Online Appendix: Reputation Building under Observational Learning

Harry PEI

July 5, 2022

This document consists of four sections. In Section A, I show the *no-herding result* Proposition 2. In Section B, I examine the patient player's undiscounted average payoff. I show Proposition 3 in Section B.1, and establish the tightness of the lower bound in Proposition 3 in Section B.2. I show that the conclusion of Proposition 3 fails when both K and M are finite in Section B.3, under a mild regularity condition on players' stage-game payoffs. I show Theorems 2 and 3 in Section C. I also establish the existence of equilibrium when the private signal is unboundedly informative. In Section D, I use a counterexample to show that the conclusion of Proposition 1 fails when the short-run players can directly observe calendar time.

A Proof of Proposition 2

First, I establish the result when $M=+\infty$. Suppose by way of contradiction that player 2s herd on $b \neq b^*$ at h^t, then the strategic type has no intertemporal incentive at h^t and at every h_*^t that differs from h^t only in $\{a_0,...,a_{t-K}\}$. In equilibrium, strategic-type player 1 plays his myopic best reply to b at those histories. Consider two cases. First, suppose $\mathrm{BR}_1(b)=\{a^*\}$, then in equilibrium, both types of player 1 play a^* at h^t and at every h_*^t that differs from h^t only in $\{a_0,...,a_{t-K}\}$. As a result, player 2_t has a strict incentive to play b^* instead of b at h^t. This contradicts the hypothesis that $b \neq b^*$. Second, suppose $\mathrm{BR}_1(b) \neq \{a^*\}$, then in equilibrium, the strategic type has no incentive to play a^* at h^t and at every h_*^t that differs from h^t only in $\{a_0,...,a_{t-K}\}$. Since $\pi(h^t)>0$, player 2_{t+1}'s belief assigns probability 1 to the commitment type if she observes $a_t=a^*$, and player 1's actions from period $t-K+1$ to $t-1$ and player 2's actions from period 0 to $t-1$ are given according to h^t. Therefore, player 2_{t+1} plays b^* following the aforementioned observation, which contradicts the hypothesis that they herd on $b \neq b^*$.

Next, I establish the result when M is finite and is at least one. Suppose by way of contradiction that player 2s herd on $b \neq b^*$ at h^t. Since K and M are both finite, player 2_t's action must be measurable with respect to

$$(a_{\max\{0,t-K\}}, ..., a_{t-1}, b_{\max\{0,t-M\}}, ..., b_{t-1}).$$

For every $t \geq \max\{M, K\}$, $(a_{t-K}, ..., a_{t-1}, b_{t-M}, ..., b_{t-1})$, and $h^t \equiv (a_s, b_s)_{s \leq t-1}$, there exists $h^T \succ h^t$ such that player 2's action at h^T coincides with her action at $(a_{t-K}, ..., a_{t-1}, b_{t-M}, ..., b_{t-1})$. Therefore, player 2s herding on $b \neq b^*$ at any $h^t \equiv (a_s, b_s)_{s \leq t-1}$ implies that they play b in every period after $\max\{K, M\}$. Hence, the strategic-type player 1 has no intertemporal incentive after period T. Consider two cases. Suppose $\text{BR}_1(b) \neq \{a^*\}$, then the strategic type has no incentive to play a^*, so player 2 assigns probability 1 to the commitment type after observing a^*, which means that player 2 has a strict incentive to play b^*. This contradicts the hypothesis that they herd on action $b \neq b^*$. Suppose $\text{BR}_1(b) \neq \{a^*\}$, then in equilibrium, the strategic type has no incentive to play a^* and player 2 has a strict incentive to play b^*. This contradicts the hypothesis that they herd on action $b \neq b^*$.

B Player 1's Undiscounted Average Payoff

This Appendix consists of three parts. I establish Proposition 3 in Section B.1. I establish the tightness of the payoff lower bound in Section B.2. I show that the conclusion of Proposition 3 fails when both K and M are finite in Section B.3.

B.1 Proof of Proposition 3: Lower Bound on Undiscounted Average Payoff

I show that when $M = +\infty$ and players' payoffs satisfy Assumptions 1 and 2 in the main text, for every $\delta \in (0,1)$, $\pi_0 \in (0,1)$, and every strategy profile (σ_1, σ_2) that is part of a PBE, we have:

$$\liminf_{t \to \infty} \frac{1}{t} \mathbb{E}^{(a^*, \sigma_2)} \Big[\sum_{s=0}^{t-1} u_1(a_s, b_s) \Big] \geq \frac{K}{K+1} u_1(a^*, b^*) + \frac{1}{K+1} u_1(a^*, b'). \tag{B.1}$$

Consider the strategic type's payoff when he deviates and imitates the commitment type. For every $\beta \in \Delta(B)$ and $a \prec a^*$, Assumption 2 implies that $u_1(a^*, \beta) < u_1(a, \beta)$. Let $h^t \equiv \{a_s, b_s\}_{s=0}^{t-1}$. For every $t \in \mathbb{N}$ and $a \in A$, let $E_t(a, b^t)$ be the event that (i) player 1 plays a in period t, (ii) player 1 has played a^* from period $t - K + 1$ to $t - 1$, (iii) player 1 plays according to σ_1 starting from period $t + 1$, and (iv) the history of player 2's actions until period t is $b^t \equiv (b_0, ..., b_{t-1})$.

For every $\tau \in \{1, 2, ..., K\}$ and $h^t \equiv (a^*, ..., a^*, b^t)$, let $y_t^\tau(\cdot|a, h^t) \in \Delta(B)$ be the distribution of $b_{t+\tau}$ conditional on event $E_t(a, b^t)$, and let $y_t(\cdot|a, h^t) \in \Delta(B^K)$ be the distribution of $(b_{t+1}, ..., b_{t+K})$ conditional on event $E_t(a, b^t)$. Let $\overline{u}_1$ and $\underline{u}_1$ be player 1's highest and lowest feasible stage-game payoffs, respectively, and let $||\cdot||$ be the total variation norm. If

$$||y_t(\cdot|a^*, h^t) - y_t(\cdot|a, h^t)|| \leq \frac{1-\delta}{2\delta(\overline{u}_1 - \underline{u}_1)}\Big(u_1(a, \beta) - u_1(a^*, \beta)\Big), \tag{B.2}$$

then the strategic-type player 1 has a strict incentive to play a instead of a^* at h^t as well as at every history h^t_* that differs from h^t only in terms of $\{a_0, ..., a_{t-K}\}$. The latter is because the distribution of $\{b_{t+1}, ..., b_{t+K}\}$ does not depend on $\{a_0, ..., a_{t-K}\}$ since they cannot be observed by players 2_{t+1} to 2_{t+K}. Let

$$\Delta \equiv \frac{1-\delta}{2K\delta(\overline{u}_1 - \underline{u}_1)} \min_{\beta \in \Delta(B), a \prec a^*} \Big\{u_1(a, \beta) - u_1(a^*, \beta)\Big\}. \tag{B.3}$$

Since

$$||y_t^\tau(\cdot|a^*, h^t) - y_t^\tau(\cdot|a, h^t)|| \leq ||y_t(\cdot|a^*, h^t) - y_t(\cdot|a, h^t)|| \leq \sum_{s=1}^{K} ||y_t^s(\cdot|a^*, h^t) - y_t^s(\cdot|a, h^t)||,$$

inequality (B.2) holds when $||y_t^\tau(\cdot|a^*, h^t) - y_t^\tau(\cdot|a, h^t)|| \leq \Delta$ for every $\tau \in \{1, 2, ..., K\}$. Let $\mathcal{H}^{(a^*,\sigma_2)}$ be the set of public histories that occur with positive probability when player 1 plays a^* in every period and player 2 plays σ_2. I partition $\mathcal{H}^{(a^*,\sigma_2)}$ into two subsets, $\mathcal{H}_0^{(a^*,\sigma_2)}$ and $\mathcal{H}_1^{(a^*,\sigma_2)}$:

1. If there exists $a \prec a^*$ such that $||y_t^\tau(\cdot|a^*, h^t) - y_t^\tau(\cdot|a', h^t)|| \leq \Delta$ for every τ, then $h^t \in \mathcal{H}_0^{(a^*,\sigma_2)}$.
2. If for every $a \prec a^*$, there exists τ such that $||y_t^\tau(\cdot|a^*, h^t) - y_t^\tau(\cdot|a', h^t)|| \geq \Delta$, then $h^t \in \mathcal{H}_1^{(a^*,\sigma_2)}$.

For every $h^t \in \mathcal{H}_0^{(a^*,\sigma_2)}$, the strategic type has a strict incentive not to play a^* at h^t, which means that player 2 assigns probability 1 to the commitment type after observing a^* at h^t. For every $\tau \in \{1, 2, ..., K\}$, every on-path history $h^{t+\tau} \succ h^t$ such that a^* has been played from period t to $t+\tau-1$, player 2 has a strict incentive to play b^* at $h^{t+\tau}$. This in addition to the fact that player 2 plays an action at least as large as b' at every on-path history implies that for every $h^t \in \mathcal{H}_0^{(a^*,\sigma_2)}$, we have:

$$\frac{1}{K+1}\mathbb{E}^{(a^*,\sigma_2)}\Big[\sum_{s=t}^{t+K} u_1(a_s, b_s)\Big|h^t\Big] \geq \frac{K}{K+1}u_1(a^*, b^*) + \frac{1}{K+1}u_1(a^*, b'). \tag{B.4}$$

For every $h^t \in \mathcal{H}_1^{(a^*,\sigma_2)}$, there exists a constant $\gamma > 0$ such that for every $\alpha \in \Delta(A)$ such that $b \prec b^*$ best replies against α, we have $||y_t(\cdot|a^*, h^t) - y_t(\cdot|\alpha, h^t)|| \geq \gamma\Delta$. The Pinsker's inequality implies that

$$d\Big(y_t(\cdot|\alpha,h^t)\Big\|y_t(\cdot|a^*,h^t)\Big) \geq 2\gamma^2\Delta^2. \tag{B.5}$$

for every such $\alpha \in \Delta(A)$. For every equilibrium (σ_1,σ_2) and every $\tau \in \{0,1,...,K\}$,

$$\mathbb{E}^{(a^*,\sigma_2)}\Big[\sum_{s=0}^{\infty} d\Big(y_{s(K+1)+\tau}(\cdot|\sigma_1(h^{s(K+1)+\tau}),h^{s(K+1)+\tau})\Big\|y_{s(K+1)+\tau}(\cdot|a^*,h^{s(K+1)+\tau})\Big)\Big] \leq -\log\pi_0. \tag{B.6}$$

Inequalities (B.5) and (B.6) together imply that:

$$\mathbb{E}^{(a^*,\sigma_2)}\Big[\sum_{s=0}^{\infty}\mathbf{1}\Big\{h^{s(K+1)+\tau} \in \mathcal{H}_1^{(a^*,\sigma_2)} \text{ and } \sigma_2(h^{s(K+1)+\tau}) \prec b^*\Big\}\Big] \leq -\frac{\log\pi_0}{2\gamma^2\Delta^2} \tag{B.7}$$

