1990). Our subsequent analysis then shows the nonexistence of the D1 equilibrium for some parameter values, highlighting the value to our novel approach of equilibrium selection via perturbation methods in Sections B.2.

Consider a single-agent model in which a single player of ex-ante skill $\theta \in \{0,1\}$ operates a firm. The firm's revenue are a function $A(\cdot)$ of the expected ability of the player. The player of ability $\theta = 1$ always generates good performance, while the player of ability $\theta = 0$ generates a loss with intensity λ. The player has a binary signal about his ability, which we can identify with his private prior $\tilde{p} \in \{p^L, p^H\}$ about θ. The player has a profitable investment project that he can undertake at no cost, while continuing to operate his existing firm. The project yields expected value V that is constant and independent of the player's ability.

The expected value to the player from undertaking the project at time t depends on the beliefs of clients about which types undertake the project at that time. Denote these beliefs by k_t, while denote by q_t the belief about the agents who have not undertaken this new project at time t. The expected value for the player from undertaking the project at time t and being perceived as ex-ante type l is given by $W(p,t,l)$ defined in (B.72). This single player game is monotone in client beliefs, meaning that the three belief refinements outlined in 7 coincide with the classic notion of D1 refinement outlined in Cho and Sobel (1990). Everything we do from here onwards can be applied to this single informed-player game. The key distinction from the games considered in Cho and Sobel (1990) is the dependence of the payoff of the agent on client beliefs q_t prior to undertaking the irreversible public action. This both generates a quiet period for both players, but also limits the applicability of using the classic D1 belief refinement as we show below.

B.1.2 Divine Equilibrium Construction and Conditions for Existence

In this section, we construct the unique candidate equilibrium that survives the D1 refinement as defined by 7. Lemma B.25 shows that it must be separating and feature a quiet period $[0, t_1^*]$ during which all types are retained as long as they perform well, a churning period $[t_1^*, t_2^*]$ during which low type p^L agents are gradually let go, and a final time $\bar{t}$ when the high skilled p^H agents are let go. We show in Lemma B.26 that this equilibrium survives the divinity refinement 7 if the quiet period is not too long, as proxied by a sufficiently high intermediary outside option V. At the same time, surprisingly, if V is very low, then the initial quiet period is so long that no divine equilibrium may exist.

As in Section B.5 suppose the types are binary and characterized by $p \in \{p^L, p^H\}$. Denote by $\mathbb{T}^I(p^L)$ and $\mathbb{T}^I(p^H)$ to be the set of times during which types L and H are willing to separate. I.e., $\mathbb{T}^I(p^L)$ and $\mathbb{T}^I(p^H)$ are indifference sets for both the high and the low types.

Lemma B.23 (Final period separation)**.** *Suppose the equilibrium satisfies 7. Then there exists a $\bar{t} = \sup \mathbb{T}^I(p^H) \geq \sup \mathbb{T}^I(p^L)$. Moreover, it must be the case that $l_{\bar{t}} = p^H$ and $\bar{t} \in \mathbb{T}^I(p^L)$.*

Proof. The fact that separation times are bounded from above follows from Lemma A.26. Moreover, it can't be the case that $\sup \mathbb{T}^I(p^L) > \sup \mathbb{T}^I(p^H)$ since a low type agent would prefer to separate at $\sup \mathbb{T}^H$ together with the high type, rather than postpone separation and still be identified as a low type in equilibrium.

Suppose there is pooling at $\bar{t}$. Then consider an ε deviation to separate at $\bar{t} + \varepsilon$. Consider the indifference beliefs $d_{\bar{t}}(p^i)$ and $d_{\bar{t}}(p^H)$ that clients assign to agent-intermediary pairs that do not separate during $t \in [\bar{t}, \bar{t} + \varepsilon]$. These indifference beliefs must satisfy

$$U\left(\pi_{\bar{t}}(p^i), \pi_{\bar{t}}(l_{\bar{t}})\right) = \int_0^{\varepsilon} e^{-rt} \cdot \left(\pi_{\bar{t}}(p^i) + (1 - \pi_{\bar{t}}(p^i)) \cdot e^{-\lambda t}\right) \cdot \left[A\left(\pi_t(d_{\bar{t}}(p^i))\right) - rV\right] dt$$
$$+ e^{-r\varepsilon} \pi_{\bar{t}}(p^i) \cdot u_1\left(\pi_{\bar{t}}(d_{\bar{t}}(p^i))\right) + e^{-(r+\lambda)\varepsilon}\left(1 - \pi_{\bar{t}}(p^i)\right) \cdot u_0\left(\pi_{\bar{t}}(d_{\bar{t}}(p^i))\right).$$

It must be the case that $d_{\bar{t}}(p^i) \geq l_{\bar{t}}$ since the agent and the intermediary sacrifices rV during the period $[0, \varepsilon]$. For ε sufficiently small it must be the case that $d_{\bar{t}}(p^i)$ is close to $l_{\bar{t}}$ such that

$$A\left(\pi_t(d_{\bar{t}}(p^i))\right) - rV < A\left(\pi_t(l_{\bar{t}})\right)$$

for all $t \in [0, \varepsilon]$. Consider the cash flow process $C = (C_t)_{t \geq 0}$ defined as $C_t = A(\pi_t(l_{\bar{t}}))$ and cash flow process $\hat{C} = (\hat{C}_t)_{t \geq 0}$ defined as

$$\hat{C}_t = \begin{cases} A\left(\pi_t(d_{\bar{t}}(p^H))\right) - rV & \text{if} \quad t \in [\bar{t}, \bar{t} + \varepsilon), \\ A\left(\pi_t(d_{\bar{t}}(p^H))\right) & \text{if} \quad t \geq \bar{t} + \varepsilon. \end{cases}$$

By definition of $d_{\bar{t}}(p^H)$ the time $\bar{t}$ continuation value from the cash flows C and $\hat{C}$ is identical for type $\tilde{p}_0 = p^H$. However $C_t > \hat{C}_t$ if and only if $t \in [\bar{t}, \bar{t}+\varepsilon]$ cash flows C and $\hat{C}$ satisfy the single-crossing condition of Lemma A.16. This implies that the time $\bar{t}$ continuation value for type $p^L < p^H$ agent is strictly higher for cash flow stream C than cash flow stream $\hat{C}$. This, in turn, implies that the indifference belief $d_{\bar{t}}(p^L)$ must be strictly higher than $d_{\bar{t}}(p^H)$. This implies that if clients see an agent stay beyond time $\bar{t}$, following the D1 criterion, clients should assign the high type beliefs $l_t = p^H$ for $t > \bar{t}$. This then implies that if $\bar{t} \in \mathbb{T}^I(p^H)$, then it requires that $l_{\bar{t}} = p^H$, implying that there is no pooling at time $\bar{t}$.

