

# Building Reputations via Summary Statistics

Harry PEI\*

July 30, 2022

**Abstract:** A patient seller decides whether to build a reputation for exerting high effort in front of a sequence of consumers. Each consumer decides whether to trust the seller after she observes the number of times that the seller took each of his actions in the last  $K$  periods, but not the order with which these actions were taken. I show that (i) the seller's payoff from building a reputation is at least his commitment payoff for all  $K$  and in *all* equilibria, and (ii) the seller sustains his reputation for exert high effort in *all* equilibria *if and only if*  $K$  is below some cutoff. Although a larger  $K$  allows more consumers to observe the seller's opportunistic behavior, it weakens their incentives to punish the seller after they observe opportunistic behavior. This effect undermines the seller's reputational incentives and lowers consumers' welfare. I also provide conditions under which coarsening the summary statistics observed by the consumers encourages the seller to sustain his reputation and improves consumers' welfare.

**Keywords:** limited memory, reputation sustainability, summary statistics, robust behavioral predictions.

## 1 Introduction

In many settings of economic interest, people have limited information about their trading partners' past records and can only observe some summary statistics of their recent behaviors.

To fix ideas, consider a market where consumers mainly learn through word-of-mouth communication with other consumers. This can happen due to the lack of record-keeping institutions. After a consumer interacts with a seller, she remembers who the seller is and her experience with the seller for some time, during which she can share it with future consumers. However, it is hard for consumers to communicate more detailed information such as who bought before them and what they learnt from others. This is because information transmitted through word-of-mouth is often filtered. For example, senders may omit complicated details due to the costs of describing them (Banerjee and Fudenberg 2004, Neihaus 2011) and receivers may not have enough capacity to process detailed information. As a result, consumers may only know *the number of times* that the seller provided good and bad services within a certain time frame, but it might be hard for them to know other details of the seller's history, such as the exact timing of these actions.

---

\*Northwestern University. I thank Dan Barron, James Best, Laura Doval, Jeff Ely, Drew Fudenberg, George Georgiadis, Yingkai Li, Qingmin Liu, Wojciech Olszewski, Larry Samuelson, Ali Shourideh, Vasiliki Skreta, Alex Smolin, Takuo Sugaya, Caroline Thomas, Matthew Thomas, and Alex Wolitzky for helpful comments. I thank NSF Grant SES-1947021 for financial support.

This paper examines a patient player’s incentive to sustain his reputation when his opponents can only observe the summary statistics of his recent behaviors. Focusing on the case where players’ stage-game payoffs are supermodular, my main result shows that the patient player sustains his reputation in all equilibria *if and only if* his opponents’ memory length is below some cutoff. The intuition is that longer memory weakens the short-run players’ incentives to punish the patient player when he deviates occasionally.

My baseline model is a repeated game between a patient player (e.g., a seller) and a sequence of short-run players (e.g., a sequence of consumers). The seller discounts future payoffs since the game ends with a positive probability after each period. Each period, the seller chooses his effort and a consumer chooses the extent to which she trusts the seller after she observes *the number of times that the seller took each of his actions in the last  $K$  periods*.<sup>1</sup> The seller is either a commitment type who exerts the highest effort in every period, or an opportunistic type who maximizes his payoff. The seller’s *reputation* is the probability that the consumer’s belief assigns to the commitment type.

The consumers in my model understand that the seller’s action may depend on the game’s history, such as his reputation. However, they *cannot* observe when the game started and *cannot* observe the order with which the seller took his actions.<sup>2</sup> As a result, they cannot use strategies that depend on the details of the game’s history and longer memories have ambiguous effects on consumers’ ability to monitor the seller: When memories are longer, each consumer receives information about a larger number of actions, which implies that the seller can be punished by more consumers after he shirks. However, each consumer’s signal about the seller’s action in any given period also becomes more noisy. My model stands in contrast to Fudenberg and Levine (1989) as well as existing reputation models with limited memories such as Liu (2011), Liu and Skrzypacz (2014), and Pei (2022), where consumers *can* observe the order of the seller’s actions. My model describes situations where each consumer interacts with the seller once, but remembers the seller’s action against her for  $K$  periods and communicates information about that action to consumers who arrive in the next  $K$  periods. She, however, does not communicate other details such as the identities of her predecessors and what she learnt from them, probably due to the costs of describing these details.

My analysis focuses on games that satisfy a *monotone-supermodularity* condition: Players’ actions are strategic complements and the seller’s payoff decreases in his effort and increases in consumers’ trust.

First, I show that the seller’s best reply in the repeated game must satisfy a *no-back-loop property*: The

---

<sup>1</sup>In Section 3.5, I consider more general summary statistics, which is characterized by a partition of the seller’s action space such that each consumer only observes the number of times that the seller’s last  $K$  actions belong to each partition element.

<sup>2</sup>One potential concern is that consumers need to make Bayesian inferences about the seller’s current-period action after observing the summary statistics of the seller’s last  $K$  actions, which can be quite complicated in some equilibria. I address this concern by showing that my theorems apply even when we focus attention on equilibria where consumers’ inference problems are simple and intuitive, such as equilibria where consumers believe that the distribution of the seller’s current-period action is close to the frequency of the seller’s actions in the last  $K$  periods. I provide more a more detailed explanation in Section 3.3.

seller will not shirk when he has a positive reputation and then restore his reputation to a positive level in the future. This property applies regardless of the seller's discount factor as well as the consumers' strategy it best replies to. My no-back-loop property rules out *reputation cycles* in which the seller milks his reputation when it is strictly positive and then exerts high effort until he has a positive reputation again.

When the opportunistic-type seller's strategy satisfies the *no-back-loop property*, there is *at most one period* over the infinite horizon where he has a positive reputation but does not exert the highest effort. Since the commitment type exerts the highest effort *in every period* and consumers do not know when the game started, consumers believe that the seller will exert the highest effort with probability close to one when he has a positive reputation. This leads to Theorem 1, which shows that in every equilibrium, consumers will play a best reply to the seller's highest effort when the seller has a positive reputation. This implies that the seller can secure his commitment payoff by exerting the highest effort in every period.

Nevertheless, the fact that the seller can secure his commitment payoff by building a reputation *does not* imply that he will do so in equilibrium, since other strategies may give him a higher payoff. Focusing on games where the highest effort is the seller's optimal commitment action, Theorem 2 shows that a patient seller exerts the highest effort with frequency close to one in all equilibria *if and only if*  $K$  is below some cutoff. This implies that longer memory hurts reputational incentives and lowers consumers' welfare.

