Online Appendix: Reputation Effects under Short Memories

Harry PEI

September 24, 2023

Online Appendix C: Proof of Corollary 1

Recall that $S \equiv A^K$ and $s^* \in S$ is the state where all of player 1's actions in the last K periods were a^*. For every Nash equilibrium $\sigma \equiv (\sigma_1, \sigma_2)$, let $S'(\sigma) \subset S$ be such that $s \in S'(\sigma)$ if and only if (i) $s \neq s^*$, and (ii) there exists a pure strategy $\widehat{\sigma}_1$ that best replies to σ_2 such that s^* is reached within a finite number of periods when the initial state is s and player 1 uses strategy $\widehat{\sigma}_1$. Let $S''(\sigma) \equiv S \Big\backslash \Big(\{s^*\} \cup S'(\sigma)\Big)$.

Recall the definitions of inflow $\mathcal{I}(\cdot)$ and outflow $\mathcal{O}(\cdot)$ in equations (B.2) and (B.3) in the main text. Since player 1's equilibrium strategy σ_1 must satisfy the no-back-loop property, we have $\mathcal{I}(S'(\sigma)) = \mathcal{O}(S''(\sigma)) = 0$. Statement 1 of Theorem 2 implies that there exists a constant $C \in \mathbb{R}_+$ such that $\mu(s^*) \geq 1 - C(1-\delta)$ for every equilibrium under discount factor δ. Therefore, $\sum_{s \in S'(\sigma)} \mu(s) + \sum_{s \in S''(\sigma)} \mu(s) \leq C(1-\delta)$. Recall that $p(s)$ is the probability that the state is s conditional on calendar time being K and player 1 is the strategic type. Since $\mathcal{O}(S''(\sigma)) = 0$, we have

$$\sum_{s \in S''(\sigma)} p(s) \leq \sum_{s \in S''(\sigma)} \mu(s) \leq C(1-\delta)$$

Lemma B.2 in the main text implies that:

$$\Big|\mathcal{I}(S''(\sigma)) - \mathcal{O}(S''(\sigma))\Big| = \frac{1-\delta}{\delta}\Big|\sum_{s \in S''(\sigma)} \Big(\mu(s) - p(s)\Big)\Big| \leq \frac{C(1-\delta)^2}{\delta}. \tag{1}$$

This together with $\mathcal{O}(S''(\sigma)) = 0$ implies that $\mathcal{I}(S''(\sigma)) \leq \frac{C(1-\delta)^2}{\delta}$. For every $t \geq K$ and $S' \subset S$, let $q_t(S')$ be the probability that the state in period t belongs to S' conditional on player 1 being the strategic type. Since $\mathcal{I}(S'(\sigma)) = \mathcal{O}(S''(\sigma)) = 0$,

$$(1-\delta^{t-K})q_t(S'(\sigma)) + \delta^{t-K} q_t(S''(\sigma)) \leq \sum_{s \neq s^*} \mu(s) \leq C(1-\delta). \tag{2}$$

Suppose $t \in \mathbb{N}$ is such that $\delta^t \in (\varepsilon, 1-\varepsilon)$, and δ is above some cutoff such that $\delta^{t-K} \in (\sqrt{\varepsilon}, 1-\sqrt{\varepsilon})$, inequality (2) implies that

$$q_t(S'(\sigma)) + q_t(S''(\sigma)) \le \max\left\{\frac{C(1-\delta)}{1-\delta^{t-K}}, \frac{C(1-\delta)}{\delta^{t-K}}\right\} \le \frac{C}{\sqrt{\varepsilon}}(1-\delta),$$

which implies that $q_t(\{s^*\}) \ge 1 - \frac{C}{\sqrt{\varepsilon}}(1-\delta)$. Let r_t be the probability with which the strategic-type player 1 *does not* play a^* in period t conditional on the period t state is s^*. Inequality (1) implies that:

$$(1-\delta)\delta^{t-K} \underbrace{\Big(1 - q_t(S'(\sigma)) - q_t(S''(\sigma))\Big)}_{=q_t(\{s^*\})} r_t \le \mathcal{I}(S''(\sigma)) \le \frac{C(1-\delta)^2}{\delta}. \tag{3}$$

Dividing both sides of inequality (3) by $1-\delta$, using the conclusion that $q_t(\{s^*\}) \ge 1 - \frac{C}{\sqrt{\varepsilon}}(1-\delta)$, as well as the hypothesis that $\delta \ge \sqrt{\varepsilon}$, we obtain:

$$r_t \le \frac{C}{\sqrt{\varepsilon}\delta} \cdot \frac{1}{1-\frac{C}{\sqrt{\varepsilon}}(1-\delta)} \cdot (1-\delta) \le \underbrace{\frac{C}{\varepsilon} \cdot \frac{1}{1-\frac{C}{\sqrt{\varepsilon}}(1-\sqrt{\varepsilon})}}_{\equiv C_\varepsilon} \cdot (1-\delta).$$

This yields the desired conclusion.

Online Appendix D: Patient Player's Equilibrium Behavior

This appendix compares the predictions on the patient player's action frequencies in my baseline model and the canonical reputation model of Fudenberg and Levine (1989).

Recall that in Fudenberg and Levine (1989), every short-run player observes the *entire sequence* of the patient player's past actions. In games where players' payoffs satisfy Assumption 1 as well as a generic assumption that neither player is indifferent between any pairs of *pure* action profiles (which implies that each player has a unique best reply to any of his opponent's pure action, and that player 1 has a unique optimal pure commitment action), Li and Pei (2021) show that the frequency with which player 1 plays a^* in equilibrium can be anything between $G^*(u_1, u_2)$ and 1, where $G^*(u_1, u_2)$ is defined as the value of the following constrained optimization problem:.

$$G^*(u_1, u_2) \equiv \min_{(\alpha_1, \alpha_2, b_1, b_2, q) \in \Delta(A) \times \Delta(A) \times B \times B \times [0,1]} \Big\{ q\alpha_1(a^*) + (1-q)\alpha_2(a^*) \Big\}, \tag{4}$$

subject to $b_1 \in \arg\max_{b \in B} u_2(\alpha_1, b)$, $b_2 \in \arg\max_{b \in B} u_2(\alpha_2, b)$, and

$$qu_1(\alpha_1, b_1) + (1-q)u_1(\alpha_2, b_2) \geq u_1(a^*, b^*). \tag{5}$$

For an interpretation of the linear program that defines $G^*(u_1, u_2)$, consider an optimization problem faced by a planner who chooses a distribution over action profiles in order to minimize the expected probability of a^* subject to the constraints that (i) each action profile in the support of this distribution satisfies player 2's myopic incentive constraint, and (ii) player 1's expected payoff from this distribution is no less than his commitment payoff $u_1(a^*, b^*)$. I show that for any stage-game where $G^*(u_1, u_2) > 0$, which is the case if and only if

$$u_1(a^*, b^*) > \max_{a \neq a^*} \max_{b \in \mathrm{BR}_2(a)} u_1(a, b), \tag{6}$$

the discounted frequency with which player 1 plays a^* in my model is strictly bounded above $G^*(u_1, u_2)$ regardless of the value of K.

Proposition. *Suppose (u_1, u_2) satisfies Assumptions 1 and 2, and there do not exist a pair of action profiles $(a, b), (a', b') \in A \times B$ and $i \in \{1, 2\}$ such that $u_i(a, b) = u_i(a', b')$.*

1. *If (u_1, u_2) violates (6), then $G^*(u_1, u_2) = 0$ and the minimal frequency with which player 1 plays a^* is 0 both in my model and in Fudenberg and Levine (1989)'s model.*

2. *If (u_1, u_2) satisfies (6), then there exists $\psi > 0$ such that for every $K \in \mathbb{N}$, there exists $\underline{\delta} \in (0, 1)$ such that for every $\delta > \underline{\delta}$, and in every equilibrium σ under δ, we have:*

$$\underbrace{\sum_{b \in B} F^{\sigma}(a^*, b)}_{\text{the discounted frequency of } a^* \text{ under } \sigma} > G^*(u_1, u_2) + \psi. \tag{7}$$

Proof. In the case where a^* is not player 1's optimal pure commitment action, by definition, $G^*(u_1, u_2) = 0$ and the proof of Theorem 2 in Appendix B implies that the lowest frequency with which player 1 plays a^* in my model is also 0. This establishes the first statement of the proposition.

In what follows, I focus on the case where a^* is player 1's optimal pure commitment action. According to the definition of $G^*(u_1, u_2)$, it is without loss of generality to focus on (α_1, b_1) and (α_2, b_2) such that $u_1(\alpha_1, b_1) \geq u_1(a^*, b^*) \geq u_1(\alpha_2, b_2)$. When players' stage-game payoffs satisfy Assumption 1, every $(\alpha_1, \alpha_2, b_1, b_2, q)$ that solves the constrained optimization problem must satisfy:

1. $b_1 = b^*$,

2. α_1 is a nontrivially mixed action that assigns positive probability to a^*,

3. $u_1(\alpha_1, b_1) > u_1(a^*, b^*) > u_1(\alpha_2, b_2)$ and constraint (5) is binding.