Define $\bar{t}_L = \sup \mathbb{T}^I(p^L)$ and suppose that $\bar{t}_L < \bar{t}$. This implies that to wait until $\bar{t}$ the lowest type requires $d_{\bar{t}}(p^L) > p^H$. By continuity, it follows that if the the agent-intermediary pair choose to separate at $\bar{t} - \varepsilon$, then for a sufficiently low ε it follows that $d_{\bar{t}-\varepsilon}(p^L) > d_{\bar{t}-\varepsilon}(p^H)$. Following the D1 refinement it then must be the case that $l_{\bar{t}-\varepsilon} = p^H$, implying that type p^H agent would strictly prefer separating at $\bar{t}-\varepsilon$ at the correct belief of p^H, rather than wait until time $\bar{t}$. □

Lemma B.24 (Divine equilibrium must be separating)**.** *Suppose $A(x) = x$, $L = 0$, and $\underline{p} \geq 1/2$. Then there cannot be pooling prior to $\bar{t}$ and the high type only separates in equilibrium at time $\hat{t}$.*

Proof. Define $\hat{t} \stackrel{def}{=} \sup\left\{\mathbb{T}^I(p^H) \cap [0,\bar{t})\right\}$. Suppose $\hat{t} = \bar{t}$. This implies that there exists a sequence of times $\{t_n\} \subset \mathbb{T}^I(p^H)$ such that $t_n \to \bar{t}$. It must be the case that $\{t_n\} \subset \mathbb{T}^I(p^L)$ as well since, otherwise, it would not be incentive compatible for the p^H type to wait until $\bar{t}$. Consider the indifference belief $d_t(p^L)$ and $d_t(p^H)$ at which types p^L and p^H are willing to separate. From the previous Lemma it follows that $d_{\bar{t}}(p^L) = d_{\bar{t}}(p^H) = p^H$. It further follows that

$$\frac{d}{dt}d_t(p^i) = \frac{rV + A(\pi_t(d_t(p^i))) - A(q_t)}{\pi_t'(d_t(p^i)) \cdot \partial_2 U(\pi_t(p^i), \pi_t(d_t(p^i)))}.$$

For t close to $\bar{t}$ it follows from above that $d_t(p^L) < d_t(p^H)$. Hence it must be the case that $d_{t_n}(p^L) < d_{t_n}(p^H)$ for n sufficiently large, which contradicts with the existence of $\{t_n\} \in \mathbb{T}^I(p^L) \cap \mathbb{T}^I(p^H)$.

The above shows that $\hat{t} < \bar{t}$. It must still be the case that $\hat{t} \in \mathbb{T}^I(p^L) \cap \mathbb{T}^I(p^H)$. This implies that both types p^L and p^H are indifferent between separating at times $\hat{t}$ and $\bar{t}$. Then

$$\begin{cases} u_1(\pi_{\hat{t}}(l_{\hat{t}})) = \displaystyle\int_{\hat{t}}^{\bar{t}} e^{-r(s-\hat{t})} \cdot \big(A(q_s) - rV - A(\pi_s(p^H))\big)\, ds + u_1(\pi_{\hat{t}}(p^H)), \\ u_0(\pi_{\hat{t}}(l_{\hat{t}})) = \displaystyle\int_{\hat{t}}^{\bar{t}} e^{-(r+\lambda)(s-\hat{t})} \cdot \big(A(q_s) - rV - A(\pi_s(p^H))\big)\, ds + u_0(\pi_{\hat{t}}(p^H)). \end{cases} \tag{B.78}$$

Suppose that $l_{\hat{t}} > p^L$, meaning that there is pooling at time $\hat{t}$. Given this pooling outcome at $\hat{t}$, there exists an $\varepsilon > 0$ such that there are no separations along the equilibrium path during $(\hat{t}-\varepsilon, \hat{t}+\varepsilon)$. It must be the case that $A\left(q_{\hat{t}+}\right) - rV - A\left(\pi_{\hat{t}}(l_{\hat{t}})\right) \geq 0$ since there would otherwise be a contradiction with D1 with separations at $\hat{t}+\varepsilon$. Due to the linearity of $A(\cdot)$, the expected value of the average agent that leaves the intermediary at time $\hat{t}$ is then given by

$$\frac{\pi_{\hat{t}}\left(l_{\hat{t}}\right)}{r} \leq \frac{A\left(q_{\hat{t}+}\right) - rV}{r}.$$

The equilibrium value of the expected intermediary-agent pair is given by

$$\begin{aligned} \mathrm{E}\left[\int_{\hat{t}}^{\tau\wedge\eta} e^{-r(t-\hat{t})} \cdot (A(q_t) - rV)\, dt + \int_{\tau\wedge\eta}^{\eta} e^{-r(t-\hat{t})} A(p_t)\, dt\right] &= \mathrm{E}\left[-\int_{\hat{t}}^{\tau\wedge\eta} rV\, dt + \int_{\hat{t}}^{\eta} e^{-r(t-\hat{t})} p_t\, dt\right] \\ &> -\int_{\hat{t}}^{\infty} rV\, dt + \int_{\hat{t}}^{\infty} e^{-r(t-\hat{t})} \mathrm{E}\left[p_t\right]\, dt \end{aligned}$$

$$= \frac{q_{\hat{t}} - rV}{r}.$$

This implies that the expected value of the average intermediary-agent pair always strictly exceeds $\frac{q_t - rV}{r}$ since $A(\cdot)$ is linear. This implies that it is not incentive compatible for either type L or type H agent to separate at time $\hat{t}$, contradicting the possibility of pooling. Finally, it cannot be the case that the high type separates before all low types separate as the low type's payoff is dominated by the expected value of being perceived as the high type. □

Lemmas B.23 and B.24 above state that the low types separate prior to $\bar{t}$ and the high type separates at $\bar{t}$. Moreover, the low type agent also finds it weakly optimal to wait until time $\hat{t}$ but separates before then with certainty.