I derive a lower bound for $\liminf_{t\to\infty}\frac{1}{t}\mathbb{E}^{(a^*,\sigma_2)}\Big[\sum_{s=0}^{t-1}u_1(a_s,b_s)\Big]$ using inequalities (B.4) and (B.7). For every $\tau \in \{0,1,...,K\}$, let

$$\mathcal{H}_0^{\tau} \equiv \Big\{h^t\Big|\exists h^{s(K+1)+\tau} \in \mathcal{H}_0^{(a^*,\sigma_2)} \text{ such that } h^t \succeq h^{s(K+1)+\tau} \text{ and } t \in [s(K+1), s(K+1)+K]\Big\},$$

let

$$\mathcal{H}_1^{\tau} \equiv \Big\{h^{s(K+1)+\tau} \in \mathcal{H}_1^{(a^*,\sigma_2)}\Big|s \in \mathbb{N}\Big\},$$

and let $\mathcal{H}^{\tau} \equiv \mathcal{H}_0^{\tau} \cup \mathcal{H}_1^{\tau}$. By definition, $\mathcal{H}^{(a^*,\sigma_2)} = \bigcup_{\tau=0}^{K}\mathcal{H}^{\tau}$. An important observation is that for every $\tau,\tau' \in \{0,1,...,K\}$ with $\tau \neq \tau'$,

$$\mathcal{H}_1^{\tau} \cap \mathcal{H}_1^{\tau'} = \{\varnothing\} \text{ and } \mathcal{H}_0^{\tau} \cap \mathcal{H}_0^{\tau'} = \{\varnothing\}. \tag{B.8}$$

The former is straightforward. For the latter, suppose by way of contradiction that $h^t \in \mathcal{H}_0^{\tau} \cap \mathcal{H}_0^{\tau'}$ with $\tau < \tau'$, there exist h^s and $h^{s+\tau'-\tau}$ such that $h^t \succsim h^{s+\tau'-\tau} \succ h^s$, $h^s \in \mathcal{H}_0^{\tau}$, $t-s \leq K$, and $s-\tau$ is divisible by $K+1$. On one hand $h^s \in \mathcal{H}_0^{\tau}$ and $\tau'-\tau \leq K$ implies that $\sigma_1(h^{s+\tau'-\tau}) = a^*$. On the other hand $h^{s+1} \in \mathcal{H}_0^{\tau'}$ implies that $\sigma_1(h^{s+\tau'-\tau}) \neq a^*$. This leads to a contradiction.

For every $\tau \in \{0,1,...,K\}$, inequality (B.4) implies that player 1's expected average payoff at histories in $\mathcal{H}_0^{\tau}$ is at least the right-hand-side of (B.1). Since $\mathcal{H}_0^{\tau} \cap \mathcal{H}_0^{\tau'} = \{\varnothing\}$ for every $\tau \neq \tau'$, it implies that player 1's expected average payoff at histories in $\bigcup_{\tau=0}^{K}\mathcal{H}_0^{\tau}$ is at least the right-hand-side of (B.1). For every $\tau \in \{0,1,...,K\}$, (B.7) implies that player 1's expected average payoff at histories belonging to set $\mathcal{H}_1^{\tau}\Big\backslash\bigcup_{s=0}^{K}\mathcal{H}_0^{s}$ is at least $u_1(a^*,b^*)$. Since $\mathcal{H}_1^{\tau} \cap \mathcal{H}_1^{\tau'} = \{\varnothing\}$ for every

$\tau \neq \tau'$, it implies that player 1's expected average payoff at histories in $\bigcup_{s=0}^{K} \mathcal{H}_1^s \Big\backslash \bigcup_{s=0}^{K} \mathcal{H}_0^s$ is at least $u_1(a^*, b^*)$. The two parts imply that

$$\liminf_{t\to\infty} \frac{1}{t} \mathbb{E}^{(a^*,\sigma_2)} \Big[\sum_{s=0}^{t-1} u_1(a_s, b_s) \Big] \geq \frac{K}{K+1} u_1(a^*, b^*) + \frac{1}{K+1} u_1(a^*, b').$$

B.2 Tightness of the Lower Bound in Proposition 3

When payoffs are monotone-supermodular, (a', b') is the unique stage-game Nash equilibrium. Let $\overline{\pi}_0$ be the largest real number in $(0, 1)$ such that b' best replies against the mixed action $\overline{\pi}_0 \circ a^* + (1 - \overline{\pi}_0) \circ a'$. Consider the following construction when $\pi_0 \in (0, \overline{\pi}_0)$. At every on-path history (the set of on-path histories can be derived recursively),

- if t is divisible by $K + 1$, then player 1 plays a' and player 2 plays b' in period t;
- if t is not divisible by $K + 1$, then player 1 plays a^* and player 2 plays b^* in period t.

I partition off-path histories into three subsets. For every period t public history such that:

- (i) there exists no $r < t$, such that $b_r \neq b^*$ and r is not divisible by $K + 1$; (ii) there exists no $s < t$ such that $b_s \neq b'$ and s is divisible by $K + 1$; (iii) player 2 observes player 1 playing an off-path action in period $t - 1$, then players play (a^*, b^*) if t is divisible by $K + 1$, and play (a', b') if t is not divisible by $K + 1$.
- (i) there exists no $r < t$, such that $b_r \neq b^*$ and r is not divisible by $K+1$, but (ii) there exists $s < t$ such that $b_s \neq b'$ and s is divisible by $K + 1$. If $t - 1$ is divisible by $K + 1$, $b_{t-1} = b^*$ while $a_{t-1} \neq a^*$, then play (a', b') in period t. If $t - 1$ is divisible by $K + 1$, $b_{t-1} = b^*$ while $a_{t-1} = a^*$, then play (a^*, b^*) in period t if and only if $\xi_t > 1/2$ and play (a', b') in period t otherwise. If $t - 1$ is not divisible by $K + 1$, or $b_{t-1} \neq b^*$, then play (a^*, b^*) if t is not divisible by $K + 1$ and play (a', b') if t is divisible by $K + 1$.
- there exists $r < t$, such that $b_r \neq b^*$ and r is not divisible by $K + 1$, then play (a', b') in all subsequent periods.

Player 1's undiscounted time-average payoff from playing a^* in every period equals the right-hand-side of (B.1). I verify players' incentive constraints. Since b^* best replies to a^* and b' best replies to a', player 2's incentive constraints are satisfied. I verify player 1's incentives. At every on-path h^t,

- If $t+1$ not divisible by $K+1$ and t is not divisible by $K+1$, then the strategic type's continuation value from playing a^* in period t is at least

$$V \equiv \frac{u_1(a',b') + \delta u_1(a^*,b^*) + \delta^2 u_1(a^*,b^*) + ... + \delta^K u_1(a^*,b^*)}{1+\delta+...+\delta^K}, \tag{B.9}$$

  while his continuation value from playing any other action is $u_1(a',b')$. This verifies his incentive to play a^* when δ is above some cutoff.

- If $t+1$ not divisible by $K+1$ and t is divisible by $K+1$, then the strategic type's continuation values from playing a^* and a' are the same, equal V, while his continuation value from playing other actions is $u_1(a',b')$. He has a strict incentive to play a' since a' best replies to b'.

- If $t+1$ is divisible by $K+1$, then the strategic type's continuation value from playing a^* in period t is at least V. If he deviates and plays a_t, then consider his incentive in period $t+1$ at off-path history $(h^t, a_t, b_t = b^*)$.

  Since player 2 plays b^* in period $t+1$ after observing player 1's deviation in period t, player 1's continuation value from playing a^* in period $t+1$ is at least $\frac{1}{2}V + \frac{1}{2}u_1(a',b')$. This is because player 2 will play b^* with probability $1/2$ in period $t+2$, after which player 1 will be forgiven for his deviation. Player 1's continuation value from playing actions other than a^* in period $t+1$ is $u_1(a',b')$. Therefore, he has a strict incentive to play a^* in period $t+1$ following his deviation in period t, and his continuation value in period t when he deviates is strictly lower than V.

B.3 Asymptotic Payoff under Finite K and M

I examine player 1's undiscounted average payoff when he imitates the commitment type given that every player 2 can only observe player 1's actions in the last K periods and player 2's actions in the last M periods, where both K and M are finite. In contrast to the conclusion of Proposition 3, I show that under a mild regularity condition on players' stage-game payoffs (that is satisfied by the product choice game and many other examples), there exist equilibria in which player 1's undiscounted average payoff from imitating the commitment type equals his minmax payoff $u_1(a',b')$.

First, I introduce the regularity condition. Since players' payoffs are monotone-supermodular, $u_1(a^*,b^*) > u_1(a',b')$. Without loss of generality, I normalize player 1's payoff so that $u_1(a',b') = 0$

and $u_1(a^*, b^*) = 1$. Let $\underline{q}$ be the largest $q \in [0, 1]$ such that b' is not player 2's strict best reply to $qa^* + (1 - q)a'$. Let $\overline{q}$ be the smallest $q \in [0, 1]$ such that b^* is not player 2's strict best reply to $qa^* + (1 - q)a'$. Since b' is a strict best reply to a' and b^* is a strict best reply to a^*, there exist $b^{**} \neq b'$ and $b'' \neq b^*$ such that $\{b^{**}, b'\} \subset \mathrm{BR}_2(\underline{q}a^* + (1 - \underline{q})a')$ and $\{b^*, b''\} \subset \mathrm{BR}_2(\overline{q}a^* + (1 - \overline{q})a')$. Assumption 2 in the main text implies that either $b^* = b^{**}$ and $b' = b''$, or $b^* \succ b'' \succ b^{**} \succ b'$, or $b^* \succ b'' = b^{**} \succ b'$.

Definition. *Players' stage-game payoffs are irregular if $b'' = b^{**}$ and $u_1(a^*, b^{**}) < -1$. Players' stage-game payoffs are regular if they are not irregular.*

One can verify that players' payoffs are regular in the product choice game since $b^* = b^{**} \succ b'' = b'$. Players' payoffs are also regular in general monotone-supermodular games where player 1's cost of playing a^* is not too large when player 2 plays actions in between b' and b^*. An example of a game where players' payoffs are irregular is given by:

$-$	b^*	b''	b'
a^*	$1, 1$	$-2, 0$	$-3, -2$
a'	$2, -2$	$1, 0$	$0, 1$

Proposition. *Suppose players' payoffs are monotone-supermodular and regular. For every $\varepsilon > 0$, there exists $\underline{\delta} \in (0, 1)$ such that when $\delta > \underline{\delta}$, there exist equilibria in which*

$$\limsup_{t \to \infty} \frac{1}{t} \mathbb{E}^{(a^*, \sigma_2)} \Big[\sum_{s=0}^{t-1} u_1(a_s, b_s) \Big] \leq \varepsilon. \tag{B.10}$$

My proof of this proposition considers three cases separately.

Case 1: $\mathbf{u_1(a^*, b^{**}) > 0}$ Consider the following strategy profile in which player 2_t's action depends only on (a_{t-1}, b_{t-1}), and the rational type player 1 mixes between a^* and a' such that player 2_t is indifferent between b^{**} and b'. Later on, I will verify that such mixing probabilities exist.

1. When $(a_{t-1}, b_{t-1}) \notin \{(a^*, b'), (a^*, b^{**}), (a', b^{**})\}$, player 2_t plays b^{**} with probability $r(a', b')$.