This is because:

1. The first requirement is implied by a^* being player 1's optimal pure commitment action.

2. The second requirement is implied by Proposition 1 in Li and Pei (2021).

3. Since $u_1(a, b)$ is strictly decreasing in a, requirement 2 implies that $u_1(\alpha_1, b_1) > u_1(a^*, b^*)$. Suppose by way of contradiction that $u_1(a^*, b^*) = u_1(\alpha_2, b_2)$, then the fact that a^* being player 1's optimal pure commitment action implies that α_2 assigns positive probability to a^* (this is because otherwise, actions lower than a^* can also induce player 2 to play b^*, in which case committing to this lower action gives player 1 a strictly higher payoff compared to committing to a^*). Consider another solution $(\alpha_1', \alpha_2', b_1', b_2', q')$ where $(\alpha_1', b_1') = (\alpha_1, b_1)$, α_2' assigns probability 1 to player 1's lowest action, b_2' best replies to α_2', and q' is chosen such that constraint (5) binds. Compared to $(\alpha_1, \alpha_2, b_1, b_2, q)$, $(\alpha_1', \alpha_2', b_1', b_2', q')$ increases the value of (4) without violating any constraint, which contradicts the optimality of $(\alpha_1, \alpha_2, b_1, b_2, q)$.

4. Suppose by way of contradiction that constraint (5) is not binding. Consider two cases separately. First, suppose $\alpha_1(a^*) > \alpha_2(a^*)$. One can decrease q to make (5) binding, which increases the value of (4). This is a contradiction. Second, suppose $\alpha_1(a^*) \leq \alpha_2(a^*)$. Consider another solution $(\alpha_1', \alpha_2', b_1', b_2', q')$ where $(\alpha_1', b_1') = (\alpha_1, b_1)$, α_2' assigns probability 1 to player 1's lowest action, b_2' best replies to α_2', and q' is chosen such that constraint (5) binds. Compared to $(\alpha_1, \alpha_2, b_1, b_2, q)$, $(\alpha_1', \alpha_2', b_1', b_2', q')$ increases the value of (4) without violating any constraint. This contradicts the optimality of $(\alpha_1, \alpha_2, b_1, b_2, q)$ in the program that defines $G^(u_1, u_2)$.

Let $S \equiv A^K$. Let us partition the entire set of states S according to player 2's information structure $S \equiv \cup_{j=1}^N S_j$, where $N \in \mathbb{N}$ can be computed from K and the number of actions in A. Recall the definition of $\mu \in \Delta(S)$. Let $\mu(S_j) \equiv \sum_{s \in S_j} \mu(s)$. Let $\alpha_j \in \Delta(A)$ be player 1's expected action conditional on S_j and $\beta_j \in \Delta(B)$ be player 2's action at S_j. We have

$$\Big| \sum_{b \in B} F^\sigma(a^*, b) - \sum_{j=1}^N \mu(S_j) \alpha_j(a^*) \Big| \leq 1 - \delta^K. \tag{8}$$

and

$$(1-\delta^K)\max_{(a,b)\in A\times B} u_1(a,b)+\delta^K\sum_{j=1}^{N}\mu(S_j)u_1(\alpha_j,\beta_j)\geq\sum_{t=0}^{+\infty}(1-\delta)\delta^t u_1(a_t,b_t). \tag{9}$$

Suppose by way of contradiction that for every $\underline{\delta}\in(0,1)$, there exist $\delta>\underline{\delta}$ and an equilibrium σ under δ such that $\sum_{b\in B}F^\sigma(a^*,b)<G^*(u_1,u_2)+\varepsilon$. When $\underline{\delta}$ is close to 1, there exists an element of player 2's information partition $S_j\in\{S_1,...,S_N\}$ such that (i) $\mu(S_j)$ is bounded away from 0, (ii) β_j assigns probability close to 1 to b^* and α_j is close to one of the optimal solutions to (4), i.e., the probability α_j assigns to a^* is bounded away from 1.

First, I show that S_j cannot be the partition element that contains the history where all of the last K actions were a^*. This is because the probability α_j assigns to a^* is bounded away from 1, which implies that if S_j contains the history where all of the last K actions were a^*, it must be the case that $\mathcal{O}(S_j)\geq\mu(S_j)(1-\alpha_j(a^*))$. Since (u_1,u_2) satisfies Assumptions 1 and 2, Lemma 3.4 in the main text implies that $\mathcal{O}(S_j)\leq\frac{2(1-\delta)}{\delta}$ if S_j contains the history where all of the last K actions were a^*. This leads to a contradiction.

Next, for every $a\in A$, let $K_j(a)$ be the number of times action a occurred in histories that belong to S_j. Suppose player 1 deviates and plays action a for $K_j(a)$ times every K periods for every $a\in A$. Under this deviation, player 1's discounted average payoff is close to

$$\frac{1}{K}\sum_{a\in A}K_j(a)u_1(a,\beta_j)\approx\frac{1}{K}\sum_{a\in A}K_j(a)u_1(a,b^*). \tag{10}$$

When δ is close to 1, (10) is bounded away from $u_1(a^*,b^*)$ since $K_j(a^*)\leq K-1$ and $u_1(a,b)$ is strictly decreasing in a. Hence, for every $\varepsilon>0$, there exists $\underline{\delta}\in(0,1)$ such that when $\delta>\underline{\delta}$, player 1's equilibrium payoff is at least $\frac{1}{K}\sum_{a\in A}K_j(a)u_1(a,b^*)-\varepsilon$. According to (9), we have

$$(1-\delta)^K\max_{(a,b)\in A\times B}u_1(a,b)+\delta^K\sum_{j=1}^{N}\mu(S_j)u_1(\alpha_j,\beta_j)\geq\frac{1}{K}\sum_{a\in A}K_j(a)u_1(a,b^*)-\varepsilon.$$

Since constraint (5) must be binding in the optimal solution, there is no equilibrium where the frequency of a^* is close to $G^*(u_1,u_2)$ when δ is close to 1. Hence, there exists $\eta>0$ and $\underline{\delta}\in(0,1)$ such that for every $\delta>\underline{\delta}$ and every Nash equilibrium σ under δ, we have $\sum_{b\in B}F^\sigma(a^*,b)>G^*(u_1,u_2)+\eta$.

$\square$

Online Appendix E: Learning from Noisy Signals

I study settings where the short-run players can only observe the summary statistics of some noisy signal about the patient player's action. Let $\widetilde{a}_t \in A$ be a signal of player 1's action a_t such that $\widetilde{a}_t = a_t$ with probability $1-\varepsilon$, and $\widetilde{a}_t$ is drawn according to $\alpha \in \Delta(A)$ with probability ε. Player 1 observes the entire history $h^t = \{a_s, b_s, \widetilde{a}_s\}_{s=0}^{t-1}$. Player 2_t only observes the number of times that each *signal realization* occurred in the last $\min\{t, K\}$ periods.

I show that a version of my no-back-loop lemma holds for all small enough ε. For every $t \geq K$, player 1's incentive in period t depends on his history only through $(\widetilde{a}_{t-K}, ..., \widetilde{a}_{t-1})$. Let $S \equiv A^K$ be the set of signal vectors of length K with a typical element denoted by $s \in S$, which I call a *state*. Without loss of generality, I focus on player 1's strategies that are measurable with respect to the state, that is, $\sigma_1 : S \to \Delta(A)$. I say that a pure strategy $\widehat{\sigma}_1 : S \to A$ induces an *ε-back-loop* if there exist a subset of states $\{s_0, ..., s_M\} \subset S$ such that $s_0 = s_M = (a^*, ..., a^*)$ and for every $i \in \{0, 1, ..., M-1\}$, if player 1 plays $\widehat{\sigma}_1(s_i)$ in state s_i, then it reaches state s_{i+1} in the next period with probability more than $1-\varepsilon$.

Proposition. *There exists $\varepsilon > 0$ such that for every $\alpha \in \Delta(A)$ and $\sigma_2 : \mathcal{H}_2 \to \Delta(B)$, if a pure strategy $\widehat{\sigma}_1$ best replies to σ_2, then $\widehat{\sigma}_1$ does not induce any ε-back-loop.*[1]

The proof can be found by the end of this section. After establishing this general no-back-loop lemma for all small ε, one can use the same argument to show Theorem 1, namely, player 1 can approximately secure payoff $u_1(a^*, b^*)$ in all equilibria when δ is close enough to 1 and ε is small enough. This is because conditional on observing $(\widetilde{a}_{t-K}, ..., \widetilde{a}_{t-1}) = (a^*, ..., a^*)$, the probability that player 2 assigns to player 1 playing a^* in the current period is close to 1. This implies that player 2's action is at least b^*, and therefore, player 1's payoff is approximately $u_1(a^*, b^*)$ when he plays a^* in every period. Theorems 2 and 3 can also be extended to environments where ε is small, i.e., the same cutoff $\overline{K}$ applies as long as ε is small enough. Intuitively, this is because when ε is small, the occupation measure of states other than $(\widetilde{a}_{t-K}, ..., \widetilde{a}_{t-1}) = (a^*, ..., a^*)$ is close to 0 since one can show that (i) when K is below the cutoff, the probability that other states being generated by player 1's deliberate behavior is close to 0, and (ii) when ε is close to 0, the probability that other states being generated by noise in player 2's signal is also close to 0.

Proof. Let $S \equiv A^K$ with a typical element denoted by $s \in S$. As in the proof of the no-back-loop lemma in Appendix A, it is without loss of generality to focus on player 1's strategies that depend only on the signal realizations in the last K periods. Therefore, I write player 1's strategy as $\sigma_1 : S \to \Delta(A)$.

[1] If player 2 has three or more actions and u_2 satisfies Assumption 2, then this conclusion holds for all $\sigma_2 : \mathcal{H}_2 \to \mathcal{B}^*$.