Define by t_1^* the candidate first time when it is locally optimal for the low type to leave the intermediary:

$$t_1^* \overset{def}{=} \inf \left\{ t : \quad \pi_t(q_0) - \pi_t\left(p^L\right) - rV = 0 \right\}. \tag{B.79}$$

Define by t_2^* the candidate last time when it is locally optimal for the low type to leave the intermediary:

$$t_2^* \overset{def}{=} \inf \left\{ t : \quad 1 - \pi_t\left(p^L\right) - rV = 0 \right\}. \tag{B.80}$$

Lemma B.25 (Necessary equilibrium structure)**.** *Suppose $A(x) = x$ is linear, $L = 0$, and $\underline{p} \geq 1/2$. Then in any equilibrium satisfying the D1 refinement 7 it must be the case that $\mathbb{T}^L = \{t_1^*, t_2^*\} \cup \{\bar{t}\}$ and all the low type agents separate from the intermediary gradually during $[t_1^*, t_2^*]$.*

Proof. Following Lemma B.24, the Perfect Bayesian Equilibrium satisfying the D1 refinement must be separating in types. Taken together with Lemma B.23, it follows that all low types must separate prior to $\bar{t}$, while still being indifferent to waiting until $\bar{t}$.

Consider the first time that the low type agent separates from the intermediary. Since in equilibrium it must happen at the separating belief, the intermediary-agent pair solves a stopping problem

$$\max_T \left\{ \int_0^T e^{-rt} \left(p^L + (1-p^L)e^{-\lambda t}\right) \left(A(\pi_t(q_0)) - rV\right) dt + \int_T^\infty e^{-rt} \left(p^L + (1-p^L)e^{-\lambda t}\right) A(\pi_t(p^L))\, dt \right\}.$$

Under the stated assumptions, the revenue function $A(\pi_t(p))$ is concave in t, implying that there exists a unique t_1^*, defined by (B.79) when it is optimal to let go of the low type agent.

Similarly, consider the last time when it is optimal for the low type to leave the intermediary. Since, following Lemma B.23 all low type agents must leave prior to $\bar{t}$, it must be the case that at this date $\hat{t}$ the remaining types are only the high types, meaning that $q_{\hat{t}} = 1$. Local optimality then implies that this date $\hat{t}$ must coincide with t_2^* defined in (B.80).

Finally, since $A(x) \equiv x$, it follows that $A(\pi_t(p))$ is concave in t for $p \geq 1/2$ and $t \geq 0$. This implies that there cannot be a positive gap in separations during $[t_1^*, t_2^*]$. The reason is that if there ever was a gap $(\hat{t}_1, \hat{t}_2)$, then local optimality at $\hat{t}_1$ implies that $A(q_{\hat{t}_1}) - A(\pi_{\hat{t}_1}(p^L)) = rV$. The concavity property of $A(\pi_t(p))$ then would imply that $A(\pi_{t-\hat{t}_1}(q_{\hat{t}_1})) - A(\pi_t(p^L)) < rV$ for $p \in (\hat{t}_1, \hat{t}_2]$, implying a contradiction with the optimality of waiting to separate at $\hat{t}_2$ for the low type agent. □

The above lemmas show that in order for the Perfect Bayesian Equilibrium to satisfy the D1 refinement, it must be the case that the high type separates at time $\bar{t}$, the low types separate gradually during $[t_1^*, t_2^*]$, and the low type finds it weakly optimal to wait until $\bar{t}$. In order for this to be an equilibrium, it must be the case that the beliefs disciplined by D1 during the quiet period $[0, t_1^*)$ are consistent with the above strategies.

For the equilibrium to satisfy the D1 refinement, it must be the case that for $t \in [0, t_1^*) \cup (t_2^*, \bar{t}) = [0, \bar{t}] \setminus \mathbb{T}^L$ the low type is weakly more likely to deviate than the high type, meaning that $d_t(p^L) \leq d_t(p^H)$. If this were not at satisfied at any point $\hat{t}$, then the D1 refinement would specify a high type off-equilibrium belief $k_{\hat{t}} = p_{\hat{t}}^H$, which would break the equilibrium conjectured above by making all types willing to separate at $\hat{t}$.

Lemma B.26 (Existence and nonexistence of a Divine Equilibrium)**.** *Suppose $A(\pi_t(p))$ is weakly concave in t. Then there exist two thresholds $\underline{V} < \overline{V}$. If the intermediary's outside option $V > \overline{V}$ then there exists a unique divine equilibrium characterized in Lemma B.25 and the clients' belief $k_t = p_t^L$ for all $t < \bar{t}$. If the intermediary's outside option $V < \underline{V}$, however, then a divine equilibrium does not exist.*

Proof. Lemma B.25 specifies the unique separation dynamics that may take place in a divine equilibrium. We need only verify if such separation dynamics are consistent with off path beliefs specified by the D1 refinement in (B.77).

Case $V > \overline{V}$. In the equilibrium constructed in Lemma B.25 the low type is indifferent between separating at time t_1^* and waiting until pooling with type $\bar{t}$. At the same time, the high type finds it strictly optimal to wait until separation at time $\bar{t}$. Consider $V = A(q_0) - A(p_0^L)$. This implies that the game begins immediately with the churning period. At the start of this churning period, by construction, $d_0^L = p_0^L$, while $d_0^H = p_0^H$. By continuity of the solution to (B.74), this implies that there exists $\overline{V} < A(q_0) - A(p_0^L)$ such that there is a positive length quiet period, i.e., $t_1^* > 0$ and $d_t^L < d_t^H$ for $t < t_1^*$.

Case $V < \underline{V}$. Index the churning period $[t_1^*(V), t_2^*(V)]$, the average belief process $q_t(V)$, and the indifference beliefs $d_t(p^L; V)$ and $d_t(p^H; V)$ solving (B.74) by the intermediary's outside option V. At $V = 0$ the intermediary equilibrium features no churning and, consequently, $d_t(p^L; 0) = d_t(p^H; 0) = \pi_t(q_0)$. In what follows, we show that $\frac{d}{dV} d_t(p^H; V)\big|_{V=0} < \frac{d}{dV} d_t(p^L; V)\big|_{V=0}$, implying that the equilibrium constructed in Lemma B.25 does not satisfy the divinity refinement at the start of the quiet period when V is sufficiently low.