2. When $(a_{t-1}, b_{t-1}) = (a^*, b')$, player 2_t plays b^{**} with probability $r(a^*, b')$.

3. When $(a_{t-1}, b_{t-1}) = (a', b^{**})$, player 2_t plays b^{**} with probability $r(a', b^{**})$.

4. When $(a_{t-1}, b_{t-1}) = (a^*, b^{**})$, player 2_t plays b^{**} with probability $r(a^*, b^{**})$.

Let

$$X \equiv \max\Big\{-\frac{1-\delta}{\delta}u_1(a^*,b'),\ (1-\delta)u_1(a',b^{**})\Big\}. \tag{B.11}$$

Player 1's continuation values are $V(a',b')=0$, $V(a^*,b')=-\frac{1-\delta}{\delta}u_1(a^*,b')$, and

$$V(a,b^{**})=\frac{X-(1-\delta)u_1(a,b^{**})}{\delta} \text{ for every } a\in\{a',a^*\}. \tag{B.12}$$

Let $r(a',b')=0$. For every $(a,b)\in\{(a^*,b'),(a^*,b^{**}),(a',b^{**})\}$, let

$$r(a,b)=\frac{V(a,b)}{X}. \tag{B.13}$$

I verify that $V(a,b)\leq X$ so that $r(a,b)$ is a well-defined probability. First, $V(a^*,b')\leq X$ by definition. Second, I show that $V(a',b^{**})<V(a^*,b^{**})\leq X$. The first inequality is because $u_1(a',b^{**})>u_1(a^*,b^{**})$ and (B.12). The second inequality is equivalent to

$$\frac{X-(1-\delta)u_1(H,T)}{\delta}\leq X \Leftrightarrow X<u_1(a^*,b^{**}).$$

The last inequality is satisfied when δ is close to 1 since X converges to 0 and $u_1(a^*,b^{**})>0$.

According to the construction of these continuation values, we have

$$(1-\delta)u_1(a^*,b^{**})+\delta V(a^*,b^{**})=(1-\delta)u_1(a',b^{**})+\delta V(a',b^{**}),$$

and

$$(1-\delta)u_1(a^*,b^{'})+\delta V(a^*,b^{'})=(1-\delta)u_1(a',b^{'})+\delta V(a',b^{'}),$$

which means that player 1 is indifferent regardless of player 2's action, and therefore, he is indifferent between a^* and a' at every (a_{t-1},b_{t-1}). Since a' is the lowest action and player 1's continuation value at (a,b) is the same as his lowest continuation value $V(a',b')$ for every b and $a\notin\{a^*,a'\}$, player 1 strictly prefers a' to actions other than a' and a^* at every history.

Then I verify player 1's mixed strategies is well-defined by showing that $\pi_t\leq q^*/2$ at every history. Let

$$L\equiv\min\Big\{\frac{r(a^*,b')}{r(a^*,b^{**})},\frac{r(a^*,b^{**})}{r(a^*,b')}\Big\}. \tag{B.14}$$

According to the expressions for $r(a^*, b^{**})$ and $r(a^*, b')$, we have

$$\frac{r(a^*, b')}{r(a^*, b^{**})} = \frac{-u_1(a^*, b')}{\max\Big\{-\frac{u_1(a^*, b')}{\delta}, u_1(a', b^{**})\Big\} - u_1(a^*, b^{**})}.$$

Both the denominator and the numerator of the above expression are bounded away from 0 and are bounded from above for δ close to 1, which implies that L is bounded away from 0. Let $\overline{\pi}_0$ be defined via the following equation:

$$\frac{\overline{\pi}_0}{1-\overline{\pi}_0} = \frac{q^*/2}{1-q^*/2}\Big(\frac{q^*}{2}\Big)^{K+M+1} L^M. \tag{B.15}$$

I show by induction that $\pi_t \leq q^*/2$ for every $t \in \mathbb{N}$ if $\pi_0 < \overline{\pi}_0$. Without loss of generality, we only need to consider histories where $(a_{t-K}, ..., a_{t-1}) = (a^*, ..., a^*)$ and $(b_{t-M}, ..., b_{t-1}) = (b^{**}, ..., b^{**})$. First, condition (B.15) implies that $\pi_0 \leq q^*/2$. Second, suppose $\pi_s \leq q^*/2$ for every $s \leq t-1$, then the rational type plays a^* with probability at least $q^*/2$ at every history from period 0 to period $t-1$. Let P^c be the probability measure over histories induced by the commitment type and let P^r be the probability measure over histories induced by the strategic type. According to Bayes rule, we have

$$\frac{\pi_t}{1-\pi_t} = \frac{\pi_0}{1-\pi_0}\Pi_{i=1}^K \frac{P^c(a_{t-i} = a^*|a_{t-i+1}, ..., a_{t-1})}{P^r(a_{t-i} = a^*|a_{t-i+1}, ..., a_{t-1})}\Pi_{i=1}^M \frac{P^c(b_{t-i} = b^{**}|a_{t-i+1}, ..., a_{t-1}, b_{t-i+1}, ..., b_{t-1})}{P^r(b_{t-i} = b^{**}|a_{t-i+1}, ..., a_{t-1}, b_{t-i+1}, ..., b_{t-1})}.$$

According to the induction hypothesis, we have

$$\frac{P^c(a_{t-i} = a^*|a_{t-i+1}, ..., a_{t-1})}{P^r(a_{t-i} = a^*|a_{t-i+1}, ..., a_{t-1})} \leq (q^*/2)^{-1}. \tag{B.16}$$

Since the rational type player 1 plays a^* with probability at least $q^*/2$ in every period before t and conditional on playing a^*, the probability of b^{**} is at least $\min\{r(a^*, b'), r(a^*, b^{**})\}$. Therefore,

$$\frac{P^c(b_{t-i} = b^{**}|a_{t-i+1}, ..., a_{t-1}, b_{t-i+1}, ..., b_{t-1})}{P^r(b_{t-i} = b^{**}|a_{t-i+1}, ..., a_{t-1}, b_{t-i+1}, ..., b_{t-1})} \leq (q^*/2)^{-1}L^{-1}. \tag{B.17}$$

Plugging inequalities (B.16) and (B.17) into the expression for π_t, we have $\pi_t \leq q^*/2$.

In the last step, I compute player 1's undiscounted time average payoff by playing a^* in every period, which induces a 2-state Markov Chain with transition probabilities $\Pr(b^{**}|b^{**}) = r(a^*, b^{**})$ and $\Pr(b'|b') = r(a^*, b')$. The stationary distribution attaches probability $\frac{r(a^*,b')}{1-r(a^*,b^{**})+r(a^*,b')}$ to state

b^{**}. Player 1's undiscounted average payoff from playing a^* in every period is

$$\frac{r(a^*,b')}{1-r(a^*,b^{**})+r(a^*,b')}u_1(a^*,b^{**})+\frac{1-r(a^*,b^{**})}{1-r(a^*,b^{**})+r(a^*,b')}u_1(a^*,b'). \tag{B.18}$$

Plugging in the expressions for $r(a^*,b')$ and $r(a^*,b^{**})$ and using the observation that $X \to 0$ as $\delta \to 1$, we obtain that the above equation is close to 0 as $\delta \to 1$.

Case 2: $\mathbf{u_1(a^*,b^{**}) \leq 0}$ and $\mathbf{b^* \succ b'' \succ b^{**} \succ b'}$ Consider the following strategy profile:

1. When $(a_{t-1},b_{t-1}) \notin \{(a^*,b'),(a^*,b^{**}),(a^*,b''),(a^*,b^*),(a',b^*),(a',b^{**})\}$, player 2_t plays b' and the rational type player 1 mixes between a^* and a' with probabilities such that the unconditional probability of a^* is $\underline{q}$.

2. When $(a_{t-1},b_{t-1})=(a^*,b')$, player 2_t plays b^{**} with probability $r(a^*,b')$ and plays b' with complementary probability. The rational type player 1 mixes between a^* and a' with probabilities such that the unconditional probability of a^* is $\underline{q}$.

3. When $(a_{t-1},b_{t-1})=(a',b^{**})$, player 2_t plays b^{**} with probability $r(a',b^{**})$ and plays b' with complementary probability. The rational type player 1 mixes between a^* and a' with probabilities such that the unconditional probability of a^* is $\underline{q}$.

4. When $(a_{t-1},b_{t-1})=(a^*,b^{**})$, player 2_t plays b^* with probability $r(a^*,b^{**})$ and plays b'' with complementary probability. The rational type player 1 mixes between a^* and a' with probabilities such that the unconditional probability of a^* is $\overline{q}$.

5. When $(a_{t-1},b_{t-1})=(a^*,b'')$, player 2_t plays b^{**} with probability $r(a^*,b'')$ and plays b' with complementary probability. The rational type player 1 mixes between a^* and a' with probabilities such that the unconditional probability of a^* is $\underline{q}$.

6. When $(a_{t-1},b_{t-1})=(a',b^*)$, player 2_t plays b^* with probability $r(a',b^*)$ and plays b'' with complementary probability. The rational type player 1 mixes between a^* and a' with probabilities such that the unconditional probability of a^* is $\overline{q}$.

7. When $(a_{t-1},b_{t-1})=(a^*,b^*)$, player 2_t plays b^* with probability $r(a^*,b^*)$ and plays b'' with complementary probability. The rational type player 1 mixes between a^* and a' with probabilities such that the unconditional probability of a^* is $\overline{q}$.

The rational type player 1's continuation value satisfies $V(a', b') = 0$, $V(a', b'') = 0$,

$$0 = (1-\delta)u_1(a^*, b') + \delta V(a^*, b') = (1-\delta)u_1(a', b') + \delta V(a', b'), \tag{B.19}$$

$$X(b^{**}) \equiv (1-\delta)u_1(a^*, b^{**}) + \delta V(a^*, b^{**}) = (1-\delta)u_1(a', b^{**}) + \delta V(a', b^{**}), \tag{B.20}$$

$$Y(b'') \equiv (1-\delta)u_1(a^*, b'') + \delta V(a^*, b'') = (1-\delta)u_1(a', b^{''}) + \delta V(a', b''), \tag{B.21}$$

and

$$Y(b^*) \equiv (1-\delta)u_1(a^*, b^*) + \delta V(a^*, b^*) = (1-\delta)u_1(a', b^*) + \delta V(a', b^*), \tag{B.22}$$

where $Y(b'') \equiv (1-\delta)u_1(a', b'')$,

$$X(b^{**}) \equiv \max\Big\{(1-\delta)u_1(a', b''), (1-\delta)u_1(a', b^{**}), -\frac{1-\delta}{\delta}u_1(a^*, b')\Big\},$$

and

$$Y(b^*) \equiv 2\max\Big\{\frac{X(b^{**}) - (1-\delta)u_1(a^*, b^{**})}{\delta}, (1-\delta)(\delta u_1(a', b'') + u_1(a^*, b^*)), (1-\delta)u_1(a', b^*) + \delta X(b^*)\Big\}.$$

In order to deliver these continuation values, we need

$$Y(b'') \le V(a^*, b^{**}) = \frac{X(b^{**}) - (1-\delta)u_1(a^*, b^{**})}{\delta} \le Y(b^*), \tag{B.23}$$

$$0 \le V(a^*, b'') = \frac{Y(b'') - (1-\delta)u_1(a^*, b'')}{\delta} \le X(b^{**}), \tag{B.24}$$

$$Y(b'') \le V(a^*, b^*) = \frac{Y(b^*) - (1-\delta)u_1(a^*, b^*)}{\delta} \le Y(b^*), \tag{B.25}$$

$$Y(b'') \le V(a', b^*) = \frac{Y(b^*) - (1-\delta)u_1(a', b^*)}{\delta} \le Y(b^*), \tag{B.26}$$

and

$$0 \le V(a', b^{**}) = \frac{X(b^{**}) - (1-\delta)u_1(a', b^{**})}{\delta} \le X(b^{**}). \tag{B.27}$$

All of these conditions are satisfied when δ is close to 1 given the values of $X(b^{**})$, $Y(b'')$, and $Y(b^{**})$. As a result, there exist $r(a^*, b')$, $r(a^*, b^{**})$, $r(a^*, b'')$, $r(a^*, b^*)$, $r(a', b^*)$, and $r(a', b^{**})$ that deliver player 1 continuation values $V(a^*, b')$, $V(a^*, b^{**})$, $V(a^*, b'')$, $V(a^*, b^*)$, $V(a', b^*)$, and $V(a', b^{**})$. Furthermore, the definition of $Y(b^*)$ implies that $r(a^*, b^*), r(a^*, b^{**}), r(a', b^*)$ are less than 1/2.