The rationale behind Theorem 2 is that having a longer memory undermines consumers' incentives to punish a seller who shirks occasionally. For a heuristic explanation, suppose consumers believe that the opportunistic type will shirk in periods  $0, K, 2K, \dots$  and will exert the highest effort in other periods. Since consumers cannot observe calendar time, they believe that the seller will shirk with probability close to  $\frac{1}{K}$  after they observe that the seller shirked once in the last  $K$  periods. When  $K$  is large,  $\frac{1}{K}$  is small, each consumer believes that it is unlikely that the seller's current-period effort is low, so she has no incentive to punish the seller even though she knows that the seller is opportunistic. If this is the case, then the seller prefers *shirking once every  $K$  periods* to *exerting the highest effort in every period*, making consumers' beliefs self-fulfilling. This leads to an equilibrium where the opportunistic-type seller's reputation is always zero. One can also construct equilibria where the seller shirks  $n$  times every  $K$  periods, provided that consumers have no incentive to punish him when he shirks with probability  $\frac{n}{K}$ .

My main technical contribution is to show that if  $K$  is small, then in *every* equilibrium, consumers have no incentive to play their best reply to the highest effort when the seller's reputation is zero,<sup>3</sup> i.e., the seller is *guaranteed to be punished* after losing his reputation. The seller sustains his reputation for exerting high

---

<sup>3</sup>My technique can analyze other repeated games with limited records and incomplete information. For example, in a work in progress, I use these techniques to analyze anonymous random matching games between a population of patient players. I provide conditions under which patient players cooperate in all equilibria, which is novel compared to the existing results in this literature.

effort in all equilibria since he cannot obtain his commitment payoff via other strategies.

When  $K$  is small, Theorem 2 delivers a sharp prediction on the patient player's behavior that applies to *all* equilibria. This stands in contrast to most of the existing results in the reputation literature that focus exclusively on equilibrium payoffs or players' behaviors in some particular equilibria. It stands in contrast to the behavioral predictions in Fudenberg and Levine (1989), where Li and Pei (2021) show that there are equilibria where the seller shirks with frequency bounded away from zero. This is because when consumers can observe the entire history of play, the seller may still receive his commitment payoff after he separates from the commitment type. The lack of punishment leads to equilibria where the seller shirks. In my model with a small  $K$ , the seller sustains his reputation since he is guaranteed to receive his commitment payoff when he has a positive reputation and is guaranteed to be punished after he loses his reputation.

From the above perspective, Theorem 2 contributes to the discussions on the sustainability of reputations. Instead of focusing on the patient player's long-run behavior as in Cripps, Mailath and Samuelson (2004), I examine the *discounted frequency* with which the patient player plays his commitment action. Compared to the criteria for reputation sustainability in Cripps, Mailath and Samuelson (2004), my criteria has a comparative advantage in evaluating the uninformed players' welfare and social welfare.<sup>4</sup> My main result provides a tight condition under which the patient player sustains his reputation in *all* equilibria.

Theorem 2 also has implications on societies' ability to sustain cooperation when people have limited information about others' past behaviors. Most of the existing works on this topic focus on repeated games with anonymous random matching such as Kandori (1992) and Ellison (1994). A notable exception is Bhaskar and Thomas (2019) who study a repeated trust game *without* commitment types between a patient player and a sequence of myopic players, each has a finite memory. They show that cooperation cannot be sustained in any purifiable equilibrium when the short-run players can perfectly observe past outcomes, but can be sustained in some purifiable equilibria when past outcomes are observed with noise. By contrast, I study a reputation model and examine the effects of memory length on a patient player's incentive to sustain cooperation when his opponents can only observe the summary statistics of his recent behaviors. I provide conditions under which the patient player sustains his reputation in all equilibria.

I also study an extension where consumers learn from *coarse summary statistics*, in the sense that there is a partition of the seller's action space such that each consumer only observes the number of times that the seller's last  $K$  actions belong to each element of that partition. This assumption fits when consumers cannot precisely communicate the seller's actions to future consumers and can only tell which of the several broad

---

<sup>4</sup>Ekmekci, Gossner and Wilson (2012) and Liu and Skrzypacz (2014) propose another criteria for reputation sustainability, that the patient player can secure his commitment payoff *at every history* in every equilibrium. My Theorem 1 implies that for every  $K \geq 1$ , the patient player can secure his commitment payoff at every on-path history of every Nash equilibrium.

categories the seller's actions belong to (e.g., *good actions*, *bad actions*). I show that when  $K$  is intermediate and the seller has at least three actions, coarsening the summary statistics may improve the welfare of the consumers. Intuitively, consider a partition where each consumer only knows the number of times that the seller exerted the highest effort in the last  $K$  periods but cannot tell the difference between other actions. Under such a partition, it is never optimal for the seller to take any action other than the highest effort and the lowest effort. Ruling out intermediate effort levels provides consumers a stronger incentive to punish the seller after he loses his reputation. The threat of punishment encourages the seller to exert the highest effort.

## 2 Baseline Model

Time is indexed by  $t = 0, 1, \dots$ . A long-lived player 1 (e.g., seller) interacts with a different player 2 (e.g., consumer) in each period. After each period, the game ends with probability  $1 - \delta$  with  $\delta \in (0, 1)$ . I assume that player 1 is indifferent between receiving one unit of utility in the current period and receiving one unit of utility in the next period. Under this assumption, the seller discounts his future payoffs by  $\delta$ .

In period  $t$ , player 1 chooses  $a_t \in A$  and player 2 chooses  $b_t \in B$  from finite sets  $A$  and  $B$ . Their stage-game payoffs are  $u_1(a_t, b_t)$  and  $u_2(a_t, b_t)$ . I introduce two assumptions on players' stage-game payoffs. I provide examples that satisfy them later in this section. First, I assume that players' actions are strategic complements and that player 1 prefers to lower his action but prefers his opponents to raise their actions.

**Assumption 1.** *There exist a complete order on  $A$ ,  $\succ_A$ , and a complete order on  $B$ ,  $\succ_B$ , such that:*

1.  $u_1(a, b)$  and  $u_2(a, b)$  have strictly increasing differences in  $a$  and  $b$ .
2.  $u_1(a, b)$  is strictly increasing in  $b$  and is strictly decreasing in  $a$ .

Let  $a^* \equiv \max A$  be the highest element in  $A$  under  $\succ_A$ . Let  $b^*$  be player 2's lowest best reply to  $a^*$  under  $\succ_B$ . Let  $u_1(a^*, b^*)$  be player 1's *commitment payoff*.

I introduce a condition in the comparative statics literature which ensures that player 2 has *single-peaked preferences* over her actions regardless of her belief about player 1's action. Let  $B \equiv \{b^1, \dots, b^m\}$ , where  $b^m \succ_B \dots \succ_B b^1$ . For every  $i \leq m - 1$  and  $a \in A$ , let  $\gamma_a(i) \equiv u_2(a, b^i) - u_2(a, b^{i+1})$  be player 2's gain from locally decreasing her action from  $b^{i+1}$  to  $b^i$  when player 1's action is  $a$ . Let  $I \equiv \{1, 2, \dots, m - 1\}$ .