During the quiet period $t \in [t_1^*(V), t_2^*(V)]$ it follows that

$$A(q_t(V)) - A(p_t^L) = rV. \tag{B.81}$$

The indifference condition to separate at time $\bar{t}(V)$ is given by

$$\begin{aligned}&\int_0^{t_1^*(V)} e^{-rs} \cdot \left(p^L + (1-p^L) \cdot e^{-\lambda s}\right) \cdot \left(A(q_s(V)) - rV - A(p_s^L)\right) \, ds + U(p^L, p^L) \\ =&\int_0^{\bar{t}(V)} e^{-rs} \cdot \left(p^L + (1-p^L) \cdot e^{-\lambda s}\right) \cdot \left(A(q_s(V)) - rV - A(p_s^H)\right) \, ds + U(p^L, p^H).\end{aligned}$$

Given the indifference condition (B.81), the derivative with respect to V of both the left and right-hand side is given by

$$\begin{aligned}&\int_0^{t_1^*(V)} e^{-rs} \cdot \left(p^L + (1-p^L) \cdot e^{-\lambda s}\right) \cdot \left(A'(q_s(V)) \cdot q_s'(V) - r\right) \, ds \\ =&\bar{t}'(V) \cdot \left(p^L + (1-p^L) \cdot e^{-\lambda \bar{t}(V)}\right) \cdot \left(A(q_{\bar{t}(V)}(V)) - rV - A(p_{\bar{t}(V)}^H)\right) \\ &+ \int_0^{\bar{t}(V)} e^{-rs} \cdot \left(p^L + (1-p^L) \cdot e^{-\lambda s}\right) \cdot \left(A'(q_s(V)) \cdot q_s'(V) - r\right) \, ds\end{aligned} \tag{B.82}$$

Note that $q_{\bar{t}(V)} = p_{\bar{t}(V)}^H$. Consequently (B.82) can be simplified to

$$\begin{aligned}&\int_0^{t_1^*(V)} e^{-rs} \cdot \left(p^L + (1-p^L) \cdot e^{-\lambda s}\right) \cdot \left(A'(q_s(V)) \cdot q_s'(V) - r\right) \, ds \\ =& -\bar{t}'(V) \cdot \left(p^L + (1-p^L) \cdot e^{-\lambda \bar{t}(V)}\right) \cdot rV \\ &+ \int_0^{\bar{t}(V)} e^{-rs} \cdot \left(p^L + (1-p^L) \cdot e^{-\lambda s}\right) \cdot \left(A'(q_s(V)) \cdot q_s'(V) - r\right) \, ds\end{aligned} \tag{B.83}$$

Consider now $V = 0$. In this case $t_1^*(V) = +\infty$, $q_s(V) = \pi_s(q_0)$, and $q_s'(0) = 0$. Then (B.83) becomes

$$\begin{aligned}&\int_0^{\infty} e^{-rs} \cdot \left(p^L + (1-p^L) \cdot e^{-\lambda s}\right) \cdot (A(\pi_s(q_0)) \cdot 0 - r) \, ds \\ =& -\bar{t}'(0) \cdot \left(p^L + (1-p^L) \cdot e^{-\lambda \bar{t}(0)}\right) \cdot r \\ &+ \int_0^{\infty} e^{-rs} \cdot \left(p^L + (1-p^L) \cdot e^{-\lambda s}\right) \cdot \left(A'(\pi_s(q_0)) \cdot 0 - r\right) \, ds\end{aligned}$$

The above indifference condition implies that

$$\lim_{V \to 0} \bar{t}'(V) \cdot \left(p^L + (1-p^L) \cdot e^{-\lambda \bar{t}(V)}\right) \cdot rV = 0. \tag{B.84}$$

Consider now the indifference beliefs $d_0(p^L; V)$ and $d_0(p^H; V)$ for the low and high types, respectively. The indifference belief $d_0^i(V)$ solves

$$U(p^i, d_0^i(V)) = \int_0^{\bar{t}(V)} e^{-rs} \cdot \left(p^i + (1-p^i) \cdot e^{-\lambda s}\right) \cdot \left(A(q_s(V)) - rV - A(p_s^H)\right) \, ds + U(p^i, p^H).$$

Differentiating both sides with respect to V obtain

$$\begin{aligned}\partial_2 U(p^i, d_0^i(V)) \cdot \frac{d}{dV} d_0^i(V) = & \int_0^{\bar{t}(V)} e^{-rs} \cdot \left(p^i + (1-p^i) \cdot e^{-\lambda s}\right) \cdot \left(A(q_s(V)) \cdot q_s'(V) - r\right) \, ds \\ & -\bar{t}'(V) \cdot e^{-r\bar{t}(V)} \cdot \left(p^i + (1-p^i) \cdot e^{-\lambda \bar{t}(V)}\right) \cdot rV.\end{aligned}$$

It follows from (B.84) that at $V = 0$ obtain

$$\partial_2 U(p^i, q_0) \cdot \frac{d}{dV} d_0^i(0) = -\int_0^{\infty} e^{-rs} \cdot \left(p^i + (1-p^i) \cdot e^{-\lambda s}\right) \cdot r \, ds$$

where we used the fact that $d_0^i(0) = q_0$. For the divine equilibrium to not exist we need $\frac{d}{dV} d_0^H(0) < \frac{d}{dV} d_0^L(0)$, which can be expressed as