Let

$$L \equiv \min\left\{\frac{\min\{r(a^*,b^*),r(a^*,b^{**}),r(a',b^*)\}}{\max\{r(a^*,b^*),r(a^*,b^{**}),r(a',b^*)\}},\frac{\min\{r(a^*,b''),r(a',b^*),r(a',b^{**})\}}{\max\{r(a^*,b''),r(a',b^*),r(a',b^{**})\}}\right\}.$$

By definition, L is bounded away from 0. Let $\overline{\pi}_0 \in (0,1)$ be defined via the following equation:

$$\frac{\overline{\pi}_0}{1-\overline{\pi}_0} = \frac{\underline{q}/2}{1-\underline{q}/2}\Big(\frac{\underline{q}}{2}\Big)^{K+M+1} L^M. \tag{B.28}$$

The same argument as before implies that player 2's posterior belief attaches probability less than $\underline{q}/2$ to the commitment type if her prior belief satisfies $\pi_0 \leq \overline{\pi}_0$.

When player 1 plays a^* in every period, he induces a Markov Chain with four states b^*, b^{**}, b'', and b', which is communicating. Since his discounted average payoff is 0, his undiscounted average payoff is close to 0 when δ is close to 1.

Case 3: $\mathbf{u_1(a^*,b^{**}) \leq 0,\ b^* \succ b'' = b^{**} \succ b',}$ and $\mathbf{u_1(a^*,b^{**}) \in (-1,0]}$ Consider the following strategy profile:

1. When $(a_{t-1},b_{t-1}) \notin \{(a^*,b'),(a^*,b^{**}),(a^*,b^*),(a',b^*),(a',b^{**})\}$, player 2_t plays b' and the rational type player 1 mixes between a^* and a' with probabilities such that the unconditional probability of a^* is $\underline{q}$.

2. When $(a_{t-1},b_{t-1}) = (a^*,b')$, player 2_t plays b^{**} with probability $r(a^*,b')$ and plays b' with complementary probability. The rational type player 1 mixes between a^* and a' with probabilities such that the unconditional probability of a^* is $\underline{q}$.

3. When $(a_{t-1},b_{t-1}) = (a^*,b^{**})$, player 2_t plays b^{**} with probability $r(a^*,b^{**})$ and plays b' with complementary probability. The rational type player 1 mixes between a^* and a' with probabilities such that the unconditional probability of a^* is $\underline{q}$.

4. When $(a_{t-1},b_{t-1}) = (a',b^{**})$, player 2_t plays b^* with probability $r(a',b^{**})$ and plays b^{**} with complementary probability. The rational type player 1 mixes between a^* and a' with probabilities such that the unconditional probability of a^* is $\overline{q}$.

5. When $(a_{t-1},b_{t-1}) = (a',b^*)$, player 2_t plays b^* with probability $r(a',b^*)$ and plays b^{**} with complementary probability. The rational type player 1 mixes between a^* and a' with probabilities such that the unconditional probability of a^* is $\overline{q}$.

6. When $(a_{t-1}, b_{t-1}) = (a^*, b^*)$, player 2_t plays b^* with probability $r(a^*, b^*)$ and plays b^{**} with complementary probability. The rational type player 1 mixes between a^* and a' with probabilities such that the unconditional probability of a^* is $\overline{q}$.

Let

$$X \equiv \max\Big\{(1-\delta)u_1(a', b^{**}), -\frac{1-\delta}{\delta}u_1(a^*, b')\Big\}. \tag{B.29}$$

Let $Y \in \mathbb{R}$ be a real number satisfying

$$\frac{X - (1-\delta)u_1(a^*, b^{**})}{\delta} < Y < (1-\delta)u_1(a^*, b^*) + \delta X. \tag{B.30}$$

Such Y exists if and only if $(1+\delta)X < u_1(a^*, b^{**}) + \delta u_1(a^*, b^*)$. When δ is close enough to 1, this is satisfied when $u_1(a^*, b^*) + u_1(a^*, b^{**}) > 0$, i.e., when payoffs are regular.

Player 1's continuation values are $V(a', b') = 0$, $V(a^*, b') = -\frac{1-\delta}{\delta}u_1(a^*, b')$, $V(a^*, b^{**})$ and $V(a', b^{**})$ are pinned down by

$$X = (1-\delta)u_1(a', b^{**}) + \delta V(a', b^{**}) = (1-\delta)u_1(a^*, b^{**}) + \delta V(a^*, b^{**}), \tag{B.31}$$

and $V(a', b^*)$ and $V(a^*, b^*)$ are pinned down by

$$Y = (1-\delta)u_1(a', b^*) + \delta V(a', b^*) = (1-\delta)u_1(a^*, b^*) + \delta V(a^*, b^*). \tag{B.32}$$

According to the construction of X and Y, we know that $V(a^*, b') \in (0, X)$, $V(a', b^{**}) \in (0, X)$, $V(a^*, b^{**}) \in (X, Y)$, $V(a', b^*) \in (0, X)$ and $V(a^*, b^*) \in (0, X)$. The strategy of playing a^* in every period induces a Markov Chain with three states b', b^*, and b^{**}, that is communicating. Since player 1's discounted average payoff from playing a^* in every period is 0, his undiscounted average payoff is 0 when δ is close to 1.

C Proofs of Theorem 2 and Theorem 3

I establish Theorem 2 in Section C.1. In Section C.2, I establish the existence of equilibrium when the private signal is unboundedly informative, $M = +\infty$, and δ being close enough to 1. I establish Theorem 3 in Section C.3.

C.1 Proof of Theorem 2

I start from Lemma C.1 which shows that in every equilibrium, if player 1 plays a^* in every period, then there exists $\eta > 0$ that depends only on the distribution over private signals and the prior probability of commitment type π_0, such that the probability with which player 2 plays b^* with probability at least η in every period is close to 1.

Lemma C.1. *Suppose the private signal is unboundedly informative about a^*. For every $\pi_0 > 0$ and $\varepsilon > 0$, there exists $\eta > 0$, such that in every equilibrium (σ_1, σ_2),*

$$\Pr\left\{ \Pr(b_t = b^*) \geq \eta \text{ for every } t \in \mathbb{N} \middle| (a^*, \sigma_2) \right\} \geq 1 - \varepsilon. \tag{C.1}$$

Proof. Let $p^* \in (0,1)$ be such that player 2 has a strict incentive to play b^* when she believes that player 1 plays a^* with probability more than p^*. For every $\pi > 0$, there exists $M(\pi) > 0$ such that when the prior belief assigns probability more than π to a^* and the signal realization s is such that $f(s|a^*) > M(\pi) f(s|a)$ for every $a \neq a^*$, the posterior belief after observing s assigns probability more than p^* to a^*. Let $l_0 \equiv \frac{1-\pi_0}{\pi_0}$, $l^* \equiv l_0/\varepsilon$, $\pi^* \equiv \frac{1}{l^*+1}$, let $S(\pi^*) \subset S$ be the set of signal realizations such that $f(s|a^*) > M(\pi^*) f(s|a)$ for every $a \neq a^*$, and let $\eta \equiv \sum_{s \in S(\pi^*)} f(s|a^*)$. Since the private signal is unboundedly informative, $S(\pi^*)$ is non-empty and $f(s|a^*) > 0$ for every $s \in S(\pi^*)$. Therefore, $\eta > 0$.

Let π_t be the probability of commitment type after player 2_t observes $\{b_0, ..., b_{t-1}\}$, but not s_t and $\{a_{\max\{0,t-K\}}, ..., a_{t-1}\}$. Let $\widetilde{\pi}_t$ be the probability of commitment type after player 2_t observes $\{b_0, ..., b_{t-1}\}$ and $\{a_{\max\{0,t-K\}}, ..., a_{t-1}\}$, but not s_t. By definition, if $\{a_{\max\{0,t-K\}}, ..., a_{t-1}\} = \{a^*, ..., a^*\}$, then $\widetilde{\pi}_t \geq \pi_t$. Under the probability measure induced by (a^*, σ_2), $\{\frac{1-\pi_t}{\pi_t}\}_{t \in \mathbb{N}}$ is a non-negative supermartingale. The Doob's Upcrossing Inequality implies that when the prior belief is π_0, the probability of the event $\{\pi_t \geq \pi^* \text{ for all } t \in \mathbb{N}\}$ is at least $1 - \varepsilon$. Since player 2_t has a strict incentive to play b^* after she observes $s_t \in S(\widetilde{\pi}_t)$, and moreover $\widetilde{\pi}_t \geq \pi_t$, we have $S(\pi^*) \subset S(\widetilde{\pi}_t)$ when $\pi_t \geq \pi^*$. The probability of event $\{\Pr(b_t = b^*) \geq \eta \text{ for every } t \in \mathbb{N}\}$ is at least $1 - \varepsilon$. □

Next, I show that in every period where the probability of commitment type is more than π^* but player 2 plays b^* with ex ante probability less than $1 - \nu$, one can bound the informativeness of b_t about player 1's type from below by a strictly positive function of ν.

Lemma C.2. *Suppose the private signal is unboundedly informative about a^*, and satisfies MLRP. For every $\pi^* \in (0,1)$, there exists $c > 0$ such that for every $\nu \in (0,1)$, $\alpha \in \Delta(A)$*

with $\alpha(a^*) > \pi^*$, *and* $\beta : S \to \Delta(B)$ *that best replies to* α. *If* $\gamma(a^*,\beta)[b^*] < 1-\nu$, *then* $d\big(\gamma(\alpha,\beta)\big\|\gamma(a^*,\beta)\big) > 2c\nu^2$.

Proof. Since $u_2(a,b)$ has strictly increasing differences and the distribution over private signals satisfies MLRP, Topkis Theorem implies that every β that best replies to some α must be monotone, i.e., for every $s \succ s'$ and $b \in B$, if $\beta(s)$ assigns a positive probability to b, then $\beta(s')$ assigns zero probability to every b' smaller than b. Therefore, it is without loss of generality to focus on player 2's pure strategies taking the form of $\beta : S \to B$.

When $\pi_t > \pi^*$, player 2_t has a strict incentive to play b^* after observing $s \in S(\pi^*)$, where $S(\pi^*)$ is the set of signal realizations such that $f(s|a^*) > f(s|a)M(\pi^*)$ for every $a \neq a^*$. At every history h^t, there exists an interval $[\underline{s},\overline{s}] \subset S$ such that $\beta(s) = b^*$ if and only if $s \in [\underline{s},\overline{s}]$, and moreover, $\beta(s) \succ b^*$ for every $s \succ \overline{s}$, and $\beta(s) \prec b^*$ for every $s \prec \underline{s}$. By definition, $S(\pi^*) \subset [\underline{s},\overline{s}]$. Let $S^* \equiv [\underline{s}^*,\overline{s}^*]$ be a non-empty interval that is a subset of $S(\pi^*)$. Since the signal distribution satisfies MLRP, we know that $f(s|a^*) > f(s|a)M(\pi^*)$ for every $s \preceq \overline{s}^*$ and $a \succ a^*$, and $f(s|a^*) > f(s|a)M(\pi^*)$ for every $s \succeq \underline{s}^*$ and $a \prec a^*$.