Suppose by way of contradiction that there exists σ_2 and a pure strategy $\widehat{\sigma}_1 : S \to A$ that best replies to σ_2 such that $\widehat{\sigma}_1$ induces an ε-back-loop. Let $\widehat{\sigma}_1(a_{t-K}, .., a_{t-1})$ be player 1's action under $\widehat{\sigma}_1$ conditional on the last K signal realizations. Then there exist two signal realizations $a', a'' \neq a^*$ such that $\widehat{\sigma}_1(a^*, ..., a^*) = a'$ and after player 1 plays a' at $(a^*, ..., a^*)$ and uses strategy $\widehat{\sigma}_1$, he will reach $(a'', a^*, ..., a^*)$ after a finite number of periods and $\widehat{\sigma}_1(a'', a^*, ..., a^*) = a^*$. Let $\beta^* \in \Delta(B)$ be player 2's action when all of the last K signal realizations were a^* and let $\beta'' \in \Delta(B)$ be player 2's action when player 1's last K signal realizations were $(a'', a^*, ..., a^*)$. By definition, a' is optimal for player 1 at $(a^*, ..., a^*)$ and a^* is optimal for player 1 at $(a'', a^*, ..., a^*)$. Therefore, we have

$$(1-\delta)u_1(a^*, \beta'') + \delta(1-\varepsilon)V(a^*, a^*, ..., a^*, a^*) + \delta\varepsilon\mathbb{E}_{\widetilde{a}\sim\alpha}[V(a^*, ..., a^*, \widetilde{a})]. \tag{11}$$

$$\geq (1-\delta)u_1(a', \beta'') + \delta(1-\varepsilon)V(a^*, a^*, ..., a^*, a') + \delta\varepsilon\mathbb{E}_{\widetilde{a}\sim\alpha}[V(a^*, ..., a^*, \widetilde{a})],$$

and

$$(1-\delta)u_1(a^*, \beta^*) + \delta(1-\varepsilon)V(a^*, a^*, ..., a^*, a^*) + +\delta\varepsilon\mathbb{E}_{\widetilde{a}\sim\alpha}[V(a^*, ..., a^*, \widetilde{a})] \tag{12}$$

$$\leq (1-\delta)u_1(a', \beta^*) + \delta(1-\varepsilon)V(a^*, a^*, ..., a^*, a') + \delta\varepsilon\mathbb{E}_{\widetilde{a}\sim\alpha}[V(a^*, ..., a^*, \widetilde{a})].$$

Assumption 1 implies that β^* cannot strictly FOSD β''. Assumption 2 implies that β'' weakly FOSDs β^*. In what follows, I compare player 1's discounted average payoff starting from history $(a'', a^*, ..., a^*)$ under strategy $\widehat{\sigma}_1$ to his discounted average payoffs under the following two deviations. For simplicity, I use t^* to denote the calendar time that the deviation starts and use t to denote a generic calendar time.

1. Plays a' at history $(a'', a^*, ..., a^*)$ and then follows $\widehat{\sigma}_1$
2. Plays a'' at history $(a'', a^*, ..., a^*)$, and then
   - Plays a^* in period t if $t^* + 1 \leq t \leq t^* + K - 1$ and $\widetilde{a}_s = a_s$ for every $s \in \{t^*, ..., t-1\}$.
   - Follows strategy $\widehat{\sigma}_1$ otherwise.

Recall the definition of U in equation (A.3) of Appendix A. Since the state space is finite, for every $\eta > 0$, there exists $\overline{\varepsilon} > 0$ such that for every $\alpha \in \Delta(A)$ and $\varepsilon < \overline{\varepsilon}$, player 1's discounted average payoff from strategy $\widehat{\sigma}_1$ belongs to an η-neighbourhood of

$$\frac{(1-\delta)u_1(a^*, \beta'') + (1-\delta)\delta u_1(a', \beta^*) + (\delta^2 - \delta^{M-1})U}{1 - \delta^{M-1}}, \tag{13}$$

his discounted average payoff from the first deviation belongs to an η-neighbourhood of

$$\frac{(1-\delta)u_1(a',\beta'')+(\delta-\delta^{M-2})U}{1-\delta^{M-2}}, \tag{14}$$

and his discounted average payoff from the second deviation belongs to an η-neighbourhood of

$$\frac{(1-\delta)u_1(a'',\beta'')+(\delta-\delta^{K})u_1(a^*,\beta'')}{1-\delta^{K}} \tag{15}$$

where M is the length of the back loop induced by $\widehat{\sigma}_1$. Since $\widehat{\sigma}_1$ best replies to σ_2, the value of (13) must be weakly greater than the maximum of (14) and (15) plus 2η. Therefore,

$$u_1(a^*,\beta'')+\frac{1-\delta}{1-\delta^K}\underbrace{\Big\{u_1(a'',\beta'')-u_1(a^*,\beta'')\Big\}}_{>0,\text{ since } a''\prec a^* \text{ and } u_1 \text{ is decreasing in } a}-2\eta\leq U$$

$$\leq u_1(a^*,\beta'')+\delta\underbrace{\Big\{u_1(a',\beta^*)-u_1(a',\beta'')\Big\}}_{\leq 0,\text{ since } \beta''\succeq\beta^* \text{ and } u_1 \text{ is increasing in } b}+2\eta. \tag{16}$$

The left-hand-side of (16) is strictly greater than $u_1(a^*,\beta'')$ while the right-hand-side of (16) is strictly smaller than $u_1(a^*,\beta'')$. Therefore, (16) is false for η small enough. When u_1 satisfies Assumption 1 in the main text, the following expression is bounded away from 0:

$$\min_{a',a''\neq a^*,\beta''\succeq_{FOSD}\beta^*}\frac{1}{4}\Big\{\delta(u_1(a',\beta'')-u_1(a',\beta^*))+\frac{1-\delta}{1-\delta^K}(u_1(a'',\beta'')-u_1(a^*,\beta''))\Big\} \tag{17}$$

Pick a small enough $\varepsilon>0$ such that η is strictly less than (17), I can obtain that any pure strategy that induces an ε-back-loop cannot best reply to any of player 2's strategies in which she plays actions in $\mathcal{B}^*$ at every history that occurs with positive probability. □

Online Appendix F: Learning from Coarse Summary Statistics

This appendix studies an extension where the short-run players can only learn from *coarse summary statistics*. Formally, let $A_1\cup...\cup A_n$ be a partition of A. For every $t\in\mathbb{N}$, player 2_t only observes the number of times that player 1's last $\min\{t,K\}$ actions belong to each partition element.

Since there exists a complete order $\succ_A$ on A and $u_1(a,b)$ is strictly decreasing in a, for every partition element A_i, the strategic-type of player 1 will never choose action $a\in A_i$ if there exists $a'\in A_i$ such that $a\succ_A a'$. Hence, analyzing the game under an n-partition $\{A_1,...,A_n\}$ of A is equivalent to analyzing a game where player 1 chooses his action from set $\{\min A_1, ..., \min A_n\}$.

When players' stage-game payoffs satisfy Assumption 1 and inequality (6), player 2 has no incentive to play b^* unless player 1 plays a^* with positive probability. When the prior probability of commitment type π_0 is small enough such that player 2 has no incentive to play b^* when player 1 plays a^* with probability no more than π_0, the strategic-type player 1 has no incentive to play a^* and player 2 has no incentive to play b^* unless the partition element that contains a^* is a singleton. If we partition A according to $A = \{a^*\} \bigcup \Big(A \backslash \{a^*\}\Big)$, then player 2 may receive a higher welfare under some intermediate K. Intuitively, such a partition helps player 1 to credibly commit not to take any action other than his commitment action a^* and his lowest-cost action $\underline{a} \equiv \min A$. This provides player 2 a stronger incentive to punish player 1 after the latter loses his reputation, since player 2 knew that player 1 will take his lowest action as long as he does not take the highest action. This credible threat of punishment motivates player 1 to play a^* in every period.

Proposition. *Suppose players' stage-game payoffs (u_1, u_2) satisfy Assumptions 1 and 2,*

1. *If the partition element that contains a^* is not a singleton, then the discounted frequency with which the strategic-type of player 1 plays a^* is 0 in all Nash equilibria.*

2. *If the partition element that contains a^* is a singleton (without loss of generality, let $A_1 \equiv \{a^*\}$), then when $\delta > \underline{\delta}(\pi_0)$, the strategic-type player 1's payoff is at least $(1-\delta^K)u_1(a^*, \underline{b}) + \delta^K u_1(a^*, b^*)$ in every Nash equilibrium. Furthermore, there exists an integer $\overline{K} \in \mathbb{N}$ such that*

   (i) There exists $C \in \mathbb{R}_+$ that is independent of δ such that for every $1 \leq K < \overline{K}$, we have $F^\sigma(a^*, b^*) \geq 1 - (1-\delta)C$ for every Nash equilibrium σ under K and δ.

   (ii) There exists $\eta > 0$ such that for every $K \geq \overline{K}$, there exists $\underline{\delta} \in (0,1)$ such that for every $\delta > \underline{\delta}$, there exists a PBE such that $\sum_{b \in B} F^\sigma(a^*, b) \leq 1 - \eta$.

The proof of this proposition follows directly from that of Theorems 1 and 2 in the main text, which I omit in order to avoid repetition. The way to compute the cutoff $\overline{K}$ is similar to that in the baseline model. If inequality (6) is violated, then $\overline{K} = 1$ and there exists an equilibrium in which the strategic type plays a^* with zero frequency. If inequality (6) is satisfied, then $\overline{K}$ is the smallest K such that b^* best replies to the mixed action $\frac{K-1}{K} a^* + \frac{1}{K} \min_{j \in \{2,...,n\}} \{\min A_j\}$.