$$\begin{aligned}
\frac{-\int_0^\infty e^{-rs} \cdot \left(p^H + (1-p^H) \cdot e^{-\lambda s}\right) \cdot r \, ds}{\partial_2 U(p^H, q_0)} &< \frac{-\int_0^\infty e^{-rs} \cdot \left(p^L + (1-p^L) \cdot e^{-\lambda s}\right) \cdot r \, ds}{\partial_2 U(p^L, q_0)} \\
\frac{\int_0^\infty e^{-rs} \cdot \left(p^H + (1-p^H) \cdot e^{-\lambda s}\right) \, ds}{\partial_2 U(p^H, q_0)} &> \frac{\int_0^\infty e^{-rs} \cdot \left(p^L + (1-p^L) \cdot e^{-\lambda s}\right) \, ds}{\partial_2 U(p^L, q_0)} \\
\partial_2 U(p^L, q_0) \cdot \int_0^\infty e^{-rs} \cdot \left(p^H + (1-p^H) \cdot e^{-\lambda s}\right) \, ds &> \partial_2 U(p^H, q_0) \cdot \int_0^\infty e^{-rs} \cdot \left(p^L + (1-p^L) \cdot e^{-\lambda s}\right) \, ds \\
\partial_2 U(p^L, q_0) \cdot \left(\frac{p^H}{r} + \frac{1-p^H}{r+\lambda}\right) &> \partial_2 U(p^H, q_0) \cdot \left(\frac{p^L}{r} + \frac{1-p^L}{r+\lambda}\right) \\
\left(p^L \cdot u_1'(q_0) + (1-p^L) \cdot u_0'(q_0)\right) \cdot \left(\frac{p^H}{r} + \frac{1-p^H}{r+\lambda}\right) &> \left(p^H \cdot u_1'(q_0) + (1-p^H) \cdot u_0'(q_0)\right) \cdot \left(\frac{p^L}{r} + \frac{1-p^L}{r+\lambda}\right) \\
\left(p^L \left(\frac{p^H}{r} + \frac{1-p^H}{r+\lambda}\right) - p^H \left(\frac{p^L}{r} + \frac{1-p^L}{r+\lambda}\right)\right) u_1'(q_0) &< \left((1-p^H)\left(\frac{p^L}{r} + \frac{1-p^L}{r+\lambda}\right) - (1-p^L)\left(\frac{p^H}{r} + \frac{1-p^H}{r+\lambda}\right)\right) u_0'(q_0) \\
\frac{p^L - p^H}{r+\lambda} \cdot u_1'(q_0) &> \frac{p^L - p^H}{r} \cdot u_0'(q_0) \\
r \cdot u_1'(q_0) &< (r+\lambda) \cdot u_0'(q_0). \qquad \text{(B.85)}
\end{aligned}$$

Note that

$$\begin{aligned}
r \cdot u_1'(q_0) &= r \cdot \int_0^\infty e^{-rt} \cdot A'(\pi_t(q_0)) \cdot \frac{e^{-\lambda t}}{(q_0 + (1-q_0) \cdot e^{-\lambda t})} \, dt \\
&= r \cdot \int_0^\infty e^{-rt} \cdot \frac{1}{q_0(1-q_0)} \cdot \frac{d}{dt} A(\pi_t(q_0)) \, dt \\
&= -\frac{1}{q_0(1-q_0)} \cdot \int_0^\infty \frac{d}{dt} A(\pi_t(q_0)) \, de^{-rt}
\end{aligned}$$

$$= \frac{1}{q_0(1-q_0)} \cdot \frac{d}{dt} A(\pi_t(q_0))\Big|_{t=0} + \frac{1}{q_0(1-q_0)} \cdot \int_0^\infty e^{-rt} \underbrace{\frac{d^2}{dt^2} A(\pi_t(q_0))}_{\leq 0} dt$$

$$< \frac{1}{q_0(1-q_0)} \cdot \frac{d}{dt} A(\pi_t(q_0))\Big|_{t=0} + \frac{1}{q_0(1-q_0)} \cdot \int_0^\infty e^{-(r+\lambda)t} \underbrace{\frac{d^2}{dt^2} A(\pi_t(q_0))}_{\leq 0} dt = (r+\lambda) \cdot u_0'(q_0).$$

The above chain of inequalities relies on the fact that $A(\pi_t(p))$ is weakly concave in t and $e^{-rt} > e^{-(r+\lambda)t}$. Consequently, inequality (B.85) is satisfied whenever $A(\pi_t(p))$ is weakly concave in t, which then implies that $\frac{d}{dV} d_0^H(0) < \frac{d}{dV} d_0^L(0)$. Since $d_0^L(0) = d_0^H(0) = q_0$, it then implies, by continuity, that there exists an $\underline{V}$ such that for all $V < \underline{V}$ the equilibrium constructed in Lemma B.25 does not satisfy the divinity refinement (B.77) during the initial quiet period. □

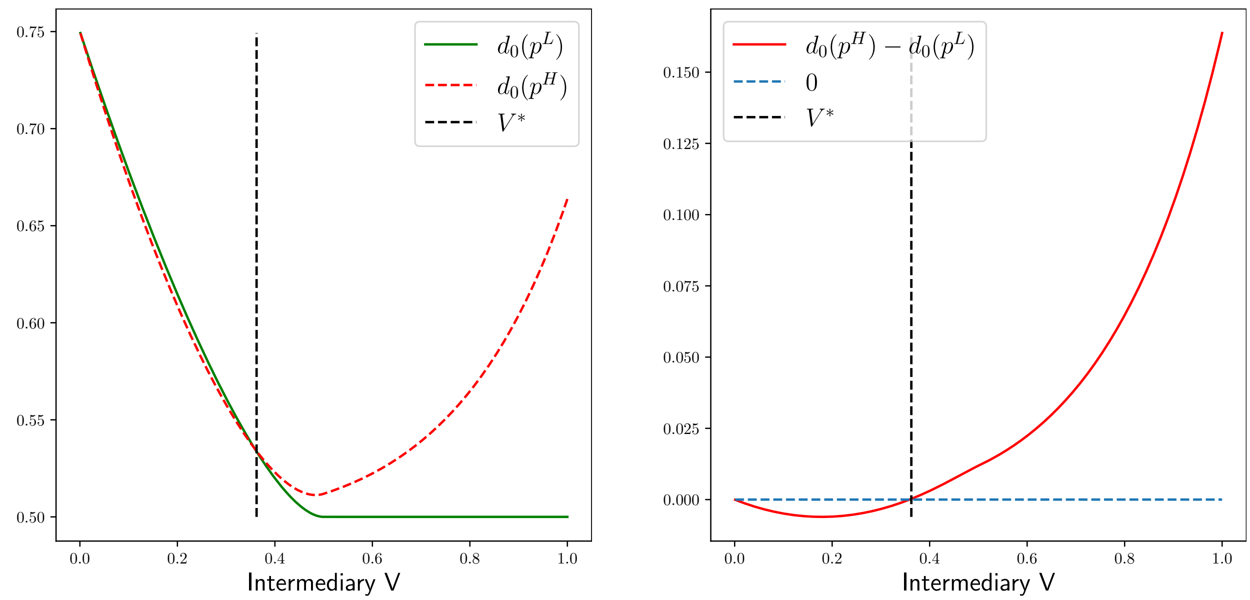
Lemma B.26 shows analytically that when a divine equilibrium exists, it takes the quiet-churning structure as the equilibrium we construct in Section 3 of the main text. We show, however, that when the intermediary's outside option V is very low, then the resulting quiet period delivers sufficient rents to low skilled agents that they become less likely to deviate. The greater propensity of higher types to separate, as captured by $d_0^H < d_0^L$ for low $V's$ occurs due to the fact that if the initial quiet period is very long, then the unique candidate divine equilibrium constructed in Section B.1.2 looks very similar to pooling from the perspective of the high type. Put differently, the churning dynamics occur sufficiently late in the game that the equilibrium rents obtained by the low-type intermediary-agent pair exceed those of the high-type intermediary-agent pair, leading the latter to be more likely to deviate. This, however, implies that a divine equilibrium may not exist since if clients attribute an off-path deviation to a high type at the start of the game, then all types would separate, resulting in immediate separation for all types. Such pooling, however is again precluded by the divinity refinement as the high type intermediary-agent pair would be more willing to delay separation. This implies that if the initial quiet period is very long, which holds if, for example, the intermediary's outside option V is very low, then there may not be a Perfect Bayesian Equilibrium satisfying the divinity refinement 7. In the following Section, we provide a semi-analytic example of the ranking of indifference beliefs $d_0(p^L; V)$ and $d_0(p^H; V)$ as a function of the intermediary's outside option.