Let $\overline{A}$ be the set of actions that are strictly higher than a^* and let $\underline{A}$ be the set of actions that are strictly lower than a^*. For every $\alpha \in \Delta(A)$, let $\alpha' \in \Delta(A)$ be the distribution over A conditional on $a \neq a^*$. If $\text{supp}(\alpha) \cap \overline{A} \neq \{\varnothing\}$, then let $\overline{\alpha} \in \Delta(A)$ be the distribution over A conditional on $a \in \text{supp}(\alpha) \cap \overline{A}$. If $\text{supp}(\alpha) \cap \underline{A} \neq \{\varnothing\}$, then let $\underline{\alpha} \in \Delta(A)$ be the distribution over A conditional on $a \in \text{supp}(\alpha) \cap \underline{A}$. By definition, there exists $\lambda \in [0,1]$ such that $\alpha' = \lambda\overline{\alpha} + (1-\lambda)\underline{\alpha}$.

Suppose $\gamma(a^*,\beta)[b^*] < 1$ and $||\gamma(\alpha',\beta) - \gamma(a^*,\beta)|| = D$, then

$$\sum_{s \succ \overline{s}} f(s|a^*) \geq -D + \lambda \sum_{s \succ \overline{s}} f(s|\overline{\alpha}), \quad \sum_{s \prec \underline{s}} f(s|a^*) \geq -D + (1-\lambda) \sum_{s \prec \underline{s}} f(s|\underline{\alpha}), \tag{C.2}$$

and

$$-D + \sum_{s \in [\underline{s},\overline{s}] \backslash S^*} f(s|a^*) + \sum_{s \in S^*} f(s|a^*)$$
$$\leq \lambda \sum_{s \in S^*} f(s|\overline{\alpha}) + (1-\lambda) \sum_{s \in S^*} f(s|\underline{\alpha}) + \lambda \sum_{s \in [\underline{s},\overline{s}] \backslash S^*} f(s|\overline{\alpha}) + (1-\lambda) \sum_{s \in [\underline{s},\overline{s}] \backslash S^*} f(s|\underline{\alpha}).$$

Let $\eta \equiv \sum_{s \in S^*} f(s|a^*)$. Since $f(s|a^*) > f(s|a)M(\pi^*)$ for every $s \in S^*$ and $a \neq a^*$,

$$-D + \eta\Big(1 - \frac{1}{M(\pi^*)}\Big) + \sum_{s \in [\underline{s},\underline{s}^*)} f(s|a^*) + \sum_{s \in (\overline{s}^*,\overline{s}]} f(s|a^*) \leq \lambda \sum_{s \in [\underline{s},\overline{s}] \backslash S^*} f(s|\overline{\alpha}) + (1-\lambda) \sum_{s \in [\underline{s},\overline{s}] \backslash S^*} f(s|\underline{\alpha}). \tag{C.3}$$

Since the distribution over private signals satisfies MLRP,

$$\frac{\sum_{s\succ\overline{s}} f(s|a^*)}{\sum_{s\succ\overline{s}} f(s|\overline{\alpha})} \leq \frac{\sum_{s\in(\overline{s}^*,\overline{s}]} f(s|a^*)}{\sum_{s\in(\overline{s}^*,\overline{s}]} f(s|\overline{\alpha})} \quad \text{and} \quad \frac{\sum_{s\prec\underline{s}} f(s|a^*)}{\sum_{s\prec\underline{s}} f(s|\underline{\alpha})} \leq \frac{\sum_{s\in[\underline{s}^*,\underline{s})} f(s|a^*)}{\sum_{s\in[\underline{s}^*,\underline{s})} f(s|\underline{\alpha})}.$$

The above inequalities together with (C.2) imply that

$$\sum_{s\in(\overline{s}^*,\overline{s}]} f(s|a^*) \geq \frac{\sum_{s\in(\overline{s}^*,\overline{s}]} f(s|\overline{\alpha}) \sum_{s\succ\overline{s}} f(s|a^*)}{\sum_{s\succ\overline{s}} f(s|\overline{\alpha})} \geq \lambda \frac{\sum_{s\succ\overline{s}} f(s|a^*)}{D+\sum_{s\succ\overline{s}} f(s|a^*)} \sum_{s\in(\overline{s}^*,\overline{s}]} f(s|\overline{\alpha}) \tag{C.4}$$

and

$$\sum_{s\in[\underline{s},\underline{s}^*)} f(s|a^*) \geq (1-\lambda) \frac{\sum_{s\prec\underline{s}} f(s|a^*)}{D+\sum_{s\prec\underline{s}} f(s|a^*)} \sum_{s\in[\underline{s},\underline{s}^*)} f(s|\underline{\alpha}) \tag{C.5}$$

Plugging (C.4) and (C.5) back to (C.3), we obtain

$$\eta(1-\frac{1}{M(\pi^*)}) - \lambda \sum_{s\in[\underline{s},\underline{s}^*)} f(s|\overline{\alpha}) - (1-\lambda) \sum_{s\in(\overline{s}^*,\overline{s}]} f(s|\underline{\alpha})$$

$$\leq D\Big\{1 + \frac{\lambda}{D+\sum_{s\succ\overline{s}} f(s|a^*)} + \frac{1-\lambda}{D+\sum_{s\prec\underline{s}} f(s|a^*)}\Big\}. \tag{C.6}$$

First, I show that the left-hand-side of (C.6) is greater than $\eta/2$ when M is large enough. Without loss of generality, I index the elements of S as $\{..., s_{-1}, s_0, s_1, ...\}$ such that $s_i \prec s_j$ for every $i < j$. Consider three cases, depending on the limit of set S^* as $M \to +\infty$.

1. If there exist $m, n \in \mathbb{N}$ such that $\lim_{M\to+\infty} S^* = [s_m, s_n]$, then there exists $k \in \mathbb{N}$ such that $s_k \in S^*$ for every $M \in \mathbb{R}_+$. As a result, η is bounded from below by $f(s_k|a^*)$ for every M, which implies that the left-hand-side of (C.6) is more than $\eta/2$ when M is large enough.

2. If the limit of S^* is unbounded from above, then $f(s|a^*) \geq f(s|a)M$ for every $a \succ a^*$ and $s \in S$, which leads to a contradiction unless $\overline{A}$ is empty. Therefore, $\lambda = 0$ and $(\overline{s}^*, \overline{s}]$ is an empty set, and the left-hand-side of (C.6) is $\eta(1-\frac{1}{M(\pi^*)})$, which is greater than $\eta/2$ when $M(\pi^*)$ is large.

3. If the limit of S^* is unbounded from below, then similarly, the left-hand-side of (C.6) is η.

Next, I bound the term $1 + \frac{\lambda}{D+\sum_{s\succ\overline{s}} f(s|a^*)} + \frac{1-\lambda}{D+\sum_{s\prec\underline{s}} f(s|a^*)}$ from above. Since $\{b^*\} = \text{BR}_2(a^*)$, we know that for every $b \succ b^*$, there exists $\overline{r}^* \in \mathbb{R}_+$ such that $b \in \text{BR}_2(\alpha)$ only if $\alpha(\overline{A})/\alpha(a^*) \geq \overline{r}^*$, and for every $b \prec b^*$, there exists $\underline{r}^* \in \mathbb{R}_+$ such that $b \in \text{BR}_2(\alpha)$ only if $\alpha(\underline{A})/\alpha(a^*) \geq \underline{r}^*$. When

$\alpha(a^*) \geq \pi^*$, Bayes rule implies that

$$\frac{\lambda(1-\pi^*)\sum_{s\succ\overline{s}} f(s|\overline{\alpha})}{\pi^*\sum_{s\succ\overline{s}} f(s|a^*)} \geq \overline{r}^* \text{ and } \frac{(1-\lambda)(1-\pi^*)\sum_{s\prec\underline{s}} f(s|\underline{\alpha})}{\pi^*\sum_{s\prec\underline{s}} f(s|a^*)} \geq \underline{r}^*.$$

As a result,

$$1+\frac{\lambda}{D+\sum_{s\succ\overline{s}} f(s|a^*)}+\frac{1-\lambda}{D+\sum_{s\prec\underline{s}} f(s|a^*)} \leq 1+\frac{\pi^*}{1-\pi^*}(\overline{r}^*+\underline{r}^*).$$

Let $R \equiv 1+\frac{\pi^*}{1-\pi^*}(\overline{r}^*+\underline{r}^*)$. Inequality (C.6) then implies that $||\gamma(\alpha',\beta)-\gamma(a^*,\beta)|| = D \geq \frac{\eta}{2R}$. Since $\gamma(a^*,\beta)[b^*] < 1-\nu$, then there exists $c>0$ such that $\alpha(a^*) \leq 1-c\nu$, and therefore,

$$||\gamma(\alpha,\beta)-\gamma(a^*,\beta)|| \geq c\nu||\gamma(\alpha',\beta)-\gamma(a^*,\beta)|| \geq c\nu\frac{\eta}{2R}.$$

The Pinsker's inequality leads to a lower bound on the KL-divergence between $\gamma(\alpha,\beta)$ and $\gamma(a^*,\beta)$.

□

Let $h^t \equiv \{b_0,...,b_{t-1},a_{\max\{0,t-K\}},...,a_{t-1},\xi_t\}$ be player 2_t's information before observing s_t. Let $g(h^t)$ be the probability of $b_t=b^*$ at h^t. Let $g(h^t,\omega_c)$ be the probability of $b_t=b^*$ at h^t conditional on player 1 being the commitment type.

Lemma C.2 bounds the speed of learning at h^t from below. This implies a lower bound on the speed of learning when future player 2s observe b^* in period t, given that she knew that the probability with which player 2_t plays b^* is no more than $g(h^t)$. However, future player 2s' information *does not* nest that of player 2_t's, since they do not observe $\{a_{t-K},...,a_{t-1}\}$. As a result, they cannot interpret b_t in the same way as player 2_t does.

For every $s,t \in \mathbb{N}$ with $s>t$, I provide a lower bound on the informativeness of b_t about player 1's type from the perspective of player 2_s, as a function of the informativeness of b_t from the perspective of player 2_t. This together with Lemma C.2 establishes a lower bound on the informativeness of b_t from the perspective of future player 2s as a function of the probability that $b_t \neq b^*$. Using the entropy approach in Gossner (2011), one can obtain the lower bound on player 1's equilibrium payoff.

Let $\pi(h^t)$ be the probability with which player 2's belief assigns to the commitment type at h^t. By definition, $\pi(h^0)=\pi_0$. For every strategy profile σ, let $\mathcal{P}^\sigma$ be the probability measure over $\mathcal{H}$ induced by σ, let $\mathcal{P}^{\sigma,\omega_c}$ be the probability measure induced by σ conditional on player 1 being the commitment type, and let $\mathcal{P}^{\sigma,\omega_s}$ be the probability measure induced by σ conditional on player 1 being the strategic type. One can the write the posterior likelihood ratio as

$$\frac{\pi(h^t)}{1-\pi(h^t)} \Big/ \frac{\pi_0}{1-\pi_0}$$

$$= \frac{\mathcal{P}^{\sigma,\omega_c}(b_0)}{\mathcal{P}^{\sigma,\omega_s}(b_0)} \cdot \frac{\mathcal{P}^{\sigma,\omega_c}(b_1|b_0)}{\mathcal{P}^{\sigma,\omega_s}(b_1|b_0)} \cdot ... \cdot \frac{\mathcal{P}^{\sigma,\omega_c}(b_{t-1}|b_{t-2},...,b_0)}{\mathcal{P}^{\sigma,\omega_s}(b_{t-1}|b_{t-2},...,b_0)} \cdot \frac{\mathcal{P}^{\sigma,\omega_c}(a_{t-K},...,a_{t-1}|b_t,b_{t-1},...,b_0)}{\mathcal{P}^{\sigma,\omega_s}(a_{t-K},...,a_{t-1}|b_t,b_{t-1},...,b_0)} \tag{C.7}$$

Furthermore, for every $\epsilon > 0$ and every t, we know that:

$$\mathcal{P}^{\sigma,\omega_c}\Big(\pi^\sigma(b_0,b_1,...b_{t-1}) < \epsilon\pi_0\Big) \leq \epsilon\frac{1-\pi_0}{1-\pi_0\epsilon}, \tag{C.8}$$

in which $\pi^\sigma(b_0,b_1,...b_{t-1})$ is player 2's belief about player 1's type after observing $(b_0,...,b_{t-1})$ but before observing player 1's actions and s_t. For every $\epsilon > 0$, let

$$\rho^*(\epsilon) \equiv \frac{\epsilon\pi_0}{1-c\epsilon}. \tag{C.9}$$

If $\pi^\sigma(b_0,b_1,...b_{t-1}) \geq \epsilon\pi_0$ and player 2_t believes that $b_t = b^*$ occurs with probability less than $1-\epsilon$ after observing $(a_{\max\{0,t-K\}},...,a_{t-1}) = (a^*,...,a^*)$, then under probability measure $\mathcal{P}^\sigma$, the probability of $(a_{\max\{0,t-K\}},...,a_{t-1}) = (a^*,...,a^*)$ conditional on $(b_0,...,b_{t-1})$ is at least $\rho^*(\epsilon)$.