Online Appendix G: Observing the Exact Sequence of Actions

This appendix studies a model in which players' stage-game payoffs satisfy Assumption 1 but player 2 can *perfectly* observe the exact sequence of player 1's last K actions. The main result is stated as follows:

Proposition. *Suppose (u_1, u_2) satisfies Assumption 1 and inequality (6), and that for every $t \in \mathbb{N}$, player 2_t can observe player 1's last $\min\{t, K\}$ actions including the exact sequence of these actions.*

1. *For every $K \geq 1$, there exists $\underline{\delta} \in (0,1)$ such that when $\delta > \underline{\delta}$, there exists a PBE where player 1 obtains payoff $u_1(a^*, b^*)$ and player 2 obtains payoff $u_2(a^*, b^*)$.*

2. *In the following supermodular product choice game,*

-	T	N
H	$1, 1$	$-c_N, x$
L	$1 + c_T, -x$	$0, 0$

with $c_N > c_T > 0$ and $x \in (0,1)$,

there exist $\underline{K} \in \mathbb{N}$, $\overline{\pi} \in (0,1)$, and $\eta > 0$ such that when $\pi_0 \in (0, \overline{\pi})$ and $K \geq \underline{K}$,[2] *for every δ large enough,*[3] *there exists a PBE where player 1's payoff is no more than $u_1(a^*, b^*) - \eta$ and player 2's payoff is no more than $u_2(a^*, b^*) - \eta$.*

Proof of Statement 1: For any (u_1, u_2) that satisfies Assumption 1 and inequality (6), I construct an equilibrium in which (a^*, b^*) occurs with probability 1 at every on-path history. Since a^* is player 1's optimal pure commitment action and $u_1(a, b)$ is strictly increasing in b and is strictly decreasing in a, player 2's best reply to any $a \neq a^*$ is strictly smaller than b^*. Let a' be player 1's lowest action and let b' be player 2's lowest best reply to a'. Since best reply correspondences are upper-hemi-continuous, when δ is close enough to 1, there exist $\lambda \in (0,1)$ and $\beta \in \Delta(B)$ such that

1. $b^* \succ_{FOSD} \beta \succ_{FOSD} b'$,
2. β best replies to the mixed action $\lambda a^* + (1 - \lambda) a'$,
3. $u_1(a', \beta) = (1 - \delta) u_1(a^*, \beta) + \delta u_1(a^*, b^*)$.

This is because when (u_1, u_2) satisfies Assumption 1 and a^* is player 1's optimal pure commitment action, we have $u_1(a', b^*) > u_1(a^*, b^*) > u_1(a', b')$.

Consider the following PBE. At every history h^t such that either $t = 0$ or $t \geq 1$ and $a_{t-1} = a^*$, player 1 plays a^* and player 2 plays b^*. At every history with $t \geq 1$, $a_{t-1} \neq a^*$, and $a_{t-k} \neq a^*$ for every $k \in \{1, 2, ..., K\}$, then player 1 plays the mixed action $\lambda a^* + (1 - \delta) a'$ and player 2 plays β. At

[2] The requirement that π_0 being small enough is necessary for the existence of a low-payoff equilibrium. This is because when π_0 is large enough, player 2 has a strict incentive to play T upon observing $(H, ..., H)$ given that the commitment type plays H in every period, in which case player 1 can secure his commitment payoff by playing H in every period.

[3] As in Liu and Skrzypacz (2014), reputation cycles can occur only if δ is large enough. This is because restoring reputation requires player 1 to play the strictly dominated action H. Hence, he has no incentive to restore his reputation when δ is low.

any other history, player 1 plays a' and player 2 plays b'. Since player 1 takes action a^* in period 0, player 1's action is a^* and player 2's action is b^* at every on-path history. Player 2's action at every history is her myopic best reply, so her incentive constraints are satisfied. At any history, player 1 strictly prefers a' to any action that is neither a^* nor a' since player 2 treats any action that is not a^* as a' and a' leads to a strictly higher stage-game payoff compared to any other action. Player 1 is indifferent between a^* and a' at any history where $t \geq 1$, $a_{t-1} \neq a^*$, and $a_{t-k} \neq a^*$ for every $k \in \{1, 2, ..., K\}$ since $u_1(a', \beta) = (1-\delta)u_1(a^*, \beta) + \delta u_1(a^*, b^*)$. Player 1 strictly prefers a^* to a' when $t = 0$ or when $t \geq 1$ and $a_{t-1} = a^*$ since $u_1(a', \beta) = (1-\delta)u_1(a^*, \beta) + \delta u_1(a^*, b^*)$ and u_1 having strictly increasing differences imply that $u_1(a^*, b^*) > (1-\delta)u_1(a', b^*) + \delta u_1(a', \beta)$, and the right-hand-side is an upper bound on player 1's discounted average payoff after he plays any action that is not a^*. This verifies players' incentive constraints.

Proof of Statement 2: I construct for K large enough, a PBE in which (i) player 1 induces a back loop in equilibrium, (ii) player 1's equilibrium payoff is bounded away from his commitment payoff $u_1(a^*, b^*)$, and (iii) the frequency with which player 1 plays his commitment action a^* being bounded below 1. Consider a strategy profile such that for every $t \geq K$,

1. When $(a_{t-K}, ..., a_{t-1}) = (H, H, ..., H)$, the strategic-type player 1 plays H with probability $y \in (0, x)$ and player 2_t plays T with probability $\beta_H \in (0, 1)$.
2. When $(a_{t-K}, ..., a_{t-1}) = (L, L, ..., L)$, the strategic-type player 1 plays H with probability x and player 2_t plays T with probability $\beta_L \in (0, 1)$.
3. When $(a_{t-K}, ..., a_{t-1}) \in \Big\{(L, ..., L, H), (L, ..., L, H, H), ..., (L, H, ..., H)\Big\}$, the strategic-type player 1 plays H and player 2_t plays T.
4. At all other histories, the strategic-type player 1 plays L and player 2_t plays N.

When $t = 0$, the player 2 plays T with probability β_L and the strategic-type player 1 plays H with probability $\frac{x-\pi_0}{1-\pi_0}$. For every $t \in \{1, 2, ..., K-1\}$, players behave as if actions before period 0 were L.

I pin down β_H and β_L using player 1's incentive constraints. I show that when π_0 is small, there exists $y \in (0, x)$ under which player 2_t is indifferent between T and N when $(a_{t-K}, ..., a_{t-1}) = (H, H, ..., H)$. Player 1's indifference at $(a_{t-K}, ..., a_{t-1}) = (H, H, ..., H)$ implies that:

$$V(H, H, ..., H) = \beta_H(1 + c_N) - c_N = (1-\delta)\beta_H(1 + c_T) + \delta^K \beta_L(1 + c_T). \tag{18}$$

Player 1's indifference at $(a_{t-K}, ..., a_{t-1}) = (L, L, ..., L)$ implies that:

$$V(L, L, ..., L) = r_L(1 + c_T) = (1 - \delta)\Big\{\beta_L(1 + c_N) - c_N\Big\} + (\delta - \delta^K) + \delta^K\Big\{\beta_H(1 + c_N) - c_N\Big\}. \quad (19)$$

Solving this system of linear equations, I obtain:

$$\beta_L\Big\{\frac{(1 + ... + \delta^{2K-1})(1 + c_N)(1 + c_T) - (1 + c_T)^2}{(1 + c_N) - (1 - \delta)(1 + c_T)} - (1 + c_N)\Big\}$$

$$= -c_N + \delta(1 + ... + \delta^{K-2}) + \frac{\delta^K c_N(1 + c_T)}{(1 + c_N) - (1 - \delta)(1 + c_T)}$$

and

$$\beta_H = \frac{\delta^K(1 + c_T)}{(1 + c_N) - (1 - \delta)(1 + c_T)}\beta_L + \frac{c_N}{(1 + c_N) - (1 - \delta)(1 + c_T)}.$$

Since both β_L and β_H are continuous functions of $\delta \in [0, 1]$, as $\delta \to 1$, β_L and β_H converge to

$$\beta_L^* = \frac{c_N(c_T - c_N) + (K - 1)(1 + c_N)}{2K(1 + c_N)(1 + c_T) - (1 + c_T)^2 - (1 + c_N)^2}, \quad (20)$$

$$\beta_H^* = \frac{2Kc_N(1 + c_T) + (K - 1)(1 + c_T) - c_N(c_N + c_T + 2)}{2K(1 + c_N)(1 + c_T) - (1 + c_T)^2 - (1 + c_N)^2}. \quad (21)$$

If $c_N > c_T > 0$, there exists $\underline{K} \in \mathbb{N}$ such that when $K > \underline{K}$, both β_H^* and β_L^* are strictly between 0 and 1.

Given the strategic-type player 1's equilibrium behavior, he plays L from period 0 to $K - 1$ with positive probability. Hence, there exists $\underline{p} > 0$ independent of δ such that conditional on player 1 being the strategic type and $t \geq K$, the probability that $(a_{t-K}, ..., a_{t-1}) = (H, ..., H)$ is more than $\underline{p}$. Hence, when π_0 is small enough, there exists $y \in (0, x)$ such that when the strategic-type of player 1 plays H with probability y when $(a_{t-K}, ..., a_{t-1}) = (H, ..., H)$, player 2_t's belief assigns probability x to $a_t = H$ and hence, has an incentive to mix between T and N. Since it is optimal for the strategic-type player 1 to play L from period 0 to $K - 1$, β_H^* is bounded away from 1 as $\delta \to 1$, and

$$V(L, L, ..., L) \leq (1 - \delta^K) + \delta^K V(H, ..., H) = (1 - \delta^K) + \delta^K\Big(\beta_H(1 + c_N) - c_N\Big),$$

the strategic-type player 1's continuation value in period 0 is bounded away from 1 even as $\delta \to 1$, and the discounted frequency with which he plays H is also bounded away from 1.