B.1.3 Numerical Illustration of Divine Equilibrium Existence and Nonexistence

Figure B.1 provides an numerical illustration of this argument by plotting the indifference beliefs $d_0(p^L; V)$ and $d_0(p^H; V)$ at the start of the game of the low and high type agents as a a function of the intermediary's outside option V. We obtain Figure B.1 in a semi-analytic parametrization of the model that we derive below in order to maximize tractability and precision of the numerical exercise.

Figure B.1a plots the indifference beliefs $d_0(p^L)$ and $d_0(p^H)$, whereas Figure B.1b plots the difference

$d_0(p^H) - d_0(p^L)$ as a function of V. We see in Figure B.1b that there exists an $\underline{V} = \overline{V}$ such that $d_0(p^H; V) < d_0(p^L; V)$ for $V < \underline{V}$ and $d_0(p^H; V) > d_0(p^L; V)$ for $V > \underline{V}$. For the equilibrium to survive the divinity refinement 7 it must then be the case that clients attribute separations at $t = 0$ to the high type agent, meaning that $l_0 = p^H$. This, however, violates the incentive compatibility of the unique candidate divine equilibrium constructed in Section B.1.2 since all intermediary-agent pairs would then prefer to separate immediately at $t = 0$. Following the arguments of Section B.1.2, however, such pooling at $t = 0$ also violates the divinity refinement 7 as it then creates an incentive for higher-skilled agents to delay separation to signal their ability.



(a) Indifference beliefs $d_0(p^L)$ and $d_0(p^H)$ at $t = 0$ as a function of the intermediary's outside option V.

(b) The difference in indifference beliefs. The equilibrium does not survive the divinity criterion whenever $d_0(p^H) - d_0(p^L) < 0$.

Figure B.1: Indifference beliefs and equilibrium nonexistence for low intermediary values V. Model parameters: $r = \lambda = 0.5$, $p^L = 0.5$, $p^H = 1$, $q_0 = 0.75$, $A(x) \equiv x$.

The intuition is that for a low $V < \underline{V}$, the quiet period is very long. Such a long pooling period is beneficial for the low types and makes them less willing to deviate at $t = 0$ relative to the high types. The intuition for why D1 breaks down this initial long pooling period is identical to why D1 breaks pooling in the final period – the high type is willing to sacrifice some short-term rates to avoid prolonged pooling with the low type, regardless of whether it takes place early or later in the game. Consequently, even though the divinity criterion disciplines the game starting from the churning period, as we have shown by Lemma B.25 above, the backward induction implications result in an inconsistency at the very start of the game. The contradiction illustrated by Figure B.1 for $V < \underline{V}$ highlights the challenges in applying the divinity refinement to our setting due to the endogenous rents obtained by the low type agent from the necessary period of initial pooling. At the same time, Figure B.1 does show that for $V > \underline{V}$ the quiet period is sufficiently short that the higher skilled agents become less likely to deviate, as proxied by $d_t(p^H; V) > d_t(p^L; V)$ impying that the equilibrium constructed in Lemma B.25 survives the D1 refinement 7.

Parameters and functional forms. In what follows, we provide additional details on the model parametrization used to obtain Figure B.1 – we consider a rather intuitive semi-analytic parametrization of the model in order to maximize tractability and, ultimately, numerical precision. We consider a special case of the discount rate r equal to the public news informativeness λ, i.e., $r = \lambda$, to obtain analytical expressions for the agent's outside option $U(p, k)$. Also, for simplicity, we assume that $L = 0$.[21] None of the previous results relied on the relative comparison of r and λ, so constructing such an example is without loss.[22] Under such parametrization, the expected value to $\theta = 1$ agent from reputation k is equal to

$$\begin{aligned} u_1(k) &= \int_0^\infty e^{-rt} \cdot \frac{k}{k + (1-k) \cdot e^{-\lambda t}} \, dt = \int_0^\infty e^{-\lambda t} \cdot \frac{k}{k + (1-k) \cdot e^{-\lambda t}} \, dt \\ &= -\frac{1}{\lambda} \int_0^\infty \frac{k}{k + (1-k) \cdot e^{-\lambda t}} \, de^{-\lambda t} = \frac{1}{\lambda} \int_0^1 \frac{k}{k + (1-k) \cdot x} \, dx \overset{\lambda = r}{=} -\frac{1}{r} \frac{k}{1-k} \ln(k). \end{aligned}$$