**Definition 1.**  $\gamma_a : I \rightarrow \mathbb{R}$  has *single crossing property* if  $\gamma_a(i) \geq 0$  implies that  $\gamma_a(j) > 0, \forall j > i$ .

**Definition 2.**  $\gamma_a : I \rightarrow \mathbb{R}$  and  $\gamma_{\bar{a}} : I \rightarrow \mathbb{R}$  satisfy *signed-ratio monotonicity* if:

1. For any  $i \in I$  such that  $\gamma_a(i) > 0$  and  $\gamma_{\bar{a}}(i) < 0$ , we have  $-\frac{\gamma_{\bar{a}}(i)}{\gamma_a(i)} \geq -\frac{\gamma_{\bar{a}}(j)}{\gamma_a(j)}$  for every  $j > i$ .

2. For any  $i \in I$  such that  $\gamma_a(i) < 0$  and  $\gamma_{\hat{a}}(i) > 0$ , we have  $-\frac{\gamma_a(i)}{\gamma_{\hat{a}}(i)} \geq -\frac{\gamma_a(j)}{\gamma_{\hat{a}}(j)}$  for every  $j > i$ .

**Assumption 2.** Player 2's stage-game payoff function  $u_2(a, b)$  satisfies:

1. For every  $a \in A$ ,  $\gamma_a$  has single-crossing property.
2. For every  $(a, \hat{a}) \in A \times A$ ,  $\gamma_a$  and  $\gamma_{\hat{a}}$  satisfy signed-ratio monotonicity.

Notice that Assumption 2 is trivially satisfied when  $|B| = 2$ . Assumption 2 has bite only when  $|B| \geq 3$ , in which case it is satisfied by a non-degenerate set of parameters.

According to Theorem 1 in Quah and Strulovici (2012), single-crossing functions  $f$  and  $g$  satisfying signed-ratio monotonicity is *necessary and sufficient* for all convex combinations of  $f$  and  $g$  to have the single-crossing property. Therefore, if  $u_2$  satisfies Assumption 2, then for every  $\alpha \in \Delta(A)$ ,  $\sum_{a \in A} \gamma_a(i) \alpha(a)$  has the single-crossing property. This implies that  $u_2(\alpha, b)$  is *single-peaked* with respect to  $b$ , in which case *either* player 2 has a unique pure-strategy best reply to  $\alpha$ , *or* she has two pure-strategy best replies to  $\alpha$  that are adjacent elements under  $\succ_B$ . Therefore, if  $u_2$  satisfies Assumption 2, then any pair of elements in

$$\mathcal{B}^* \equiv \left\{ \beta \in \Delta(B) \mid \text{there exists } \alpha \in \Delta(A) \text{ such that } \beta \text{ best replies to } \alpha \right\} \quad (2.1)$$

can be ranked according to FOSD. That is to say, player 2's mixed-strategy best replies to player 1's actions can be completely ranked according to FOSD.

Before choosing  $a_t$ , player 1 observes all the past actions  $h^t \equiv \{a_s, b_s\}_{s=0}^{t-1}$  and his perfectly persistent type  $\omega \in \{\omega_s, \omega_c\}$ . Let  $\omega_c$  stand for a *commitment type* who plays his highest action  $a^* \equiv \max A$  in every period. Let  $\omega_s$  stand for a *strategic type* who maximizes his discounted average payoff  $\sum_{t=0}^{\infty} (1 - \delta) \delta^t u_1(a_t, b_t)$ . Let  $\pi_0 \in (0, 1)$  be the prior probability of the commitment type. For future reference, I call  $a^*$  player 1's *commitment action* and actions that belong to  $A \setminus \{a^*\}$  *opportunistic actions*.

Before choosing  $b_t$ , player 2<sub>t</sub> observes the number of times that player 1 took each of his actions in the last  $\min\{t, K\}$  periods,<sup>5</sup> where  $K \in \mathbb{N}$  is a strictly positive integer that measures memory length. In contrast to the reputation models of Fudenberg and Levine (1989, 1992) and existing reputation models with limited memories such as Liu (2011), Liu and Skrzypacz (2014), and Pei (2022), player 2 *cannot* observe the order of player 1's actions. For example, if  $K = 3$ , then player 2 who arrives in period 4 cannot distinguish between  $(a_1, a_2, a_3) = (a^*, a^*, a')$ ,  $(a_1, a_2, a_3) = (a^*, a', a^*)$ , and  $(a_1, a_2, a_3) = (a', a^*, a^*)$ .

<sup>5</sup>Section 3.5 studies an extension where player 2 learns from coarse summary statistics. Each *coarse summary statistics* is characterized by a partition of  $A \equiv A_1 \cup \dots \cup A_n$  and player 2<sub>t</sub> only observes the number of times that player 1's last  $\min\{t, K\}$  actions belong to each partition element. My baseline model corresponds to the finest partition of  $A$ .

The short-run players *cannot* directly observe calendar time. They have a common prior and update their beliefs after observing their histories.<sup>6</sup> Since the game ends with probability  $1 - \delta$  after each period, for every  $t \in \mathbb{N}$ , the probability player 2's prior assigns to calendar time being  $t + 1$  equals  $\delta$  times the probability her prior assigns to calendar time being  $t$ . Therefore, player 2's prior assigns probability  $(1 - \delta)\delta^t$  to calendar time being  $t$ .<sup>7</sup> In Section 4.2, I extend my theorems to other prior beliefs about calendar time, which can distinguish the role of player 1's patience and the role of player 2's prior belief about calendar time.

Let  $\mathcal{H}_1 \equiv \left\{ (a_s, b_s)_{s=0}^{t-1} \text{ s.t. } t \in \mathbb{N} \text{ and } (a_s, b_s) \in A \times B \right\}$  be the set of player 1's histories. Let

$$\mathcal{H}_2 \equiv \left\{ (n_1, \dots, n_{|A|}) \in \mathbb{N}^{|A|} \text{ s.t. } n_1 \geq 0, \dots, n_{|A|} \geq 0 \text{ and } n_1 + \dots + n_{|A|} \leq K \right\} \quad (2.2)$$

be the set of player 2's histories. Strategic-type player 1's strategy is  $\sigma_1 : \mathcal{H}_1 \rightarrow \Delta(A)$ . Let  $\Sigma_1$  be the set of player 1's strategies. Player 2's strategy is  $\sigma_2 : \mathcal{H}_2 \rightarrow \Delta(B)$ . Let  $\Sigma_2$  be the set of player 2's strategies. Let  $\sigma \equiv (\sigma_1, \sigma_2)$  be a typical strategy profile. The solution concept is Perfect Bayesian equilibrium, under which it is without loss of generality to focus on  $\sigma_2 : \mathcal{H}_2 \rightarrow \mathcal{B}^*$ , i.e., player 2 plays a best reply to some  $\alpha \in \Delta(A)$  at every history. My results extend to the solution concept of Nash equilibrium.