Online Appendix H: Monotone-Submodular Payoffs

I relax the assumption that $u_1(a,b)$ has strictly increasing differences while maintaining all other assumptions in my baseline model. I focus on the following *submodular* product choice game:

seller \ consumer	**T**rust	**N**o Trust
High Effort	$1,1$	$-c_N, x$
Low Effort	$1+c_T, -x$	$0,0$

with $c_T \geq c_N > 0$ and $x \in (0,1)$.

where the weakly decreasing difference condition translates into $c_T \geq c_N$.

Proposition. *In the product choice game where $u_1(a,b)$ has weakly decreasing differences.*

1. *For every $K \geq 1$ and $c_T \geq c_N > 0$, there exists $\underline{\delta} \in (0,1)$ such that for every $\delta > \underline{\delta}$, there exist σ_2 and a pure strategy $\widehat{\sigma}_1$ that best replies to σ_2 such that the no-back-loop property fails under $(\widehat{\sigma}_1, \sigma_2)$.*

2. *Suppose c_T is large enough such that $1 + c_T > K(1 + c_N)$. There exist $\underline{\delta} \in (0,1)$, $\overline{\pi}_0 > 0$, and $\eta > 0$ such that for every $\delta > \underline{\delta}$ and $\pi_0 < \overline{\pi}_0$, there exists a PBE where player 1's payoff is lower than $u_1(a^*, b^*) - \eta$ and the discounted frequency with which he plays a^* is no more than $1 - \eta$.*

Proof. For every $k \in \{0,..,K\}$, let β_k be player 2's mixed action when k of the last K actions were L. Player 2's strategy σ_2 is characterized by $(\beta_0, ..., \beta_K)$ and her behavior in the first K periods. The latter is irrelevant for the proof of the first statement but is relevant for the proof of the second statement. I use $u_1(a, \beta)$ to denote player 1's stage-game payoff when his action is a and player 2 plays T with probability β.

Proof of Statement 1: I show that for every $u_1(a,b)$ that has weakly decreasing differences, there exist a strategy for player 2, σ_2, as well as a pure strategy for player 1, $\widehat{\sigma}_1$, such that $\widehat{\sigma}_1$ best replies to σ_2 and $\widehat{\sigma}_1$ induces a back loop. I consider two cases separately, depending on the relative magnitude of c_T and c_N.

First, I consider the case in which $c_T \geq K(1+c_N)$. Let σ_2 be such that $\beta_0 = 1$ and $\beta_1 = ... = \beta_K = 0$. In the first step, I show that at history $(H, H, ..., H)$, it is optimal for player 1 to play L since his discounted average payoff from doing so is at least

$$\frac{u_1(L, \beta_0) + (\delta + ... + \delta^K) u_1(H, \beta_1)}{1 + \delta + ... + \delta^K} \tag{22}$$

which is attained if he plays L and then plays H in the next K periods. For every $\delta \in (0,1)$, the value of (22) is strictly greater than $\frac{1+c_T-Kc_N}{K+1}$. Player 1's payoff from playing H at $(H,H,...,H)$ is 1. Since

$c_T \geq K(1+c_N)$, we have $\frac{1+c_T-Kc_N}{K+1} \geq 1$, which implies that the value of (22) is strictly greater than 1. In the second step, I show that at any state $(a_{t-K},...,a_{t-1}) \neq (H,H,...,H)$, it is optimal for player 1 to return to state $(H,...,H)$ in finite time. This is because player 1's stage-game payoff when $(a_{t-K},...,a_{t-1}) \neq (H,H,...,H)$ is at most 0, which is strictly less than his payoff when he plays H at $(H,H,...,H)$. Hence when δ is large enough, player 1 has an incentive to return to $(H,...,H)$ in finite time at every $(a_{t-K},...,a_{t-1}) \neq (H,H,...,H)$. This implies that player 1 has a pure-strategy best reply to σ_2 that violates the no-back-loop property.

Next, I consider the case in which $c_T < K(1+c_N)$. Let σ_2 be such that $\beta_k = 0$ for every $k > 1$ and β_0, β_1 satisfy

$$(\delta + ... + \delta^K)(\beta_0 - \beta_1)(1+c_N) = \beta_0 c_T + (1-\beta_0)c_N. \tag{23}$$

When $c_T < (K+1)c_N$ and δ is close to 1, there exist $0 < \beta_1 < \beta_0 < 1$ that satisfy (23). It is weakly optimal for player 1 to play L at $(H,...,H)$. This is because player 1's discounted average payoff is $u_1(H,\beta_0)$ if he plays H at $(H,...,H)$, and his discounted average payoff is

$$\frac{u_1(L,\beta_0) + (\delta + ... + \delta^K)u_1(H,\beta_1)}{1+\delta+...+\delta^K} \tag{24}$$

if he plays L at $(H,...,H)$ and then plays H for K consecutive periods. The value of (24) equals $u_1(H,\beta_0)$ under (23), which verifies player 1's incentive to play L at $(H,...,H)$. Using the same argument as in the first case, when δ is large enough, player 1 has an incentive to return to either $(H,...,H)$ or histories where only one of the last K actions were L in finite time. What remains to be verified is that player 1 has an incentive to play H at histories where only one of the last K actions was L. Since $\beta_2 = ... = \beta_K = 0$, player 1's incentive constraint is the tightest when L occurred K periods ago. At history $(L,H,...,H)$, player 1's discounted average payoff from playing H is

$$\frac{u_1(H,\beta_1) + \delta u_1(L,\beta_0) + (\delta^2 + ... + \delta^K)u_1(H,\beta_1)}{1+\delta+...+\delta^K}, \tag{25}$$

and his discounted average payoff from playing L and then playing H for $K-1$ consecutive periods is

$$\frac{u_1(L,\beta_1) + (\delta + ... + \delta^{K-1})u_1(H,\beta_1)}{1+\delta+...+\delta^{K-1}}. \tag{26}$$

The value of (25) is greater than (26) if and only if

$$(\delta + ... + \delta^K)(\beta_0 - \beta_1)(1+c_T) \geq \beta_1 c_T + (1-\beta_1)c_N. \tag{27}$$

Since $c_T \geq c_N$ and $\beta_1 < \beta_0$, we know that $\beta_1 c_T + (1-\beta_1)c_N \leq \beta_0 c_T + (1-\beta_0)c_N$. Moreover,

$$(\delta + ... + \delta^K)(\beta_0 - \beta_1)(1 + c_N) \leq (\delta + ... + \delta^K)(\beta_0 - \beta_1)(1 + c_T).$$

Therefore, inequality (23) implies (27), which verifies that playing L at $(H, ..., H)$ and then playing H for K consecutive periods is player 1's best reply to σ_2, and such a best reply induces a back loop.

Proof of Statement 2: I consider three cases separately. First, I study the case in which $K = 1$. Next, I study the case in which $K \geq 2$ and $x \leq 1/2$. Then, I study the case in which $K \geq 2$, $x > 1/2$, and c_T is large enough relative to c_N in the sense that $1 + c_T > K(1 + c_N)$.

Case 1: $K = 1$ In period 0, player 1 plays L and player 2_0 plays N. For every $t \geq 1$, (i) if $a_{t-1} = L$, player 1 mixes between H and L and player 2_t plays N, and (ii) if $a_{t-1} = H$, player 1 plays L and player 2_t plays T with probability $\frac{c_N}{\delta(1+c_T)}$. Under this strategy profile, player 1's continuation value satisfies $V(L) = (1-\delta)u_1(L, N) + \delta V(L)$ and $V(H) = (1-\delta)\frac{c_N}{\delta(1+c_T)}u_1(L, T) + (1-\delta)(1 - \frac{c_N}{\delta(1+c_T)})u_1(L, N) + \delta V(L)$, which implies that $V(L) = 0$ and $V(H) = \frac{1-\delta}{\delta}c_N$. Player 1 is indifferent between H and L when $a_{t-1} = L$ since $(1-\delta)u_1(L, N) + \delta V(L) = (1-\delta)u_1(H, N) + \delta V(H)$. Player 1 prefers L to H when $a_{t-1} = H$ since $c_T \geq c_N > 0$ and player 2 plays T with higher probability when $a_{t-1} = H$. Player 2 has an incentive to play N when $a_{t-1} = L$ since $\frac{\pi_0}{\delta(1-\pi_0)-\pi_0} \leq \frac{1}{2}$. I claim that there exists a mixing probability of player 1 at $a_{t-1} = L$ that makes player 2 indifferent between T and N when $a_{t-1} = H$. This is because the commitment type plays H at $(H, ..., H)$ and the strategic type plays L at $(H, ..., H)$, such a mixing probability exists as long as π_0 is not too large.

Case 2: $K \geq 2$ **and** $x \leq 1/2$ From period 0 to $K-1$, player 2 plays T in period 0 as well as at histories where L has never occurred before, and she plays N otherwise. From period 0 to $K-1$, player 1 plays H on the equilibrium path and plays L at histories where L has occurred at least once before.

Starting from period K, the strategic type of player 1 mixes between $K+1$ pure strategies, indexed by $k \in \{0, 1, ..., K\}$, such that strategy with index k is described as follows:

- Player 1 plays L at any history where H has occurred weakly more than k times in the last K periods. At any history where H occurred strictly fewer than k times in the last K periods, player 1 plays H if $(a_{t-K}, ..., a_{t-1}) = (L, L, ..., L, H, ..., H)$, i.e., all L occurred before H in the last K periods, and plays L otherwise.

Under each of these pure strategies, player 1 plays L from period K to $2K-1$. Starting from period $2K$,

- If player 1 follows strategy 0, then he plays L in every period, and remains in state $(L, L, .., L)$.
- If player 1 follows strategy $k \in \{1, 2, ..., K\}$, then he plays H for k periods, plays L for K periods, plays H for k periods, and so on.