In order to show this, suppose by way of contradiction that the probability that $(a_{t-K},...,a_{t-1}) = (a^*,...,a^*)$ is strictly less than $\rho^*(\epsilon)$ conditional on $(b_0,...,b_{t-1})$. According to (C.9), after observing $(a_{t-K},...,a_{t-1}) = (a^*,...,a^*)$ in period t and given that $\pi^\sigma(b_0,b_1,...b_{t-1}) \geq \epsilon\pi_0$, $\pi(h^t)$ assigns probability strictly more than $1-c\epsilon$ to the commitment type. As a result, player 2 in period t believes that a^* is played with probability at least $1-c\epsilon$ at h^t. This contradicts hypothesis that she plays b^* with probability less than $1-\epsilon$.

Next, I study the believed distribution of b_t from the perspective of player 2_s conditional on the event that $\pi^\sigma(b_0,b_1,...b_{t-1}) \geq \epsilon\pi_0$. Let $\mathcal{P}(\sigma,t,s) \in \Delta(\Delta(A^K))$ be player 2's signal structure in period $s(\geq t)$ about $(a_{t-K},...,a_{t-1})$ under equilibrium σ. For every small enough $\eta > 0$, given that $\mathcal{P}(\sigma,t)$ assigns probability at least $\rho^*(\epsilon)$ to $(a_{\max\{0,t-K\}},...,a_{t-1}) = (a^*,...,a^*)$, the probability with which $\mathcal{P}(\sigma,t,s)$ assigns to event $(a_{\max\{0,t-K\}},...,a_{t-1}) = (a^*,...,a^*)$ occurring with probability less than $\eta\rho^*(\epsilon)$ is bounded from above by:

$$\frac{\eta\rho^*(\epsilon)(1-\rho^*(\epsilon))}{(1-\eta\rho^*(\epsilon))\rho^*(\epsilon)} = \eta\frac{1-\rho^*(\epsilon)}{1-\rho^*(\epsilon)\eta}. \tag{C.10}$$

Let $g(t|h^s)$ be player 2's belief about the probability with which b^* is played in period t when she observes h^s. Let $g(t, \omega_c|h^s)$ be her belief about the probability with which b^* is played in period t conditional on player 1 being committed. When player 2_t believes that $(a_{\max\{0,t-K\}}, ..., a_{t-1}) = (a^*, a^*, ..., a^*)$ occurs with probability more than $\eta\rho^*(\epsilon)$, we have:

$$g(t|h^s) \leq 1 - \epsilon\eta\rho^*. \tag{C.11}$$

Applying (C.11), we obtain a lower bound on the KL-divergence between $g(t, \omega_c|h^s)$ and $g(t|h^s)$. This is the lower bound on the speed with which player 2_s at h^s will learn through $b_t = b^*$ about player 1's type, which applies to all events except for one that occurs with probability less than $\eta\frac{1-\rho^*}{1-\rho^*\eta}$. Therefore, for every ϵ and π_0, there exists $\delta^* \in (0, 1)$ such that when $\delta > \delta^*$, strategic-type player 1's discounted average payoff by playing a^* in every period is at least:

$$\Big(1 - \epsilon - \epsilon\frac{1-\pi_0}{1-\pi_0\epsilon}\Big)u_1(a^*, b^*) + \Big(\epsilon + \epsilon\frac{1-\pi_0}{1-\pi_0\epsilon}\Big)\min_{b\in B} u_1(a^*, b) - \epsilon. \tag{C.12}$$

Let $\epsilon \to 0$ and $\delta \to 1$, (C.12) implies that with probability at least $1 - \varepsilon$, player 1's discounted average payoff from playing a^* in every period is at least $(1 - \epsilon)u_1(a^*, b^*)$. Take $\epsilon \to 0$, one can obtain that the patient player's discounted average payoff is at least $u_1(a^*, b^*)$ in every equilibrium.

C.2 Existence of Equilibrium

I establish the existence of equilibrium when the private signal is unboundedly informative about a^*, $K \geq 1$, and δ is large enough. For every $s \in S$, let $a(s) \equiv \min_{a\in A}\{f(s|a) > 0\}$ and let $b(s) \in B$ be player 2's strict best reply to $a(s)$. For every $a \in A$, let $v(a) \equiv \sum_{s\in S} f(s|a)u_1(a, b(s))$. Let

$$S' \equiv \Big\{s \in S \Big| \exists a \prec a^* \text{ such that } f(s|a) > 0\Big\} \text{ and } S^* \equiv \Big\{s \in S \Big| f(s|a^*) > 0\Big\}.$$

When $S' \cap S^* \neq \{\varnothing\}$, we have $\sum_{s\in S'} f(s|a) > 0$ for every $a \preceq a^*$, and let $p^* \equiv \min_{a\preceq a^*} \sum_{s\in S'} f(s|a)$. I show that the following strategy profile and belief constitute a Perfect Bayesian equilibrium.

- If $t = 0$, or $t \geq 1$, $(b_0, ..., b_{t-1}) = (b^*, ..., b^*)$ and $a_{t-1} = a^*$, then player 1 plays a^*, player 2_t believes that $a_t = a^*$ upon receiving any $s_t \in S^*$ and plays b^*, and believes that $a_t = a(s_t)$ upon receiving any $s_t \notin S^*$ and plays $b(s_t)$.

- At any other history, player 2_t believes that $a_t = a(s_t)$ upon receiving any $s_t \in S$, and plays

$b(s_t)$. Player 1 plays $\arg\max_{a\in A} v(a)$ in period t if there exists $\tau < t$ such that $b_\tau \neq b^*$. At histories where there exists no $\tau < t$ such that $b_\tau \neq b^*$ but $a_{t-1} \neq a^*$, player 1 plays a^* if

$$(1-\delta)v(a^*) + \delta \sum_{s\in S'} f(s|a^*) \max_{a\in A} v(a) + \delta \sum_{s\notin S'} f(s|a^*) u_1(a^*, b^*)$$

$$\geq \max_{\widetilde{a}\neq a^*} \Big\{ \frac{(1-\delta)v(\widetilde{a}) + \delta \sum_{s\in(S\backslash S^*)\cup S'} f(s|\widetilde{a}) \max_{a\in A} v(a)}{1-\delta\sum_{s\in S^*\backslash S'} f(s|\widetilde{a})} \Big\}, \tag{C.13}$$

and plays the action defined by the following expression if inequality (C.13) is violated:

$$\arg\max_{\widetilde{a}\neq a^*} \Big\{ \frac{(1-\delta)v(\widetilde{a}) + \delta \sum_{s\in(S\backslash S^*)\cup S'} f(s|\widetilde{a}) \max_{a\in A} v(a)}{1-\delta\sum_{s\in S^*\backslash S'} f(s|\widetilde{a})} \Big\}.$$

Player 2's strategy is optimal given her belief. Player 2's belief at on-path history respects Bayes Rule since every period t on-path history satisfies $(b_0, ..., b_{t-1}) = (b^*, ..., b^*)$ and $a_{t-1} = a^*$, in which case both types of player 1 play a^* and player 2_t believes that $a_t = a^*$ upon observing any $s_t \in S^*$. I verify player 1's incentive constraints by considering two cases separately.

1. Suppose $S' \cap S^* = \{\varnothing\}$, i.e., the distribution over private signals is such that $f(s|a) = 0$ for every $a \prec a^*$ and $s \in S$ satisfying $f(s|a^*) > 0$. In period t, player 1's stage-game payoff from playing a^* is $u_1(a^*, b^*)$. When he plays any $a \neq a^*$, player 2_t plays $a(s_t)$ at any history after observing any s_t that occurs with positive probability under a, from which player 1's stage-game payoff is no more than $u_1(a, \mathrm{BR}_2(a))$, which is no more than $u_1(a^*, b^*)$.

2. Suppose $S' \cap S^* \neq \{\varnothing\}$. Player 1's continuation value from playing a^* is $u_1(a^*, b^*)$ at every on-path history. Suppose he makes a one-shot deviation and plays $a \succ a^*$ at an on-path history, then his stage-game payoff is no more than $\max\{u_1(a, b^*), u_1(a, \mathrm{BR}_2(a))\}$, which is no more than $u_1(a^*, b^*)$, and his continuation value is no more than $u_1(a^*, b^*)$, which means that he cannot strictly profit from such a deviation. Suppose he makes a one-shot deviation and plays $a \prec a^*$ at an on-path history, then his stage-game payoff is no more than $u_1(a', b^*)$ and his continuation value is at most

$$\max \Big\{ \max_{a\succ a^*} u_1(a, b^*), \quad (1-\delta)u_1(a', b^*) + \delta p^* \max_{a\in A} v(a) + \delta(1-p^*) u_1(a^*, b^*) \Big\}, \tag{C.14}$$

where the first term is player 1's maximal continuation value when he plays $a \succ a^*$ at histories where player 2 has not played actions other than b^* but player 1's action in the previous period is not a^*, and the second term is player 1's maximal continuation value when he plays $a \preceq a^*$ at such histories. The value of $\max_{a \succ a^*} u_1(a, b^*)$ is strictly less than $u_1(a^*, b^*)$ since $u_1(a, b)$ strictly decreases in a, the value of $\max_{a \in A} v(a)$ is strictly less than $u_1(a^*, b^*)$ since a^* is player 1's unique Stackelberg action, $S^* \cap S' \neq \{\varnothing\}$, and $u_1(a, b)$ strictly increases in b. Therefore, (C.14) is strictly less than $u_1(a^*, b^*)$ when δ is large enough. It implies that when δ is large enough, playing a' is not a profitable one-shot deviation.

When $a_{t-1} \neq a^*$ but there is no $\tau < t$ such that $b_\tau \neq b^*$, notice that the left-hand-side of (C.13) is player 1's continuation value from playing a^*, and the right-hand-side is his continuation value from playing $\widetilde{a} \neq a^*$. This verifies his incentive constraint. When there exists $\tau < t$ such that $b_\tau \neq b^*$, player 2 plays $b(s)$ upon observing s, and it is optimal for player 1 to play $\arg\max_{a \in A} v(a)$.