**Examples:** I present a few examples where players' payoffs satisfy Assumptions 1 and 2. My leading example is the product choice game in Mailath and Samuelson (2015) where players' payoffs are:<sup>8</sup>

seller \ consumer	Trust	No Trust
High Effort	1, 1	$-c_N, x$
Low Effort	$1 + c_T, -x$	0, 0

Once players' actions are ranked according to  $H \succ_A L$  and  $T \succ_B N$ , both assumptions are satisfied when  $0 < c_T < c_N$  and  $x \in (0, 1)$ . That is, the seller's cost of effort is lower when consumers trust him and consumers are willing to trust the seller if and only if the probability of high effort is at least  $x$ .

For an example where players have more than two actions, suppose  $A \subset \mathbb{R}_+$  and  $B \subset \mathbb{R}_+$  are finite sets, where  $a \in A$  is the seller's effort and  $b \in B$  is the extent to which consumers trust the seller. Each consumer's payoff function is  $-(a - b)^2$ . The seller's payoff function  $u_1(a, b)$  is strictly decreasing in  $a$ , strictly

<sup>6</sup>Because the first  $K$  short-run players observe fewer than  $K$  actions, they can perfectly infer calendar time based on their histories. All other short-run players know that calendar time is at least  $K$  after they observe their histories.

<sup>7</sup>A similar assumption on the prior belief about calendar time is made in Cripps and Thomas (2019). Section 4 in Hu (2020) provides two interpretations for such a prior belief. The first interpretation is that the short-run players enter in a fixed order, and they are uncertain about their own identities and they hold an identical prior over their own identities. The second interpretation is that the short-run players know their own identities, but are uncertain about their entering periods. Hu (2020) also constructs a distribution over the entry process under which any two agents who observe the same history share the same belief.

<sup>8</sup>Following Mailath and Samuelson (2015, page 168), I interpret "Trust" as purchasing a premium product or a customized product and "No Trust" as purchasing a standardized product. Under this interpretation, future consumers may observe the seller's effort even when the current-period consumer does not trust the seller.

increasing in  $b$ , and has strictly increasing differences. Verifying that these payoffs satisfy Assumption 1 is straightforward. To see why they satisfy Assumption 2, notice that for every  $\alpha \in \Delta(A)$ , player 2's best replies are the elements in  $B$  that are closest to the expected value of player 1's action under  $\alpha$ . Therefore, player 2 either has a unique pure best reply, or has two pure best replies that are adjacent elements in  $B$ .

Although Assumption 1 restricts attention to games in which player 1's actions can be completely ranked, both of my theorems can be generalized to the case where  $(A, \succ_A)$  is a *lattice* with a unique highest element  $a^*$ . This fits situations where the seller's effort is multi-dimensional, e.g., the seller chooses both the quality of his product and the quality of his customer service. It also fits when the seller interacts with  $m \in \mathbb{N}$  consumers in each period: The seller chooses his effort against each consumer from a lattice, the commitment type chooses the highest effort against all consumers, and every consumer observes the number of times that the seller took each of his actions against consumers who arrived in the last  $K$  periods. That is to say, every consumer who arrives after period  $K$  can observe the summary statistics of  $Km$  actions.

### 3 Results

Section 3.1 establishes a *no-back-loop property* that applies to each of the long-run player's best reply under any discount factor. Section 3.2 uses this property to show that the long-run player receives at least his commitment payoff in all equilibria when his discount factor is above some cutoff. Section 3.3 shows that the commitment outcome  $(a^*, b^*)$  occurs with frequency arbitrarily close to 1 in all equilibria *if and only if*  $K$  is below some cutoff. I discuss the effects of memory length on consumers' welfare and whether consumers can attain a higher welfare when they observe coarser summary statistics of the seller's history.

#### 3.1 The No-Back-Loop Property

Let  $\pi_t \in [0, 1]$  be the probability player 2's belief assigns to the commitment type after she observes the summary statistics of player 1's actions in the last  $\min\{t, K\}$  periods. I call  $\pi_t$  player 1's *reputation* in period  $t$ . For every  $t \geq K$ ,  $\pi_t > 0$  if and only if player 1's history  $h^t \equiv (a_s, b_s)_{s=0}^{t-1}$  belongs to

$$\mathcal{H}_1^* \equiv \left\{ (a_s, b_s)_{s=0}^{t-1} \mid t \geq K \text{ and } (a_{t-K}, \dots, a_{t-1}) = (a^*, \dots, a^*) \right\}. \quad (3.1)$$

Intuitively,  $\mathcal{H}_1^*$  is the subset of histories where player 1 took his commitment action in each of the last  $K$  periods. Let  $\mathcal{H}_1(\sigma_1, \sigma_2)$  be the set of player 1's histories that occur with positive probability under  $(\sigma_1, \sigma_2)$ . Let  $U_1^\delta(\sigma_1, \sigma_2)$  be player 1's discounted average payoff under  $(\sigma_1, \sigma_2)$  when his discount factor is  $\delta$ .



For every  $\sigma_2$  and every pure strategy  $\hat{\sigma}_1 : \mathcal{H}_1 \rightarrow A$ , I say that  $\hat{\sigma}_1$  satisfies the *no-back-loop property* with respect to  $\sigma_2$  if there exists *no*  $h^t \in \mathcal{H}_1(\hat{\sigma}_1, \sigma_2) \cap \mathcal{H}_1^*$  such that when player 1 uses strategy  $\hat{\sigma}_1$ , he plays an action that is not  $a^*$  at  $h^t$  and reaches another history that belongs to  $\mathcal{H}_1(\hat{\sigma}_1, \sigma_2) \cap \mathcal{H}_1^*$  in the future. I say that  $\hat{\sigma}_1$  best replies to  $\sigma_2$  if  $\hat{\sigma}_1 \in \arg \max_{\sigma_1 \in \Sigma_1} U_1^\delta(\sigma_1, \sigma_2)$ . Recall the definition of set  $\mathcal{B}^*$  in (2.1).