Note that only strategy K violates the no-back-loop property. The probabilities with which player 1 play these strategies are pinned down by player 2's indifference conditions: Player 2 is indifferent between T and N at all histories after period K except for history $(L, ..., L)$. Player 2 being indifferent between T and N at history $(H, ..., H)$ together with the probability of commitment type π_0 pins down the probability of strategy K, since conditional on player 1 being the strategic type, $(H, ..., H)$ occurs on the equilibrium path after period $2K$ only if player 1 uses strategy K, and he plays L at $(H, ..., H)$ under that strategy. Player 2 being indifferent between T and N at histories where L occurred once and the probability of strategy K pins down the probability of strategy $K-1$, since the probability of H at those histories is strictly greater than $1/2$ under strategy K, the probability of H at those histories is close to 0 under strategy $K-1$, and $x < 1/2$.... Iterate this procedure K times, we obtain the probabilities of strategy K to strategy 1. Notice that the probabilities of strategy 1 to K, which are pinned down by player 2's indifference conditions, are decreasing in π_0 and converge to 0 as $\pi_0 \to 0$. Hence, for π_0 small enough, the probability of strategy 0 is large enough so that player 2 has a strict incentive to play N at $(L, .., L)$.

Since player 2 has a strict incentive to play N at $(L, ..., L)$, we have $\beta_K = 0$. I use player 1's indifference conditions to pin down $(\beta_0, ..., \beta_K)$, i.e., the probabilities with which player 2 plays T after period K. If player 1 mixes between these $K+1$ strategies, then his payoff starting from $(L, ..., L)$ is 0 no matter which strategy he uses, and moreover, he is indifferent between H and L at histories $(L, ..., L)$, $(L, ..., L, H)$,...,$(L, H, ..., H)$. Player 1 being indifferent between strategy 0 and strategy 1 implies that

$$0 = u_1(H, \beta_K) + \sum_{j=1}^{K} u_1(L, \beta_{K-1}). \tag{28}$$

For every $k \geq 1$, player 1 being indifferent between strategy 0 and strategy k implies that

$$0 = \sum_{j=0}^{k-1} \delta^j u_1(H, \beta_{K-j}) + \sum_{j=k}^{K} \delta^j u_1(H, \beta_{K-k}) + \sum_{j=K+1}^{K+k} \delta^j u_1(H, \beta_{j-k}). \tag{29}$$

Instead of solving (28) and (29), I solve the following auxiliary system of linear equations:

$$c_N = K u_1(L, \beta^*_{K-1}), \tag{30}$$

$$c_N = (K-k)u_1(L, \beta^*_{K-k-1}) + \sum_{j=K-k}^{K-1} \Big(u_1(L, \beta^*_j) + u_1(H, \beta^*_j) \Big) \text{ for every } k \in \{1, ..., K-1\}. \quad (31)$$

Later I show that as $\delta \to 1$, the unique solution of (30) and (31) is the unique limit point of the solution to (28) and (29). It is easy to see that linear system (30) and (31) has a unique solution $(\beta^*_{K-1}, ..., \beta^*_0)$. Comparing (30) to (31) when $k = 1$, and comparing (31) when $k = j$ to $k = j+1$ for every $j \in \{1, 2, ..., K-2\}$, we obtain

$$\beta^*_{K-j-1} - \beta^*_{K-j} = \frac{c_N - (1+c_N)\beta_{K-j}}{(1+c_T)(K-j)} \text{ for every } j \in \{1, 2, ..., K-1\}, \quad (32)$$

from which we obtain that $\beta^*_{K-j-1} - \beta^*_{K-j} > 0$ if and only if $c_N - (1+c_N)\beta^*_{K-j} > 0$, or equivalently $u_1(H, \beta^*_{K-j}) < 0$. I show that $\beta^*_{K-1} < \beta^*_{K-2} < ... < \beta^*_1 < \beta^*_0 < 1$ by showing that all of them are strictly smaller than $\frac{c_N}{1+c_N}$. First, (30) implies that $\beta^*_{K-1} = \frac{c_N}{K(1+c_T)}$, and given that $c_T \geq c_N$ and $K \geq 2$, we have $\frac{c_N}{K(1+c_T)} < \frac{c_N}{1+c_N}$. Suppose by way of contradiction that there exists k such that $\beta^*_k \geq \frac{c_N}{1+c_N}$. Without loss of generality, let k be the largest $\widetilde{k} \in \mathbb{N}$ such that $\beta^*_{\widetilde{k}} \geq \frac{c_N}{1+c_N}$. Then it must be the case that $\beta^*_{k+1} \in (0, \frac{c_N}{1+c_N})$ and

$$\beta^*_{k+1} + \frac{c_N - (1+c_N)\beta_{k+1}}{(1+c_T)(k+1)} \geq \frac{c_N}{1+c_N}. \quad (33)$$

Since the LHS of (33) is linear in β^*_{k+1}, it is either maximized when $\beta^*_{k+1} = 0$, in which case its value is $\frac{c_N}{(k+1)(1+c_T)}$, or maximized when $\beta^*_{k+1} = \frac{c_N}{1+c_N}$, in which case its value is $\frac{c_N}{1+c_N}$. Since $\frac{c_N}{(k+1)(1+c_T)} < \frac{c_N}{1+c_N}$, we know that whenever $\beta^*_{k+1} < \frac{c_N}{1+c_N}$, it must be the case that $\beta^*_k < \frac{c_N}{1+c_N}$. Equation (32) then implies that $0 < \beta^*_{K-1} < \beta^*_{K-2} < ... < \beta^*_1 < \beta^*_0 < \frac{c_N}{1+c_N}$.

As $\delta \to 1$, $(\beta_0, ..., \beta_{K-1})$ that solves (28) and (29) converges to $(\beta^*_0, ...\beta^*_{K-1})$. First, β_{K-1} converges to β^*_{K-1} as $\delta \to 1$. This is because β^*_{K-1} and β_{K-1} are pinned down by (30) and (28), respectively, and all the linear coefficients in (28) converge to those in (30) as $\delta \to 1$. Plugging in the values of β^*_{K-1} and β_{K-1} into equations (31) and (29) when $k = 1$, they become linear equations of β^*_{K-2} and β_{K-2}, which pin down the values of β^*_{K-2} and β_{K-2}. Since all the linear coefficients in (29) converge to those in (31) as $\delta \to 1$, β_{K-2} converges to β^*_{K-2}... Iterate this procedure, we obtain that $(\beta_0, ..., \beta_{K-1})$ converges to $(\beta^*_0, ...\beta^*_{K-1})$ as $\delta \to 1$. Hence, $0 < \beta_{K-1} < \beta_{K-2} < ... < \beta_1 < \beta_0 < \frac{c_N}{1+c_N}$ for every δ close enough to 1.

Player 1's incentive constraints are satisfied since the construction of $\beta_0, ..., \beta_K$ implies that player 1 is indifferent between strategies 0 to K, and that he is indifferent between H and L at histories where in the last K periods, there is at least one L and every H occurred after all the L. Furthermore, $\beta_0 < \frac{c_N}{1+c_N}$ implies that player 1 strictly prefers these strategies to playing H in every period. At every history where L occurred at least once and some H occurred before some L, player 1's incentive to play H is weaker compared to a history where the number of L is the same but all H occurred after L. This is because β_k is strictly decreasing in k and playing H at the latter history lowers the index of future histories in FOSD.

Case 3: $K \geq 2$**,** $x > 1/2$**, and** $1 + c_T > K(1 + c_N)$ From period 0 to $K - 1$, player 2 plays T in period 0 as well as at histories where L has never occurred before, and she plays N otherwise. From period 0 to $K - 1$, player 1 plays H on the equilibrium path and plays L at histories where L has occurred at least once before. After period K, player 2 plays N unless the history is $(H, H, ..., H)$, that is $\beta_1 = ... = \beta_K = 0$ and $\beta_0 \in (0, 1)$. Starting from period K, player 1 mixes between the following two pure strategies:

> **Strategy 1:** Plays L in every subsequent period.
>
> **Strategy 2:** Plays L at $(H, H, ..., H)$, plays H at $(L, ..., L)$, at other histories, plays L if his action one period ago was L, and plays H if his action one period ago was H.

The probability with which he uses these two strategies are pinned down by player 2's indifference condition: Player 2 is indifferent between T and N upon observing all of player 1's last K actions being H. Since the commitment type plays H at $(H, ..., H)$ and the strategic type plays L at $(H, ..., H)$, such a probability exists as long as π_0 is not too large, and the probability of the first strategy converges to 1 as $\pi_0 \to 0$. A useful observation is that given player 2's strategy, starting from period K, player 1's incentive at any history h^t with $a_{t-1} = L$ coincides with his incentive at history $(L, ..., L)$. This is because $\beta_1 = ... = \beta_K = 0$, i.e., player 2 plays T with positive probability only if all of player 1's last K actions were H.