C.3 Proof of Theorem 3

I establish Theorem 3 by modifying the constructive proof of Theorem 1. Without loss of generality, I focus on signal distributions such that $f(\cdot|a') \neq f(\cdot|a^*)$. This is because when a^* and a' generates the same distribution over private signals, the constructive proof of Theorem 1 still applies. In order to avoid repetition, I focus on the case in which $b^* = b^{**}$ and $b' = b''$. The other two cases can be shown similarly. When $b^* = b^{**}$ and $b' = b''$, there exists $q^* \in (0,1)$ such that b^* is a strict best reply to $qa^* + (1-q)a'$ if and only if $q > q^*$, and b' is a strict best reply to $qa^* + (1-q)a'$ if and only if $q < q^*$, and player 2 is indifferent between b^* and b' when player 1's action is $q^*a^* + (1-q^*)a'$. Without loss of generality, I adopt the normalization that $u_1(a^*, b^*) = 1$ and $u_1(a', b') = 0$. Let

$$S' \equiv \left\{ s \in S \middle| f(s|a') > 0 \right\}.$$

Since a^* is not strongly separable from a', $f(s|a^*) > 0$ only if $s \in S'$. Recall that S is a completely ordered set. For every $\beta : S \to \Delta\{b^*, b'\}$, I say that β is monotone if for every $s \succ s'$ with $s, s' \in S'$, $\beta(s')$ assigns positive probability to b^* implies that $\beta(s)$ assigns probability 1 to b^*, and $\beta(s)$ assigns positive probability to b' implies that $\beta(s')$ assigns probability 1 to b'. For every monotone β, let $f_1(\beta)$ be the probability of action b^* when player 1 plays a^* and player 2 responds according to β, and let $f_0(\beta)$ be the probability of action b^* when player 1 plays a' and player 2 responds according to β. Let

$$F \equiv \Big\{(f_0, f_1) \in [0,1]^2 \Big| \text{there exist } \alpha \in \Delta\{a^*, a'\} \text{ and monotone } \beta \text{ such that}$$
$$\beta \text{ best replies to } \alpha \text{ and } f_0(\beta) = f_0, \quad f_1(\beta) = f_1\Big\}.$$

Since a^* is not strongly separable from a',

1. There exists $\varepsilon > 0$ that depends only on the signal distribution such that $f_0(\beta) \geq \varepsilon f_1(\beta)$ for every monotone β, and $f_0(\beta) \leq (1-\varepsilon) f_1(\beta)$ for every monotone β satisfying $f_0(\beta) < \varepsilon$.[1]
2. For every $f_0 \in [0,1]$, there exists $f_1 \in [0,1]$ such that $(f_0, f_1) \in F$.
3. There exists a continuous and strictly increasing function $g : [0,1] \to [0,1]$ with $g(0) = 0$ and $g(1) = 1$ such that $(x, g(x)) \in F$ for every $x \in [0,1]$.
4. There exists $\underline{q} > 0$ such that when player 1 plays $\underline{q}a^* + (1-\underline{q})a'$, $\beta(s) = b'$ for all $s \in S$ is player 2's best reply.

Let Φ be the set of monotone β, let $\overline{\beta}$ be the constant mapping such that $\overline{\beta}(s) = b^*$ for every $s \in S$, and let $\underline{\beta}$ be the constant mapping such that $\underline{\beta}(s) = b'$ for every $s \in S$. Let $h^t \equiv (b_0, ..., b_{t-1})$ be the history of player 2's actions. Let $\mathcal{H}$ be the set of h^t, which also contains the null history $\varnothing$.

Consider the following strategy profile in which player 1 only plays a^* and a' on the equilibrium path. Players' on-path behaviors are characterized by $\alpha : \mathcal{H} \times \{a^*, a'\} \to \Delta\{a^*, a'\}$ and $\phi : \mathcal{H} \times \{a^*, a'\} \to \Phi$ where α is player 2_t's belief about a_t after observing $\{a_{t-K}, ..., a_{t-1}\}$ and $\{b_0, ..., b_{t-1}\}$ but before observing s_t, and ϕ is player 2_t's strategy that maps her private signals to a distribution over $\{b^*, b'\}$. Both ϕ and α depend only on the history of player 2's actions as well as player 1's action in the period before. According to the properties of monotone β, one can replace $\phi : \mathcal{H} \times \{a^*, a'\} \to \Phi$ with $f_0 : \mathcal{H} \times \{a^*, a'\} \to [0,1]$ and $f_1 : \mathcal{H} \times \{a^*, a'\} \to [0,1]$ such that $(f_0(h^t, a_{t-1}), f_1(h^t, a_{t-1})) \in F$ for every $h^t \in \mathcal{H}$ and $a_{t-1} \in \{a^*, a'\}$. Let $V(h^t, a_{t-1})$ be the strategic type player 1's continuation value at (h^t, a_{t-1}) under the above strategy profile. Similar to the proof of Theorem 1, I require functions α, ϕ, and V to satisfy the following conditions:

1. $\alpha(\varnothing) = \underline{q}a^* + (1-\underline{q})a'$, $\phi(\varnothing) = \underline{\beta}$, and $V(\varnothing) = 0$.
2. For every $h^t \in \mathcal{H}$ such that $b_{t-1} = b^*$ and b' has not occurred after the first time b^* occurred, we have $\alpha(h^t, a^*) = a^*$, $\phi(h^t, a^*) = \overline{\beta}$, and $V(h^t, a^*) = 1$.

[1] If there exists $s' \in S$ such that $f(s'|a') > 0$ and $f(s'|a^*) = 0$, then $f_0(\beta) \leq (1-\varepsilon) f_1(\beta)$ for every monotone β.

The values of functions f_0, f_1, and V at other histories are defined as follows. When $t = 1$, $V(b', a') = 0$ and $V(b', a^*) = -\frac{1-\delta}{\delta} u_1(a^*, b')$, which implies that player 1 is indifferent between a^* and a' in period 0. For every $t \geq 2$ and on-path h^t such that $b_{t-1} = b'$, player 1's incentive constraint requires him to be indifferent between a^* and a', which gives:

$$V(h^t, a) = f_0(h^t, a)\Big((1-\delta)u_1(a', b^*) + \delta \underbrace{V(h^t, b^*, a')}_{=1}\Big) + (1 - f_0(h^t, a))\Big((1-\delta)u_1(a', b') + \delta V(h^t, b', a')\Big)$$

$$= f_1(h^t, a)\Big(\underbrace{(1-\delta)u_1(a^*, b^*) + \delta V(h^t, b^*, a^*)}_{=1}\Big) + (1 - f_1(h^t, a))\Big((1-\delta)u_1(a^*, b') + \delta V(h^t, b', a^*)\Big).$$

I show that for every $V(h^t, a) \in [0, -\frac{1-\delta}{\delta} u_1(a^*, b')]$, there exist f_0, f_1, $V(h^t, b', a')$, $V(h^t, b', a^*)$ and $V(h^t, b^*, a')$ that satisfy the above incentive constraint, and moreover, $(f_0, f_1) \in F$, $V(h^t, b^*, a') = 1$ and

$$V(h^t, b', a'), V(h^t, b', a^*) \in \Big[0, -\frac{1-\delta}{\delta} u_1(a^*, b')\Big].$$

Let $f_1^* \in [0, 1]$ be such that

$$f_1^* + (1 - f_1^*)(1-\delta)u_1(a^*, b') = -\frac{1-\delta}{\delta} u_1(a^*, b'),$$

and let f_0^* be such that $(f_0^*, f_1^*) \in F$. Such f_1^* exists since $u_1(a^*, b') < u_1(a', b') = 0$. Consider two cases. First, consider the case in which

$$f_0^*((1-\delta)u_1(a', b^*) + \delta) > -\frac{1-\delta}{\delta} u_1(a^*, b'). \tag{C.15}$$

Then there exists $V(h^t, b', a^*) \in [0, -\frac{1-\delta}{\delta} u_1(a^*, b')]$ such that when $f_1(h^t, a)$ satisfies

$$f_1(h^t, a) + (1 - f_1(h^t, a))\Big((1-\delta)u_1(a^*, b') + \delta V(h^t, b', a^*)\Big) = -\frac{1-\delta}{\delta} u_1(a^*, b'), \tag{C.16}$$

and $f_0(h^t, a)$ satisfies $(f_0(h^t, a), f_1(h^t, a)) \in F$, I show that when δ is close enough to 1, we have

$$f_0(h^t, a)\Big((1-\delta)u_1(a', b^*) + \delta\Big) < -\frac{1-\delta}{\delta} u_1(a^*, b').$$

Let $v \equiv -\frac{1-\delta}{\delta} u_1(a^*, b')$ and suppose by way of contradiction that the above inequality is not true for any δ close to 1, then

$$\frac{f_0(h^t,a)}{1-f_0(h^t,a)} > \frac{v}{(1-\delta)u_1(a',b^*)+\delta-v}.$$

When $V(h^t,b',a^*) = -\frac{1-\delta}{\delta}u_1(a^*,b')$, we have $\frac{f_1(h^t,a)}{1-f_1(h^t,a)} = \frac{v}{1-v}$. This implies that

$$\frac{f_0(h^t,a)}{1-f_0(h^t,a)}\Big/\frac{f_1(h^t,a)}{1-f_1(h^t,a)} > \frac{v}{(1-\delta)u_1(a',b^*)+\delta-v}\Big/\frac{v}{1-v}, \tag{C.17}$$

with the right-hand-side converging to 1 as δ goes to 1. Since $f_0 \leq (1-\varepsilon)f_1$ for every $(f_0,f_1) \in F$ such that f_0 is small enough, and according to (C.16), $f_1(h^t,a)$ converges to 0 as $\delta \to 1$, there exists $\underline{\delta} \in (0,1)$ such that for every $\delta > \underline{\delta}$,

$$\frac{f_0(h^t,a)}{1-f_0(h^t,a)}\Big/\frac{f_1(h^t,a)}{1-f_1(h^t,a)} < 1-\frac{\varepsilon}{2}. \tag{C.18}$$

Inequalities (C.17) and (C.18) contradict each other. The intermediate value theorem implies the existence of f_0, f_1, $V(h^t,b',a')$, $V(h^t,b',a^*)$ and $V(h^t,b^*,a')$ that satisfy my requirements.

Second, consider the case in which

$$f_0^*((1-\delta)u_1(a',b^*)+\delta) \leq -\frac{1-\delta}{\delta}u_1(a^*,b').$$

I show that there exists $V(h^t,b',a') \in [0,-\frac{1-\delta}{\delta}u_1(a^*,b')]$ such that when $V(h^t,b^*,a')=1$, $V(h^t,b',a^*) = 0$, $f_1(h^t,a)$ given by

$$f_1(h^t,a)+(1-f_1(h^t,a))(1-\delta)u_1(a^*,b') = -\frac{1-\delta}{\delta}u_1(a^*,b'),$$

and $f_0(h^t,a)$ is such that $(f_0(h^t,a),f_1(h^t,a)) \in F$, the incentive constraint is satisfied. Suppose by way of contradiction that the above statement is not true, then when $V(h^t,b',a') = -\frac{1-\delta}{\delta}u_1(a^*,b')$, we have the following inequality

$$f_0(h^t,a)\Big((1-\delta)u_1(a',b^*)+\delta\Big) - (1-f_0(h^t,a))(1-\delta)u_1(a^*,b') < -\frac{1-\delta}{\delta}u_1(a^*,b'). \tag{C.19}$$

Let $v \equiv -\frac{1-\delta}{\delta}u_1(a^*,b')$, we have $f_1(h^t,a) = \frac{v(1+\delta)}{1+\delta v}$ and since a^* is not strongly separable from a', we have

$$f_0(h^t,a) \geq \varepsilon\frac{v(1+\delta)}{1+\delta v}.$$

I bound the value of the following expression from below

$$\varepsilon\frac{v(1+\delta)}{1+\delta v}\Big((1-\delta)\underbrace{u_1(a',b^*)}_{>0}+\delta\Big)+\Big(1-\varepsilon\frac{v(1+\delta)}{1+\delta v}\Big)\delta v-v,$$

which is at least

$$\varepsilon\frac{v(1+\delta)}{1+\delta v}+\Big(1-\varepsilon\frac{v(1+\delta)}{1+\delta v}\Big)\delta v-c=v\Big\{\frac{\varepsilon(1+\delta)}{1+\delta v}(1-\delta v)-(1-\delta)\Big\}$$

Since $v\to 0$ as $\delta\to 1$ and $\varepsilon>0$ is independent of δ, the right-hand-side is strictly greater than 0 when δ is close enough to 1. This contradicts the hypothesis that (C.19), and the intermediate value theorem implies that the incentive constraint can be satisfied by some $V(h^t,b',a')\in[0,-\frac{1-\delta}{\delta}u_1(a^*,b')]$, $V(h^t,b^*,a')=1$, and $V(h^t,b',a^*)=0$. The two cases together provide an algorithm that defines the continuation values such that $V=1$ when b^* was played the period before, and $V\in[0,-\frac{1-\delta}{\delta}u_1(a^*,b')]$ when b' was played the period before.