**No-Back-Loop Lemma.** *Suppose  $(u_1, u_2)$  satisfies Assumptions 1 and 2. For any  $\delta \in (0, 1)$  and player 2's strategy  $\sigma_2 : \mathcal{H}_2 \rightarrow \mathcal{B}^*$ . If player 1's pure strategy  $\hat{\sigma}_1 : \mathcal{H}_1 \rightarrow A$  best replies to  $\sigma_2$ , then  $\hat{\sigma}_1$  satisfies the no-back-loop property with respect to  $\sigma_2$ .*

The no-back-loop lemma implies that under each of the long-run player's best reply in the repeated game, *either* he does not take any opportunistic action when he has a positive reputation (i.e., he never milks his reputation), *or* his future reputation is always zero after he milks his reputation (i.e., he never restores his reputation). This conclusion applies for all discount factors and for all strategies of the short-run players where they only play (mixed) actions in  $\mathcal{B}^*$ . In Section 4.1, I explain why this lemma hinges on the combination of (i) supermodular stage-game payoffs and (ii) the short-run players not being able to observe the order of the long-run player's actions. My lemma rules out the *reputation cycles* in Liu (2011) and Liu and Skrzypacz (2014), where the long-run player milks his reputation and then restores his reputation. Those papers assume that the long-run player's stage-game payoff is strictly *submodular* and the short-run players *can* observe the order of the long-run player's actions. Both features stand in contrast to my model.

My lemma *does not* follow from existing results on supermodular games.<sup>9</sup> This is because even when players' stage-game payoffs are *monotone-supermodular*, it is not necessarily the case that when the game is played repeatedly, the long-run player has a stronger incentive to play higher actions at histories where the short-run players' actions are higher. Intuitively, the long-run player's action in the current period affects future short-run players' observations, which in turn affects future short-run players' actions as well as the long-run player's continuation value. Hence, the long-run player's incentive depends not only on the current short-run player's action, but also on future short-run players' responses after they observe his action.

The formal proof is in Appendix A. I provide a heuristic explanation that focuses on player 1's pure strategies that depend only on his actions in the last  $K$  periods, including the order of these  $K$  actions.

Suppose by way of contradiction that there exists  $\hat{\sigma}_1$  that best replies to  $\sigma_2 : \mathcal{H}_2 \rightarrow \mathcal{B}^*$  such that when player 1 plays according to  $\hat{\sigma}_1$ , he plays  $a' (\neq a^*)$  at a history where he has a positive reputation, and after a finite number of periods, reaches a history  $h^t$  that satisfies  $(a_{t-K}, \dots, a_{t-1}) = (a'', a^*, \dots, a^*)$  where  $a'' \neq a^*$  but  $a''$  and  $a'$  can be the same, and then he plays  $a^*$  at  $h^t$ , after which he has a positive reputation again. I

<sup>9</sup>Most of the existing results on supermodular games focus on one-shot games, and it is well-known from the insightful work of Echenique (2004) that a repeated supermodular game is not necessarily supermodular.

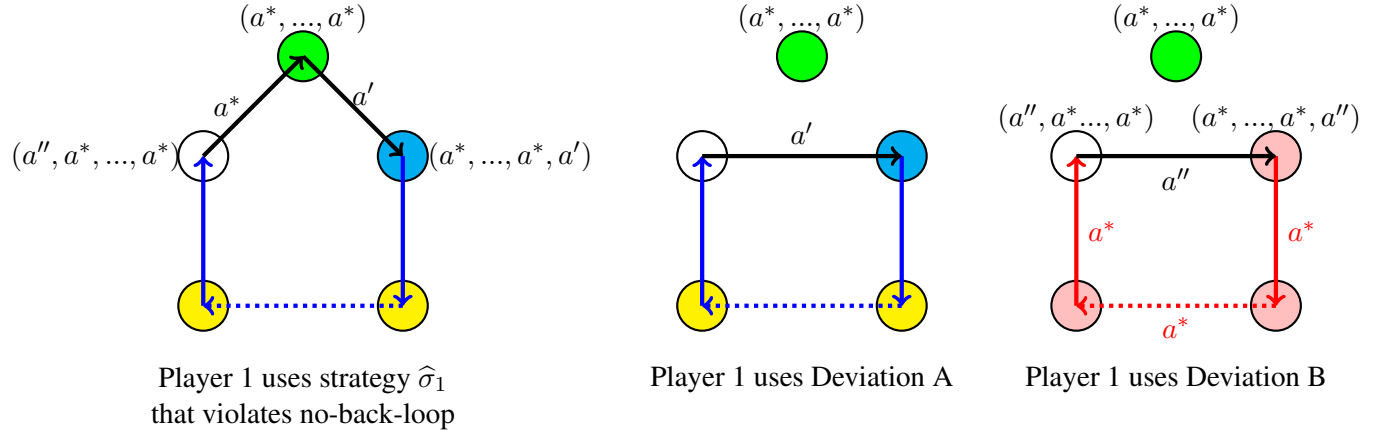


Figure 1: The green circle represents a history where player 1 has a positive reputation. The white circle represents a history where player 1 is one-period-away from having a positive reputation. The blue circle represents a history that is reached after player 1 milks his reputation. The yellow circles represent histories that are reached on the path from the blue circle to the white circle when player 1 plays  $\hat{\sigma}_1$ . The pink circles represent histories where  $a''$  was played once and  $a^*$  was played  $K - 1$  times in the last  $K$  periods.

depict player 1's on-path play under strategy  $\hat{\sigma}_1$  in the left panel of Figure 1.

If  $\hat{\sigma}_1$  best replies to  $\sigma_2$ , then player 1 prefers  $a'$  to  $a^*$  at the green circle and prefers  $a^*$  to  $a'$  at the white circle. No matter whether player 1 is currently at the green or the white circle, he will reach the green circle after playing  $a^*$  and will reach the blue circle after playing  $a'$ . Hence, the differences in player 1's incentives at the green and the white circles *cannot* be driven by his continuation value. This implies that such differences in incentives can only be driven by player 1's stage-game payoff, which is affected by the current-period short-run player's action. Since  $u_1(a, b)$  has strictly increasing differences and  $a^* \succ_A a'$ , it *cannot* be the case that player 2's mixed action at the green circle strictly FOSDs her mixed action at the white circle. When  $u_2(a, b)$  satisfies Assumption 2, any pair of elements in  $\mathcal{B}^*$  can be ranked according to FOSD. This implies that player 2's action at the white circle weakly FOSDs her action at the green circle.

I propose two deviations for player 1 starting from the white circle, which I call Deviation A and Deviation B, respectively. They are depicted in the middle and right panels of Figure 1:

- **Deviation A:** Plays  $a'$  at the white circle, and then follows strategy  $\hat{\sigma}_1$ .
- **Deviation B:** Plays  $a''$  at the white circle, then plays  $a^*$  for  $K - 1$  consecutive periods after which play will reach the white circle again, and then follows strategy  $\hat{\sigma}_1$ .