The value to $\theta = 0$ agent from reputation k is equal to

$$\begin{aligned} u_0(k) &= \int_0^\infty e^{-(r+\lambda)t} \cdot \frac{k}{k + (1-k) \cdot e^{-\lambda t}} \, dt = \int_0^\infty e^{-2\lambda t} \cdot \frac{k}{k + (1-k) \cdot e^{-\lambda t}} \, dt \\ &= -\frac{1}{\lambda} \int_0^\infty \frac{k \cdot e^{-\lambda t}}{k + (1-k) \cdot e^{-\lambda t}} \, de^{-\lambda t} = \frac{1}{\lambda} \int_0^1 \frac{k \cdot x}{k + (1-k) \cdot x} \, dx = \frac{1}{\lambda} \int_0^1 \frac{k}{1-k} \frac{k + (1-k) \cdot x - k}{k + (1-k) \cdot x} \, dx \\ &= \frac{1}{\lambda} \cdot \frac{k}{1-k} - \frac{k}{1-k} \cdot \frac{1}{\lambda} \int_0^1 \frac{k}{1-k} \frac{k}{k + (1-k) \cdot x} \, dx \\ &= \frac{1}{\lambda} \cdot \frac{k}{1-k} + \frac{k}{1-k} \cdot \frac{1}{\lambda} \frac{k}{1-k} \ln(k) \overset{\lambda = r}{=} \frac{1}{r} \cdot \frac{k}{1-k} \cdot \left(1 + \frac{k}{1-k} \cdot \ln(k)\right). \end{aligned}$$

Due to the risk-neutrality of the agent and the linearity of the production function, it follows that as long as the agent's belief about his own ability coincides with clients' belief about his ability, then his expected value is equal to $\frac{k}{r}$:

$$k \cdot u_1(k) + (1-k) \cdot u_0(k) = \frac{k}{r}.$$

Low type's separation interval $[t_1^*, t_2^*]$. The next step is to solve for t_1^* and t_2^*. For an arbitrary q and p solve for t such that

$$\pi_t(q) - \pi_t(p) = rV$$

[21] All our previous results continue to hold in this case as the low type agent is exactly indifferent between staying in the industry and leaving conditional on receiving a bad performance shock.

[22] We have constructed other, purely numerical, examples in which $r \neq \lambda$ and the results of this section are unchanged.

$$\frac{q}{q+(1-q)e^{-\lambda t}} - \frac{p}{p+(1-p)e^{-\lambda t}} = rV$$
$$q \cdot \left(p+(1-p)e^{-\lambda t}\right) - p \cdot \left(q+(1-q)e^{-\lambda t}\right) = rV \cdot \left(p+(1-p)e^{-\lambda t}\right) \cdot \left(q+(1-q)e^{-\lambda t}\right).$$

Simplify terms to obtain

$$(q-p) \cdot e^{-\lambda t} = rV \cdot \left(p+(1-p)e^{-\lambda t}\right) \cdot \left(q+(1-q)e^{-\lambda t}\right)$$
$$(q-p) \cdot e^{-\lambda t} = rV \cdot (pq + (q+p-2qp) \cdot e^{-\lambda t} + (1-p)(1-q)e^{-2\lambda t}).$$

This equation simplifies to a quadratic equation in $e^{-\lambda t}$, given by

$$e^{-2\lambda t} \cdot (1-p)(1-q) + e^{-\lambda t} \cdot \left(q+p-2qp - \frac{q-p}{rV}\right) + pq = 0,$$

which admits an analytic solution, which we denote by $t(p,q)$. Using such notation, we can express in closed form $t_1^* = t(p^L, q_0)$ and $t_2^* = t(p^L, 1)$.

High type's separation time $\bar{t}$. Having characterized t_2^* we can use it to characterize time $\bar{t}$ when the high type separates. Following Lemma B.23 the low type agent must be indifferent between separating at t_2^* and separating at $\bar{t}$:

$$U\left(p^L_{t_2^*}, p^L_{t_2^*}\right) + V = -\, p^L_{t_2^*} \cdot \int_{t_2^*}^{\bar{t}} e^{-r(t-t_2^*)} \cdot rV\, dt - (1-p^L_{t_2^*}) \cdot \int_{t_2^*}^{\bar{t}} e^{-(r+\lambda)(t-t_2^*)} \cdot rV\, dt$$
$$+\, p^L_{t_2^*} \cdot \frac{1}{r} + (1-p^L_{t_2^*}) \cdot \frac{1}{r+\lambda} + V,$$

where $p_t^L \overset{def}{=} \pi_t(p^L)$. By definition of t_2^* in (B.80) and the linearity of $A(\cdot)$ it follows that $p^L_{t_2^*} = 1 - rV$. Then

$$\frac{1-rV}{r} = -(1-rV) \cdot \int_{t_2^*}^{\bar{t}} e^{-r(t-t_2^*)} \cdot rV\, dt - rV \cdot \int_{t_2^*}^{\bar{t}} e^{-(r+\lambda)(t-t_2^*)} \cdot rV\, dt + \frac{1-rV}{r} + \frac{rV}{r+\lambda}$$
$$\frac{rV}{r+\lambda} = (1-rV) \cdot \int_{t_2^*}^{\bar{t}} e^{-r(t-t_2^*)} \cdot rV\, dt + rV \cdot \int_{t_2^*}^{\bar{t}} e^{-(r+\lambda)(t-t_2^*)} \cdot rV\, dt$$
$$\frac{r}{r+\lambda} = (1-rV) \cdot \left(1 - e^{-r(t_2^*-\bar{t})}\right) + \frac{rV}{r+\lambda} \cdot \left(1 - e^{-(r+\lambda)(\bar{t}-t_2^*)}\right).$$
$$0 = \frac{rV}{r+\lambda} \cdot e^{-(r+\lambda)(\bar{t}-t_2^*)} + (1-rV) \cdot e^{-r(\bar{t}-t_2^*)} + \frac{r}{r+\lambda} - (1-rV) - \frac{rV}{r+\lambda}$$
$$0 = \frac{rV}{r+\lambda} \cdot e^{-(r+\lambda)(\bar{t}-t_2^*)} + (1-rV) \cdot e^{-r(\bar{t}-t_2^*)} - \frac{\lambda}{r+\lambda} + rV - \frac{rV}{r+\lambda}$$
$$0 = \frac{rV}{r+\lambda} \cdot e^{-(r+\lambda)(\bar{t}-t_2^*)} + (1-rV) \cdot e^{-r(\bar{t}-t_2^*)} + \frac{rV(r+\lambda-1)-\lambda}{r+\lambda}.$$

The above is a decreasing function in $\bar{t}$, implying that there is a unique solution. Suppose that $r = \lambda$. Then the above is a quadratic equation for $e^{-r(\bar{t}-t_2^*)}$, which simplifies to

$$\frac{V}{2} \cdot e^{-2r(\bar{t}-t_2^*)} + (1 - rV) \cdot e^{-r(\bar{t}-t_2^*)} + \frac{V(2r-1)-1}{2} = 0.$$

and admits an analytic solution.