I use player 1's indifference condition at history $(L, ..., L)$ to pin down β_0. Player 1's payoff if he plays L in every period is 0 and player 1's payoff if he follows his equilibrium strategy is

$$\frac{u_1(H, 0)(1 + \delta + ... + \delta^{K-1}) + u_1(L, \beta_0)\delta^K}{1 + \delta + ... + \delta^K}.$$

If player 1 is indifferent, then the above payoff is 0, which yields

$$\beta_0 = \frac{(1 + \delta + ... + \delta^{K-1})c_N}{(1 + c_T)\delta^K} < 1, \tag{34}$$

where strict inequality comes from the requirement that player 1's equilibrium payoff being bounded away from $u_1(a^*, b^*)$. Another constraint is that player 1 has an incentive to play L at history $(H, ..., H)$, which yields:

$$\frac{u_1(L, \beta_0) + (\delta + ... + \delta^K)u_1(H, 0)}{1 + \delta + ... + \delta^K} \geq u_1(H, \beta_0). \tag{35}$$

In order to satisfy both (34) and (35) for all δ close to 1, we need

$$\frac{Kc_N}{1+c_T} < \min\Big\{1, \frac{c_N}{(K+1)(1+c_N)-(1+c_T)}\Big\},$$

which is the case if and only if $1+c_T > K(1+c_N)$. Player 2's incentive constraint is satisfied since (i) at every history before period K, player 1's action is H with probability 1 so player 2 has an incentive to play T, (ii) player 2 has an incentive to play N at $(L,...,L)$ when π_0 is small enough since the probability that player 1 plays L in every period is large enough, (iii) player 2 has an incentive to mix between T and N due to the construction of player 1's mixing probabilities, and (iv) player 2 has an incentive to play N at every history after period K with both H and L is implied by $x > 1/2$, since when δ is close to 1, player 2 assigns probability close to $1/2$ to player 1's last period action being L after observing any such history. $\square$

Online Appendix I: Sequential-Move Stage Games

In Online Appendix I.1, I extend the no-back-loop lemma to the case where player 2 can observe a noisy private signal about player 1's current-period action before choosing her action. In Online Appendix I.2, I consider the case in which player 1 can observe player 2's current-period action before choosing his action. Focusing on the case in which $K=1$, I fully characterize the patient player's equilibrium payoff set.

Online Appendix I.1: Noisy Private Signal about Player 1's Current-Period Action

This appendix extends the no-back-loop lemma when player 2_t observes a noisy private signal $y_t \sim F(\cdot|a_t)$ about a_t before choosing b_t. In this setting, player 2's action distribution depends not only on the summary statistics of player 1's last K actions but also on player 1's action in the current period. Let $\beta^*(a) \in \Delta(B)$ denote player 2's action distribution when the history belongs to $\mathcal{H}_1^*$ and player 1's current-period action is a. Let $\beta'(a) \in \Delta(B)$ denote player 2's action distribution when player 1's current-period action is a and there is one a' in the last K periods and the rest of the last K actions are a^*. Let $\beta''(a) \in \Delta(B)$ denote player 2's action distribution when player 1's current-period action is a and there is one a'' in the last K periods and the rest of the last K actions are a^*. In order to capture that the private signal about player 1's current-period action is noisy, I assume that there exists $\eta > 0$ such that $F(y|a) > \eta$ for every $y \in Y$ and $a \in A$. I show that for every $\eta > 0$, there exists $\varepsilon > 0$ such that the no-back-loop lemma holds for every distribution over private signals that satisfies $||F(\cdot|a) - F(\cdot|a')||_{TV} \leq \varepsilon$ for every $a, a' \in A$.

Without loss of generality, I normalize player 1's stage-game payoff such that it always belongs to the interval $[0,1]$. Under this normalization, for every $a_0, a_1, a_2 \in A$, $||F(\cdot|a) - F(\cdot|a')||_{TV} \leq \varepsilon$ for every $a, a' \in A$ implies that for every $a_0, a_1, a_2 \in A$, we have $|u_1(a_0, \beta^*(a_1)) - u_1(a_0, \beta^*(a_2))| \leq \varepsilon$, $|u_1(a_0, \beta'(a_1)) - u_1(a_0, \beta'(a_2))| \leq \varepsilon$ and $|u_1(a_0, \beta''(a_1)) - u_1(a_0, \beta''(a_2))| \leq \varepsilon$.

Suppose by way of contradiction that there exists $\beta : \mathcal{H} \rightarrow \mathcal{B}^*$ and a pure-strategy best reply $\widehat{\sigma}$ to β such that $\widehat{\sigma}$ induces a back loop. Following the notation and the definitions in Appendix A, player 1's incentive to play a^* at $(a'', a^*, ..., a^*)$ and to play a' at $(a^*, ..., a^*)$ imply that

$$u_1(a', \beta^*(a')) - u_1(a^*, \beta^*(a^*)) \geq \frac{\delta(V^* - V')}{1-\delta} \geq u_1(a', \beta''(a')) - u_1(a^*, \beta''(a^*)). \tag{36}$$

Let $x \equiv s - t - 1$. Player 1 prefers Strategy $*$ to Deviation A, which implies that:

$$(\delta - \delta^{x-1})U + (1-\delta^x)u_1(a', \beta''(a')) \leq (1-\delta^{x-1})u_1(a^*, \beta''(a^*)) + (\delta - \delta^x)u_1(a', \beta^*(a')). \tag{37}$$

Player 1 prefers Strategy $*$ to Deviation B, which implies that:

$$u_1(a^*, \beta''(a^*)) + \underbrace{\frac{1-\delta}{1-\delta^K}\Big\{u_1(a'', \beta''(a'')) - u_1(a^*, \beta''(a^*))\Big\}}_{\text{strictly greater than some } \Delta > 0 \text{ when } \varepsilon \text{ is small}}$$

$$\leq \frac{1-\delta}{1-\delta^x}u_1(a^*, \beta''(a^*)) + \frac{(1-\delta)\delta}{1-\delta^x}u_1(a', \beta^*(a')) + \frac{\delta^2 - \delta^x}{1-\delta^x}U. \tag{38}$$

Focusing on the case in which ε is small, inequalities (37) and (38) together imply that

$$u_1(a', \beta^*(a')) - u_1(a', \beta''(a')) \geq \Delta, \tag{39}$$

for some $\Delta > 0$, and inequality (36) implies that

$$\Big\{u_1(a', \beta^*(a')) - u_1(a', \beta''(a'))\Big\} - \Big\{u_1(a^*, \beta^*(a')) - u_1(a^*, \beta''(a'))\Big\} \geq -2\varepsilon. \tag{40}$$

I show that when (u_1, u_2) satisfies Assumptions 1 and 2 in the main text, $\eta > 0$, and ε is small, $\beta^*(a')$ and $\beta''(a')$ can be ranked according to FOSD. Assumption 2 implies that for every $\alpha \in \Delta(A)$, there exist at most 2 pure-strategy best replies to α. If b and b' are both pure-strategy best replies to α, then any action that is strictly between b and b' does not best reply to any of player 1's mixed actions. Let

$$B^* \equiv \Big\{b \in B \Big| \text{ there exists } \alpha \in \Delta(A) \text{ such that } b \text{ best replies to } \alpha\Big\}. \tag{41}$$

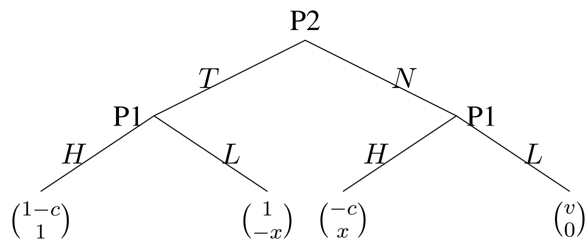

Figure 1: Sequential-Move Supermodular Product Choice Game

For any prior belief about player 1's current-period action $\alpha \in \Delta(A)$, when ε is small relative to η, the likelihood ratio $\frac{F(y|a)}{F(y|a')}$ is close to 1 for every $a, a' \in A$ and $y \in Y$. Hence, if player 2 has a unique best reply $b \in B$ to α, then she has a strict incentive to play b after observing any private signal y. If player 2 has two best replies $b, b' \in B$ to α, then regardless of the private signal she observes, either she strictly prefers to play b, or strictly prefers to play b', or is indifferent between b and b'. Therefore, $\beta^*(a')$ is either a Dirac measure on some $b \in B^*$, or assigns positive probability to two adjacent elements in B^*. The same applies to $\beta''(a')$. Therefore, $\beta^*(a')$ and $\beta''(a')$ can be ranked according to FOSD.

Sine $u_1(a, b)$ is strictly increasing in b, (39) implies that $\beta^*(a')$ strictly FOSDs $\beta''(a')$ and that $||\beta^*(a') - \beta''(a')||_{TV} \geq \Delta$. Since $u_1(a, b)$ has strictly increasing differences, the conclusion I just obtained implies that the LHS of inequality (40) is less than something proportional to $-\Delta$. This leads to a contradiction when ε is small relative to Δ, which implies that player 1's best reply cannot induce any back loop.

Online Appendix I.2: Player 1 Observing Player 2's Current-Period Action

I study a repeated sequential-move supermodular product choice game depicted in Figure 1 with $v \in (0, 1 - c)$, $c > 0$, and $x \in (0, 1)$. With probability $\pi_0 \in (0, 1)$, player 1 is a commitment type who plays H at every history, and with complementary probability, he is a strategic type who maximizes his discounted average payoff. Player 1's *commitment payoff* is $1 - c$ and his minmax value is v.

I focus on the case in which player 2 only observes player 1's action in the period before, that is, $K = 1$. Let $\Delta \equiv \frac{c}{1-v}$. Let

$$\pi^* \equiv \frac{x^2(1 - v - c)}{1 - v - cx} = \frac{x^2(1 - \Delta)}{1 - \Delta x}.$$

Let

$$u(\pi_0) \equiv \min \Big\{ v + \frac{\pi_0(1 - x)(1 - v)}{x(x - \pi_0)}, 1 \Big\}.$$

One can verify that first, $\pi^* < x$, second, $u(\pi_0) = 1 - c$ if and only if $\pi_0 \geq \pi^*$, and third, $u(\pi_0) = v$ when

$\pi_0 = 0$. I characterize the set of payoffs player 1 can attain in equilibrium as $\delta \to 1$.