Since player 2 *cannot* observe the order of player 1's actions, she takes the same action at the white circle and at every pink circle. I compare player 1's continuation value at the white circle when he uses strategy

$\hat{\sigma}_1$  to his continuation values at the white circle under Deviation A and Deviation B, respectively.

1. Compared to  $\hat{\sigma}_1$ , Deviation A takes a lower-cost action  $a'$  at the white circle, skips the green circle, and frontloads the payoffs along the blue lines (i.e., the blue, yellow, and white circles). If player 1 prefers  $\hat{\sigma}_1$  to Deviation A, then his average payoff from the circles along the blue lines (i.e., the payoff that Deviation A frontloads) must be strictly lower than his stage-game payoff at the green circle.
2. Compared to  $\hat{\sigma}_1$ , Deviation B takes a lower-cost action  $a''$  at the white circle, skips the green circle, and induces payoffs along the red lines in the next  $K - 1$  periods (i.e., the pink circles). If player 1 prefers  $\hat{\sigma}_1$  to Deviation B, then his average payoff along the red lines must be strictly smaller than a convex combination of his payoff at the green circle and his average payoff along the blue lines.

Therefore, if player 1 prefers  $\hat{\sigma}_1$  to both Deviations A and B at the white circle, then his stage-game payoff at the green circle must be strictly greater than his average payoff along the red lines. Since player 2 can only observe the summary statistics but not the order of player 1's actions, player 2's action at every circle along the red line coincides with her action at the white circle. This leads to a contradiction since player 2's action at the white circle weakly FOSDs her action at the green circle, and player 1's stage-game payoff is increasing in player 2's action. This implies that at the white circle, either Deviation A or B yields a strictly higher payoff for player 1 relative to strategy  $\hat{\sigma}_1$ . Hence,  $\hat{\sigma}_1$  cannot be a best reply to  $\sigma_2$ .

### 3.2 The Returns from Building Reputations

I use the *no-back-loop lemma* to show that for every  $K$  and at every on-path history of every equilibrium, the patient player's continuation value is at least  $u_1(a^*, b^*)$ . My proof implies that he can secure payoff  $u_1(a^*, b^*)$  by playing  $a^*$  in every period. Let  $\underline{a}$  be player 1's lowest action. For every  $\pi_0 \in (0, 1)$ , there exists  $\underline{\delta}(\pi_0) \in (0, 1)$  such that for every  $\delta > \underline{\delta}(\pi_0)$ , all of player 2's best replies to the following mixed action:

$$\left\{ 1 - \frac{(1-\delta)(1-\pi_0)}{\pi_0} \right\} a^* + \frac{(1-\delta)(1-\pi_0)}{\pi_0} \underline{a}$$

are no less than  $b^*$  under the order  $\succ_B$  defined in Assumption 1. Such  $\underline{\delta}(\pi_0)$  exists since  $b^*$  is the lowest best reply to  $a^*$ ,  $\frac{(1-\delta)(1-\pi_0)}{\pi_0} \rightarrow 0$  as  $\delta \rightarrow 1$ , and best reply correspondences are upper-hemi-continuous.

**Theorem 1.** *Suppose  $(u_1, u_2)$  satisfies Assumptions 1 and 2, and  $\delta > \underline{\delta}(\pi_0)$ . The strategic-type player 1's discounted average payoff at every on-path history of every equilibrium is at least:*

$$(1 - \delta^K) \min_{b \in B} u_1(a^*, b) + \delta^K u_1(a^*, b^*). \quad (3.2)$$

Theorem 1 suggests that when players' actions are strategic complements, the patient player can secure his commitment payoff at *every* on-path history of *every* equilibrium even when his opponents can only observe the summary statistics of his recent behaviors. My result applies as long as  $\delta$  is above some cutoff and no matter how small  $K$  is. This stands in contrast to existing reputation results that require either an infinite  $K$  (e.g., Fudenberg and Levine 1989), or a large  $K$  (e.g., Theorem 2 in Liu and Skrzypacz 2014),<sup>10</sup> or the short-run players observing the entire history of some noisy signals that can statistically identify the patient player's past actions (e.g., Fudenberg and Levine 1992, Gossner 2011, and Theorem 2 in Pei 2022).<sup>11</sup>

Each short-run player in my model receives *only one signal*, which is the summary statistics of the long-run player's recent behaviors. As a result, she *cannot* observe everything her predecessors observe. Because of this, Theorem 1 does not follow from the arguments in Fudenberg and Levine (1989, 1992) and Gossner (2011), which require each short-run player observing everything her predecessors observe.

Since each signal received by the short-run player is affected by *all* of the long-run player's actions in the last  $K$  periods, the short-run player in my model face a novel *lack-of-identification problem*. For example, when  $K = 2$ , player  $2_t$  cannot tell the difference between  $(a_{t-2}, a_{t-1}) = (a', a^*)$  and  $(a_{t-2}, a_{t-1}) = (a^*, a')$ . This stands in contrast to the lack-of-identification problems faced by the short-run players in Ely and Välimäki (2003) where the monitoring technology has a product structure, that is, each short-run player's signal is affected by *exactly one* action of the long-run player.

My proof develops a novel argument using the *no-back-loop lemma*. I show that when  $\delta > \underline{\delta}(\pi_0)$ , player 2 has no incentive to play any action lower than  $b^*$  when calendar time is at least  $K$  and player 1 has a positive reputation. This implies Theorem 1, since when the strategic-type player 1 deviates by playing  $a^*$  in every period, player 2's action is no lower than  $b^*$  starting from period  $K$ .

*Proof of Theorem 1:* Let  $\Sigma_1^*$  be the set of player 1's pure strategies that satisfy the no-back-loop property. For every integer  $t \in \mathbb{N}$ , let  $E_t$  be the event that *player 1 is strategic and has a positive reputation in period  $t$* . Fix any  $\sigma_1 \in \Sigma_1^*$  and  $\sigma_2$ , let  $p_t(\sigma_1, \sigma_2)$  be the *ex ante* probability of event  $E_t$  when the strategic-type player 1 plays  $\sigma_1$  and player 2 plays  $\sigma_2$ . Since player 1 is the commitment type with probability  $\pi_0$ , we have  $p_t(\sigma_1, \sigma_2) \leq 1 - \pi_0$  for every  $t \in \mathbb{N}$ . Let  $\mathbb{N}^*(\sigma_1, \sigma_2) \subset \mathbb{N}$  be the set of calendar time  $t$  such that  $p_t(\sigma_1, \sigma_2) > 0$  and  $t \geq K$ . For every  $t \in \mathbb{N}^*(\sigma_1, \sigma_2)$ , let  $q_t(\sigma_1, \sigma_2)$  be the probability that player 1 *does not* play  $a^*$  in period  $t$  conditional on event  $E_t$ . Since  $\sigma_1 \in \Sigma_1^*$ , if player 1 plays according to  $\sigma_1$ , as soon as he

<sup>10</sup>Theorem 2 in Liu and Skrzypacz (2014) shows that for every  $\pi_0 > 0$ , there exists  $\bar{K} \in \mathbb{N}$  such that the patient player can secure his commitment payoff at every history in every equilibrium when the memory length  $K$  is *above*  $\bar{K}$ .