Next, I specify players' strategies at off-path histories and verify that player 1 has no incentive to play any action other than a^* and a'. For every $s_t\notin S'$, player 2 believes that player 1's action is a' and plays b'. If player 2_t observes that $a_{t-1}\notin\{a^*,a'\}$, then player 2_t believes that $a_t=a'$ and plays b'. I show that under this belief and player 2's off-path strategies, player 1 does not have a strict incentive to play actions other than a^* and a' at any on-path history. When his continuation value $V(h^t,a)$ is 0, player 2 plays b' no matter which signal he observes, so player 1's payoff is strictly greater by playing his lowest action a' compared to any action $a^\dagger\notin\{a^*,a'\}$. When $V(h^t,a)=1$, player 1's continuation value is at most $-\frac{1-\delta}{\delta}u_1(a^*,b')$ in period $t+1$ if he plays $a^\dagger\notin\{a^*,a'\}$ in period t, which is strictly less than his payoff from playing a^*. Since $V(h^t,a)\in[0,-\frac{1-\delta}{\delta}u_1(a^*,b')]\cup\{1\}$ at any on-path history, I only need show that player 1 has no incentive to play $a^\dagger$ when $V(h^t,a)\in(0,-\frac{1-\delta}{\delta}u_1(a^*,b')]$. For every $(f_0,f_1)\in F$, there exists a monotone β such that $f_0(\beta)=f_0$ and $f_1(\beta)=f_1$. Let $f^\dagger(\beta)$ be the probability of b^* if player 1 plays $a^\dagger$ and player 2 plays β when $s\in S'$ and plays a' if $s\notin S'$. Since $\mathbf{f}$ satisfies MLRP, we have $f^\dagger(\beta)<f_1(\beta)$. Player 1's expected payoff from playing $a^\dagger$ is at most

$$f^\dagger(\beta)\Big((1-\delta)u_1(a^\dagger,b^*)-(1-\delta)u_1(a^*,b')\Big)-(1-f^\dagger(\beta))(1-\delta)\underbrace{u_1(a^\dagger,b')}_{<0}\qquad\text{(C.20)}$$

which is a strictly increasing function of $f^\dagger(\beta)$. Since $V(h^t,b',a^*)\le-\frac{1-\delta}{\delta}u_1(a^*,b')$, we have

$f_1(\beta) \leq V(h^t, a)$. Therefore, (C.20) is at most

$$f_1(\beta)(1-\delta)(u_1(a^\dagger, b^*) - u_1(a^*, b')).$$

The above expression is no more than

$$V(h^t, a)(1-\delta)(u_1(a^\dagger, b^*) - u_1(a^*, b')).$$

This upper bound is strictly less than $V(h^t, a)$ when δ is close to 1. This implies that player 1 has no incentive to play any action other than a^* and a'.

I verify that when the prior probability of commitment type satisfies $\pi_0 \leq \left(\frac{\underline{q}}{2}\right)^K \left(\frac{\underline{q}}{2-\underline{q}}\right)$, player 2's posterior belief is uniformly bounded below $\underline{q}/2$ at every history such that the previous period action profile is not (a^*, b^*). Recall that $M = +\infty$. When player 1 plays a^* in every period from 0 to t, the history of player 2's actions cannot switch from b^* to b'. Therefore, at every history in period $t \geq 1$ where the previous period action profile is not (a^*, b^*), player 2's posterior belief assigns positive probability to the commitment type if and only if $h^t = \{b', ..., b'\}$ and $(a_{t-K}, ..., a_{t-1}) = (a^*, ..., a^*)$. Let π_t be the posterior probability of commitment type at such a history. I show that $\pi_t \leq \underline{q}/2$ by induction on calendar time t. When $t = 0$, $\pi_0 \leq \underline{q}/2$ since $\pi_0 \leq \left(\frac{\underline{q}}{2}\right)^K \left(\frac{\underline{q}}{2-\underline{q}}\right)$. Suppose $\pi_s \leq \underline{q}/2$ for every $s \leq t-1$. Since the unconditional probability with which player 1 plays a^* is at least $\underline{q}$ in every period and the induction hypothesis requires that $\pi_s \leq \underline{q}/2$ for every $s \leq t-1$, the probability with which the strategic type plays H at each of those histories before period t must be at least $\underline{q}/2$. Let $P^{\omega_s}(\cdot)$ be the probability measure induced by the equilibrium strategy of the strategic type. Let $P^{\omega_c}(\cdot)$ be the probability measure induced by the commitment type. Let E_t be the event that $(a_{\max\{0,t-K\}}, ..., a_{t-1}) = (a^*, ..., a^*)$. Let F_t be the event that $(b_0, ..., b_{t-1}) = (b', ..., b')$. According to Bayes rule,

$$\frac{\pi_t}{1-\pi_t} \Big/ \frac{\pi_0}{1-\pi_0} = \frac{P^{\omega_c}(E_t \cap F_t)}{P^{\omega_s}(E_t \cap F_t)} = \frac{P^{\omega_c}(E_t)}{P^{\omega_s}(E_t)} \cdot \frac{P^{\omega_c}(F_t|E_t)}{P^{\omega_s}(F_t|E_t)}.$$

Since the strategic type plays a^* with probability at least $\underline{q}/2$ in every period before t and N occurs with weakly lower probability under the strategy of type ω_c compared to that under type ω_s, we have

$$\frac{P^{\omega_c}(E_t)}{P^{\omega_s}(E_t)} \leq (\underline{q}/2)^{-K} \quad \text{and} \quad \frac{P^{\omega_c}(F_t|E_t)}{P^{\omega_s}(F_t|E_t)} \leq 1.$$

Since $\pi_0 \leq \left(\frac{\underline{q}}{2}\right)^K \left(\frac{\underline{q}}{2-\underline{q}}\right)$, the above two inequalities together imply that $\pi_t \leq \underline{q}/2$.

D Counterexample: Short-Run Players Observe Calendar Time

I show that Proposition 1 in the main text relies on the assumption that the short-run players cannot directly observe calendar time. Consider the following parametric configuration of the product choice game:

-	**T**rust	**N**o Trust
High Effort	$1, 2$	$-1, 1$
Low Effort	$1.5, -1$	$0, 0$

One can verify that players' stage-game payoff functions satisfy the condition for Proposition 1 since $c_T = 0.5$ which is strictly greater than 0 and is strictly less than $c_N = 1$. I focus on the case where $\pi_0 = \frac{1}{2}$ and $(K, M) = (1, 0)$, which is allowed by Proposition 1.

Unlike Proposition 1, I assume that consumers can directly observe calendar time. Consumers' strategy is $\{\beta_t(H), \beta_t(L)\}_{t=1}^{+\infty} \cup \{\beta_0\}$ where $\beta_t(a)$ is consumer t's probability of playing T when the seller's action is a in period $t-1$, and β_0 is consumer 0's probability of playing T. Let $V_t(a)$ be the seller's continuation value in period t when his period $t-1$ action was a.

I construct a class of equilibria in which the strategic-type seller's discounted average payoff is approximately $\frac{5}{6}$ when δ is close to 1. This implies that his equilibrium payoff is bounded below his Stackelberg payoff, which equals 1.

1. In period $t \equiv 3k$ for every $k \in \mathbb{N}$, the strategic seller plays L and $\beta_0 = \beta_t(H) = \frac{1}{3\delta}$ and $\beta_t(L) = 0$. Consumers believe that the seller is committed with probability $1/2$ upon observing $a_{t-1} = H$ and is committed with probability 0 upon observing $a_{t-1} = L$.

2. In period $t \equiv 3k + 1$ for every $k \in \mathbb{N}$, the strategic seller plays H and $\beta_t(H) = \beta_t(L) = 1$. Consumers believe that the seller is committed with probability 1 upon observing $a_{t-1} = H$ and is committed with probability 0 upon observing $a_{t-1} = L$.

3. In period $t \equiv 3k + 2$ for every $k \in \mathbb{N}$, the strategic seller plays H if $a_{t-1} = H$ and plays L if $a_{t-1} = L$. Consumers' strategy is such that $\beta_t(H) = 1$ and $\beta_t(L) = 0$. Consumers believe that the seller is committed with probability $1/2$ upon observing $a_{t-1} = H$ and is committed with probability 0 upon observing $a_{t-1} = L$.

When the strategic-type seller follows his equilibrium strategy, his discounted average payoff equals

$$\frac{\frac{3}{2} \cdot \frac{1}{3\delta} + \delta + \delta^2}{1 + \delta + \delta^2},$$

which converges to $\frac{5}{6}$ as $\delta \to 1$. Verifying the consistency of consumers' beliefs as well as consumers' incentive constraints are straightforward. Next, I verify the strategic seller's incentive constraints.

1. In period $t \equiv 3k$, the seller has a strict incentive to play L since $\beta_{t+1}(H) = \beta_{t+1}(L) = 1$.

2. In period $t \equiv 3k+1$, the seller's continuation value from playing H is

$$(1-\delta) + (1-\delta)\delta + (1-\delta)\delta^2 \frac{1}{2\delta} + \delta^3 V_{t+3}(L)$$

   while his continuation value from a one-shot deviation (i.e., playing L in the current period and follow the equilibrium strategy starting from the next period) is

$$(1-\delta)\frac{3}{2} + \delta^3 V_{t+3}(L).$$

   When δ is close to 1, we have $(1-\delta) + (1-\delta)\delta + (1-\delta)\delta^2 \frac{1}{2\delta} \geq (1-\delta)\frac{3}{2}$, which verifies the seller's incentive constraint.

3. In period $t \equiv 3k+2$. If $a_{t-1} = H$, then the seller's continuation value from playing H is

$$(1-\delta) + (1-\delta)\delta \frac{1}{2\delta} + \delta^2 V_{t+2}(L),$$

   and his continuation value after a one-shot deviation is

$$(1-\delta)\frac{3}{2} + \delta^2 V_{t+2}(L).$$

   The two payoffs are equal, which implies that the seller has an incentive to play H when $a_{t-1} = H$.

   If $a_{t-1} = L$, then the seller's continuation value from playing L is $\delta^2 V_{t+2}(L)$, and his continuation value from a one-shot deviation is

$$-(1-\delta) + (1-\delta)\delta \frac{1}{2\delta} + \delta^2 V_{t+2}(L),$$

   which is strictly less than $\delta^2 V_{t+2}(L)$. This inequality verifies the seller's incentive to play L when $a_{t-1} = L$.