Expected equilibrium payoffs. The expected payoff at $t = 0$ to the intermediary-agent pair employing the low type agent is given by $W_0(p^L)$ given by

$$\begin{aligned} W_0(p^L) = &\int_0^{t_1^*} e^{-rt} \cdot \left(p^L + (1-p^L) \cdot e^{-\lambda t}\right) \cdot (A(\pi_t(q_0)) - rV)\, dt \\ &+ e^{-rt^*} \cdot p^L \cdot u_1(\pi_{t_1^*}(p^L)) + e^{-(r+\lambda)t^*} \cdot (1-p^L) \cdot u_0(\pi_{t_1^*}(p^L)) + V. \end{aligned}$$

The payoff to the high type in this candidate equilibrium is

$$\begin{aligned} W_0(p^H) &= \int_0^{\bar{t}} e^{-rt} \cdot (A(q_t) - rV)\, dt + e^{-r\bar{t}} \cdot u_1(\pi_{\bar{t}}(p^H)) + V \\ &= \int_0^{t_1^*} e^{-rt} \cdot (A(\pi_t(q_0)) - rV)\, dt + \int_{t_1^*}^{t_2^*} e^{-rt} \cdot A(\pi_t(p^L))\, dt \\ &+ \int_{t_2^*}^{\bar{t}} e^{-rt} \cdot \left(A(\pi_t(p^H)) - rV\right)\, dt + e^{-r\bar{t}} \cdot u_1(1) + V. \end{aligned}$$

Having derived the above expressions analytically, we compute the indifference beliefs $d_0(p^L)$ and $d_0(p^H)$ numerically to obtain Figure B.1.

References

Acemoglu, D. and J.-S. Pischke (1998). "Why do firms train? Theory and evidence". *The Quarterly Journal of Economics* 113.1, 79–119.

Acemoglu, D. and J.-S. Pischke (1999). "Beyond Becker: training in imperfect labour markets". *The economic journal* 109.4, 112–142.

Axelson, U. and P. Bond (2015). "Wall street occupations". *The Journal of Finance* 70.5, 1949–1996.

Bar-Isaac, H. and C. Leaver (2022). "Training, recruitment, and outplacement as endogenous adverse selection". *Economica* 89.356, 849–861.

Bar-Isaac, H. and R. Lévy (2022). "Motivating employees through career paths". *Journal of Labor Economics* 40.1, 95–131.

Barlevy, G. and D. Neal (2019). "Allocating effort and talent in professional labor markets". *Journal of Labor Economics* 37.1, 187–246.

Becker, G. S. (1962). "Investment in human capital: A theoretical analysis". *Journal of political economy* 70.5, Part 2, 9–49.

Bernhardt, D. (1995). "Strategic promotion and compensation". *The Review of Economic Studies* 62.2, 315–339.

Cho, I.-K. and D. M. Kreps (1987). "Signaling games and stable equilibria". *The Quarterly Journal of Economics* 102.2, 179–221.

Cho, I.-K. and J. Sobel (1990). "Strategic stability and uniqueness in signaling games". *Journal of Economic Theory* 50.2, 381–413.

Daley, B. and B. Green (2012). "Waiting for News in the Market for Lemons". *Econometrica* 80.4, 1433–1504.

Davis, M. H. (2018). *Markov models & optimization.* Routledge.

Diamond, P. A. (1971). "A model of price adjustment". *Journal of economic theory* 3.2, 156–168.

Farber, H. S. and R. Gibbons (1996). "Learning and wage dynamics". *The Quarterly Journal of Economics* 111.4, 1007–1047.

Ghosh, S. and M. Waldman (2010). "Standard promotion practices versus up-or-out contracts". *The RAND Journal of Economics* 41.2, 301–325.

Gibbons, R. and M. Waldman (1999). "A theory of wage and promotion dynamics inside firms". *The Quarterly Journal of Economics* 114.4, 1321–1358.

Gilson, R. J. and R. H. Mnookin (1989). "Coming of age in a corporate law firm: The economics of associate career patterns". *Stanford Law Review*, 567–595.

Greenwald, B. C. (1986). "Adverse selection in the labour market". *The Review of Economic Studies* 53.3, 325–347.

Jovanovic, B. (1979). "Job matching and the theory of turnover". *Journal of political economy* 87.5, Part 1, 972–990.

Kahn, C. and G. Huberman (1988). "Two-sided uncertainty and" up-or-out" contracts". *Journal of Labor Economics* 6.4, 423–444.

Kohlberg, E. and J.-F. Mertens (1986). “On the strategic stability of equilibria”. *Econometrica: Journal of the Econometric Society*, 1003–1037.

Kremer, I. and A. Skrzypacz (2007). “Dynamic signaling and market breakdown”. *Journal of Economic Theory* 133.1, 58–82.

McKelvey, R. D. and T. R. Palfrey (1995). “Quantal response equilibria for normal form games”. *Games and economic behavior* 10.1, 6–38.

— (1998). “Quantal response equilibria for extensive form games”. *Experimental economics* 1, 9–41.

Noldeke, G. and E. Van Damme (1990). “Signalling in a dynamic labour market”. *The Review of Economic Studies* 57.1, 1–23.

Selten, R. and R. S. Bielefeld (1988). *Reexamination of the perfectness concept for equilibrium points in extensive games*. Springer.

Spence, M. (1973). “Job Market Signaling”. *The Quarterly Journal of Economics* 87.3, 355–374.

Swinkels, J. M. (1999). “Education signalling with preemptive offers”. *The Review of Economic Studies* 66.4, 949–970.

Waldman, M. (1984). “Job assignments, signalling, and efficiency”. *The RAND Journal of Economics* 15.2, 255–267.

— (1990). “Up-or-out contracts: A signaling perspective”. *Journal of Labor Economics* 8.2, 230–250.