Proposition. *Fix any $\pi_0 \in (0,1)$.*

1. *For every $u \in [u(\pi_0), 1-c]$ and $\varepsilon > 0$, there exists $\delta^* \in (0,1)$ such that for every $\delta > \delta^*$, there exists a PBE in which player 1's payoff is within an ε-neighbourhood of u.*

2. *For every $u \notin [u(\pi_0), 1-c]$ and $\varepsilon > 0$, there exists $\delta^* \in (0,1)$ such that for every $\delta > \delta^*$, there is no Nash equilibrium in which player 1's payoff is within an ε-neighbourhood of u.*

This proposition characterizes the patient player's limiting equilibrium payoff set in a sequential-move reputation game with limited memories. It shows that any payoff between $u(\pi_0)$ and the commitment payoff $1-c$ can be attained in a PBE as $\delta \to 1$ and any payoff which does not belong to that interval cannot be attained in any Nash equilibrium when player 1 is sufficiently patient.[4]

It characterizes the patient player's *value* from having a good reputation by showing that his *lowest* equilibrium payoff is *an increasing linear function of his reputation*. It also implies that player 1 can secure his commitment payoff in all equilibria if and only if $\pi_0 \geq \pi^*$. Since x is the minimal probability with which player 1 needs to play H in order to provide player 2 an incentive to play T. If player 2 receives no information about player 1's past behavior, i.e., $K = 0$, then player 1 can secure his commitment payoff if and only if $\pi_0 > x$. The fact that $\pi^* < x$ implies that player 2's ability to observe player 1's action in the period before helps player 1 to secure a high payoff.

I discuss one aspect of my result before presenting the proof. First, unlike the case in which players move simultaneously in the stage game, allowing the patient player to observe the short-run player's current-period action (before choosing his own action) may lead to equilibria where both players receive low payoffs. This is interesting since one may expect that when actions are complements, allowing player 1 to observe player 2's action can help players to coordinate on the good outcome (H, T) due to the usual logic that sequential-move helps to solve coordination failure problems and can select equilibria with high payoffs.

The intuition is that due to the complementarity in players' actions, observing player 2's current-period action encourages player 1 to *rebuild* his reputation after milking it. In particular, player 1 will have a strong incentive to play H after observing player 2 has played T in the current period, regardless of his actions in the past. The possibility of rebuilding his reputation after milking it increases the chances that the strategic type of player 1 having a clean history, which raises player 2's suspicion after she observes a clean history.

[4] In order to make the result stronger, I use a demanding solution concept PBE for the result that establishes the existence of a certain type of equilibrium, and use a weak solution concept Nash equilibrium for the result that establishes the non-existence of a certain type of equilibrium.

Proof. Since player 1 can observe b_t before choosing a_t, he chooses a_t in order to maximize $(1-\delta)u_1(a_t, b_t) + \delta V(a_t)$, where $V(a_t)$ denotes his continuation value when his action in the period before was a_t. Since player 1's objective function does *not* depend on player 1's action in the period before, his incentive at every history depends only on player 2's current-period action. Since $u_1(a, b)$ has strictly increasing differences, player 1 has a stronger incentive to play H when player 2's action in the current period was T.

I consider three types of equilibria. In every equilibrium where player 1 strictly prefers to play H when player 2 played T in the current period, player 2 strictly prefers to play T at every on-path history of every Nash equilibrium and at every history in every PBE, since doing so leads to her highest feasible payoff 1. If player 2 plays T at every history, player 1 has no incentive to play H at any history, which contradicts the hypothesis that player 1 has a strict incentive to play H and implies that there is no such PBE. However, there exists a Nash equilibrium in which player 1 strictly prefers to play H when player 2 played T in the current period, from which player 1 obtains his commitment payoff $1 - c$.

In every equilibrium where player 1 strictly prefers to play L when player 2 played T in the current period, player 1 has a strict incentive to play L when player 2 played N in the current period given that $u_1(a, b)$ has strictly increasing differences. As a result, player 2 strictly prefers to play N at every history. However, according to Bayes rule, player 2 will learn that player 1 is the commitment type after observing his last period action being H and therefore, will have a strict incentive to play T. This leads to a contradiction which rules out such equilibria.

The rest of the proof focuses on equilibria in which player 1 is indifferent between H and L when player 2's current-period action was T. Since $u_1(a, b)$ has strictly increasing differences, player 1 strictly prefers L when player 2's current-period action was N. Let p_H and p_L be player 2's probabilities of playing T when player 1's last period action was H and L, respectively. Player 1's continuation values satisfy:

$$V(H) = p_H\Big\{(1-\delta)u_1(H,T) + \delta V(H)\Big\} + (1-p_H)\Big\{(1-\delta)u_1(L,N) + \delta V(L)\Big\}, \tag{42}$$

$$V(L) = p_L\Big\{(1-\delta)u_1(H,T) + \delta V(H)\Big\} + (1-p_L)\Big\{(1-\delta)u_1(L,N) + \delta V(L)\Big\}, \tag{43}$$

and

$$(1-\delta)u_1(H,T) + \delta V(H) = (1-\delta)u_1(L,T) + \delta V(L). \tag{44}$$

Equation (44) implies that $V(H) - V(L) = \frac{1-\delta}{\delta}c$. Subtracting (43) from (42), we obtain that

$$V(H) - V(L) = (p_H - p_L)\Big\{(1-\delta)u_1(L,T) + \delta V(L) - (1-\delta)u_1(L,N) - \delta V(L)\Big\},$$

which is equivalent to

$$p_H - p_L = \frac{c}{(1-v)\delta} = \frac{\Delta}{\delta}. \tag{45}$$

Since $v \in (0, 1-c)$, the RHS of (45) is strictly less than 1 when δ is close enough to 1. When $\delta \to 1$, player 1's equilibrium payoff depends only on p_H and p_L, which according to (45), converges to

$$p_H(1-c) + (1-p_H)(v-c). \tag{46}$$

I examine player 2's incentives and provide necessary and sufficient conditions under which there exist equilibria where player 2's strategy is parameterized by (p_H, p_L). Let q_H be the strategic type's probability of playing H when player 2's current-period action was T and his action in the period before was H. Let q_L be the strategic type's probability of playing H when player 2's current-period action was T and his action in the period before was L. Fix players' strategies $\{p_H, p_L, q_H, q_L\}$. For every $a \in \{H, L\}$, let μ_a be the probability that the last action was a conditional on player 1 being the strategic type and calendar time being at least 1. Lemma B.1 in the main text implies that there exists $y \in [0, 1]$ such that

$$\mu_H = (1-\delta)y + \delta\Big\{\mu_H p_H q_H + (1-\mu_H)p_L q_L\Big\}. \tag{47}$$

Recall that $\Delta \equiv \frac{c}{1-v}$. In equilibrium, it must be the case that $p_L = p_H - \frac{\Delta}{\delta}$ and $q_L = x$. This implies that as $\delta \to 1$, we have

$$\mu_H = \frac{(p_H - \Delta)x}{1 - p_H q_H + (p_H - \Delta)x}. \tag{48}$$

When player 1's last period action was H, player 2 has an incentive to mix between T and N only if

$$\frac{\pi_0 + (1-\pi_0)\mu_H q_H}{\pi_0 + (1-\pi_0)\mu_H} = x,$$

or equivalently,

$$(1-x)\pi_0 = \mu_H(1-\pi_0)(x - q_H). \tag{49}$$

Plugging in the expression for μ_H in (48), we know that when δ is close enough to 1, there exists an equilibrium with payoff $p_H - c$ if and only if $p_H > \Delta$ and there exists $q_H \in [0, x)$ such that

$$(1-x)\pi_0 = (1-\pi_0)(x - q_H)\frac{(p_H - \Delta)x}{1 - p_H q_H + (p_H - \Delta)x}.$$

Since the RHS is strictly decreasing in q_H and is 0 when $q_H = x$, such a q_H exists if and only if

$$(1-x)\pi_0 \leq (1-\pi_0)x\frac{(p_H-\Delta)x}{1+(p_H-\Delta)x}$$

or equivalently

$$\pi_0 \leq \frac{x^2(p_H-\Delta)}{(1-x)+x(p_H-\Delta)}. \tag{50}$$

Since the RHS of (50) is strictly increasing in p_H, we know that when $\pi_0 \geq \frac{x^2(1-\Delta)}{1-\Delta x}$, player 2 has a strict incentive to play T when player 1's action in the period before was H. As a result, $p_H = 1$ and player 1's payoff is close to his commitment payoff $1-c$ in all equilibria as $\delta \to 1$.

For every $\pi_0 < \frac{x^2(1-\Delta)}{1-\Delta x}$, the highest p_H that can be sustained in equilibrium is 1 and the *lowest* p_H that can be sustained in equilibrium is

$$\Delta + \frac{\pi_0(1-x)}{x(x-\pi_0)}.$$

Plugging this expression into the expression for player 1's equilibrium payoff as a function of p_H, given by (46), we obtain that player 1's highest equilibrium payoff is his commitment payoff $1-c$ and his lowest equilibrium payoff is $v + \frac{\pi_0(1-x)(1-v)}{x(x-\pi_0)}$.

□

References

[1] Fudenberg, Drew and David Levine (1989) "Reputation and Equilibrium Selection in Games with a Patient Player," *Econometrica*, 57, 759-778.

[2] Fudenberg, Drew and Jean Tirole (1991) "Perfect Bayesian Equilibrium and Sequential Equilibrium," *Journal of Economic Theory*, 53, 236-260.

[3] Li, Yingkai and Harry Pei (2021) "Equilibrium Behaviors in Repeated Games," *Journal of Economic Theory*, 193, 105222.

[4] Liu, Qingmin and Andrzej Skrzypacz (2014) "Limited Records and Reputation Bubbles," *Journal of Economic Theory* 151, 2-29.