<sup>11</sup>Theorem 2 in Pei (2022) establishes a reputation result when each short-run player observes all previous short-run players' actions, an unboundedly informative private signal about the long-run player's current-period action, and the long-run player's actions in the last  $K$  periods. He shows that each short-run player's action is either informative about the long-run player's action against her, or this short-run player plays her best reply to the long-run player's commitment action with probability close to 1.

plays any action that is not  $a^*$  at any history after period  $K$  where he has a positive reputation, he will never have a positive reputation in the future. This implies that  $\sum_{t \in \mathbb{N}^*(\sigma_1, \sigma_2)} p_t(\sigma_1, \sigma_2) q_t(\sigma_1, \sigma_2) \leq 1 - \pi_0$ .

Fix an equilibrium  $(\tilde{\sigma}_1, \sigma_2)$ . The no-back-loop lemma implies that  $\tilde{\sigma}_1 \in \Delta(\Sigma_1^*)$ .<sup>12</sup> For every pure strategy  $\sigma_1 \in \Sigma_1^*$ , let  $\tilde{\sigma}_1(\sigma_1)$  be the probability mixed strategy  $\tilde{\sigma}_1$  assigns to  $\sigma_1$ . This probability is well-defined since  $\Sigma_1^*$  is a countable set. Recall that player 2's prior belief assigns probability  $(1 - \delta)\delta^t$  to the calendar time being  $t$ . At any history after period  $K$  where player 1 has a positive reputation, player 2 believes that player 1's action is *not*  $a^*$  with probability

$$\frac{\sum_{\sigma_1 \in \Sigma_1^*} \tilde{\sigma}_1(\sigma_1) \sum_{t \in \mathbb{N}^*(\sigma_1, \sigma_2)} (1 - \delta)\delta^t p_t(\sigma_1, \sigma_2) q_t(\sigma_1, \sigma_2)}{\pi_0 \sum_{t=K}^{+\infty} (1 - \delta)\delta^t + \sum_{\sigma_1 \in \Sigma_1^*} \tilde{\sigma}_1(\sigma_1) \sum_{t \in \mathbb{N}^*(\sigma_1, \sigma_2)} (1 - \delta)\delta^t p_t(\sigma_1, \sigma_2)}. \quad (3.3)$$

The denominator of (3.3) is at least  $\pi_0 \delta^K$ . Since  $\sum_{t \in \mathbb{N}^*(\sigma_1, \sigma_2)} p_t(\sigma_1, \sigma_2) q_t(\sigma_1, \sigma_2) \leq 1 - \pi_0$  for every  $\sigma_1 \in \Sigma_1^*$ ,  $\Sigma_1^*$  is a countable set, and  $t \geq K$  for every  $t \in \mathbb{N}^*(\sigma_1, \sigma_2)$ , the numerator of (3.3) is no more than  $(1 - \delta)(1 - \pi_0)\delta^K$ . This suggests that the value of (3.3) is no more than  $\frac{(1-\delta)(1-\pi_0)}{\pi_0}$ . The definition of  $\underline{\delta}(\pi_0)$  together with  $u_2(a, b)$  having strictly increasing differences implies that when  $\delta > \underline{\delta}(\pi_0)$ , actions strictly lower than  $b^*$  are *not* optimal for player 2 at any history after period  $K$  and where player 1 has a positive reputation. This implies that at every on-path history of every equilibrium, if player 1 plays  $a^*$  in every period starting from that history, then he receives a continuation value of at least (3.2).  $\square$

### 3.3 Main Result: The Incentive to Sustain Reputations

The fact that player 1 can secure his commitment payoff by playing  $a^*$  in every period *does not* imply that he will do so in equilibrium. My next result focuses on player 1's behavior, and in particular, the frequency with which he plays the commitment action. It requires another assumption on players' stage-game payoffs:

**Assumption 3.** *Players' stage-game payoffs  $(u_1, u_2)$  are such that  $b^*$  is a strict best reply to  $a^*$ ,*

$$u_1(a^*, b^*) > \max_{a \in A, b \prec b^*} u_1(a, b) \quad \text{and} \quad \{a^*\} = \arg \max_{a \in A} \left\{ \min_{b \in BR_2(a)} u_1(a, b) \right\}. \quad (3.4)$$

The product choice game is a leading example that satisfies Assumption 3. The first part of Assumption 3 is generically satisfied since  $b^*$  is defined as player 2's lowest best reply to  $a^*$ . The second part requires that player 1 cannot obtain payoffs weakly greater than  $u_1(a^*, b^*)$  when player 2's action is strictly lower

<sup>12</sup>One caveat to apply the no-back-loop lemma to prove Theorem 1 for Nash Equilibrium is that it restricts attention to  $\sigma_2 : \mathcal{H} \rightarrow \mathcal{B}^*$ , while player 2 may play actions that do not belong to  $\mathcal{B}^*$  at off-path histories. However, notice that the proof of the no-back-loop lemma uses  $\sigma_2 : \mathcal{H}_2 \rightarrow \mathcal{B}^*$  only when comparing player 2's mixed actions at the green circle and at the white circle, both of which occur with positive probability under  $(\hat{\sigma}_1, \sigma_2)$ . Hence, we can show that  $\hat{\sigma}_1$  satisfies the no-back-loop property with respect to  $\sigma_2$  if (i)  $\hat{\sigma}_1$  best replies to  $\sigma_2$ , and (ii)  $\sigma_2(h_2^t) \in \mathcal{B}^*$  for every  $h_2^t$  that occurs with positive probability under  $(\hat{\sigma}_1, \sigma_2)$ .

than  $b^*$ . The third part requires  $a^*$  to be player 1's optimal pure commitment action. Since  $u_1(a, b)$  is strictly decreasing in  $a$  and  $a^*$  is the highest element of  $A$ ,  $b^*$  does not best reply to any action other than  $a^*$ .

Let  $\mathbb{E}^\sigma[\cdot]$  be the expectation operator induced by strategy profile  $\sigma \equiv (\sigma_1, \sigma_2)$  conditional on player 1 being the strategic type. Since the game continues with probability  $\delta$  after each period, the expected sum of the short-run players' payoffs depends only on the *discounted frequency* (or the *occupation measure*) of each action profile  $(a, b) \in A \times B$  under strategy profile  $\sigma$ , which is defined as

$$F^\sigma(a, b) \equiv \mathbb{E}^\sigma \left[ \sum_{t=0}^{+\infty} (1 - \delta) \delta^t \mathbf{1}\{a_t = a, b_t = b\} \right]. \quad (3.5)$$

Let

$$H^\sigma \equiv \mathbb{E}^\sigma \left[ \sum_{t=1}^{+\infty} (1 - \delta) \delta^{t-1} \mathbf{1}\{\pi_t > 0\} \right] \quad (3.6)$$

be the *discounted frequency* of histories where player 1 has a positive reputation excluding period 0.

**Theorem 2.** *Suppose  $(u_1, u_2)$  satisfies Assumptions 1, 2, and 3.*

1. *Suppose  $b^*$  does not best reply to mixed action  $\frac{K-1}{K}a^* + \frac{1}{K}a'$  for every  $a' \neq a^*$ . There exists a constant  $C \in \mathbb{R}_+$  that is independent of  $\delta$  such that  $F^\sigma(a^*, b^*) \geq 1 - (1 - \delta)C$  and  $H^\sigma \geq 1 - (1 - \delta)C$  for every equilibrium strategy profile  $\sigma$  under discount factor  $\delta$ .*
2. *Suppose  $b^*$  is a strict best reply to  $\frac{K-1}{K}a^* + \frac{1}{K}a'$  for some  $a' \neq a^*$ . There exist  $\underline{\delta} \in (0, 1)$  and  $\eta > 0$  such that for every  $\delta > \underline{\delta}$ , there is a PBE strategy profile  $\sigma$  such that  $\sum_{b \in B} F^\sigma(a^*, b) < 1 - \eta$  and  $H^\sigma = 0$ , i.e.,  $\pi_t = 0$  with probability 1 for every  $t \geq 1$  when player 1 is the strategic type.<sup>13</sup>*

The proof of the first part is in Appendix B and the proof of the second part is in Online Appendix C. Since Assumption 3 requires  $b^*$  to be player 2's strict best reply to  $a^*$ ,  $b^*$  does not best reply to  $\frac{K-1}{K}a^* + \frac{1}{K}a'$  for every  $a' \neq a^*$  if and only if  $K$  is below some cutoff  $K^*$ , where  $K^*$  is at least 2 and its value depends only on player 2's stage-game payoff function  $u_2(a, b)$ . My result implies that:

1. When  $K$  is below  $K^*$ , (i) player 1 plays  $a^*$  with frequency arbitrarily close to one in all equilibria as  $\delta \rightarrow 1$ , and (ii) histories where player 1 has a strictly positive reputation occur with frequency arbitrarily close to 1 as  $\delta \rightarrow 1$ . In Section 4.2, I discuss generalizations to situations where player 1's discount rate differs from the probability with which the game continues after each period.
2. When  $K$  is above  $K^*$ , no matter how patient player 1 is, there always exists an equilibrium in which

<sup>13</sup>I require the constructed PBE to satisfy the *no signaling what you don't know* refinement in Fudenberg and Tirole (1991).

player 1 takes opportunistic actions with frequency bounded away from zero, and starting from period 1, the strategic-type player 1's reputation is always zero on the equilibrium path.

In terms of how large  $\eta$  can be and how it is related to  $K$  and  $u_2$ , one can show that for every  $m \in \{1, 2, \dots, K-1\}$  and  $a' \neq a^*$  such that  $b^*$  is player 2's strict best reply to  $\frac{K-m}{K}a^* + \frac{m}{K}a'$ , there exists  $\underline{\delta} \in (0, 1)$  such that for every  $\delta > \underline{\delta}$ , there is an equilibrium where the discounted frequency with which the strategic-type player 1 plays  $a'$  is approximately  $\frac{m}{K}$ . That is,  $\eta$  can be as large as  $\frac{m}{K}$ .

In the product choice game of Section 2, the cutoff  $K^*$  is  $\frac{1}{1-x}$ . Theorem 2 implies that consumers having a long memory undermines the seller's incentives to sustain reputations. This is because a longer memory has two effects on the seller's reputational incentives. An obvious effect is that each of the seller's action is observed by more consumers, so that he can be punished by more consumers after he shirks. This encourages him to exert high effort. However, a longer memory also undermines consumers' incentives to punish the seller after they observe a few instances of low effort. This encourages the seller to exert low effort occasionally instead of exerting high effort in every period. Theorem 2 suggests that the second effect dominates in terms of determining whether the patient player will sustain his reputation in *all* equilibria.

In order to understand the second effect as well as how to pin down the cutoff memory length, I focus on the product choice game and provide a heuristic explanation for why the cutoff memory length is  $\frac{1}{1-x}$ .

Suppose consumers believe that the strategic-type seller will play  $L$  exactly once in every  $K$  consecutive periods. When  $\delta$  is close to 1, consumers believe that  $L$  will be played in the current period with probability close to  $\frac{1}{K}$  after they observe  $L$  being played only once in the last  $K$  periods. Hence, consumers prefer to play  $T$  at such histories when  $x < \frac{K-1}{K}$ , or equivalently, when  $K > \frac{1}{1-x}$ . If this is the case, then the seller prefers *exerting low effort once every  $K$  periods* to *exerting high effort in every period*.

This is the idea behind my proof for the second part of Theorem 2, where I construct equilibria under a large  $K$  such that on the equilibrium path, the strategic-type player 1 plays  $a' (\neq a^*)$  in periods  $0, K, 2K, \dots$ , and plays  $a^*$  in other periods.<sup>14</sup> According to Bayes rule, his reputation on the equilibrium path is always zero starting from period 1. Starting from period  $K$ , player 2 plays  $b^*$  when player 1 has a positive reputation as well as at histories where she observes  $a'$  only once and the rest of the last  $K$  actions were  $a^*$ .

Player 2's belief at on-path histories has an intuitive interpretation and her optimal action at every on-path history is easy to compute: She believes that the distribution of player 1's current-period action is close to the frequency with which player 1 played each of his actions in the last  $K$  periods. In particular, she

<sup>14</sup>Since the strategic type's reputation is always zero after period 1, his behavior in this equilibrium does not violate the no-back-loop property. The behavior I just described is qualitatively different from the reputation cycles in Liu (2011) and Liu and Skrzypacz (2014), since the strategic-type long-run player has a positive reputation at some on-path histories in their equilibria.

believes that player 1 will choose  $a^*$  in the current period after observing player 1 has chosen  $a^*$  in all of the last  $K$  periods, and she believes that player 1 will choose  $a^*$  with probability close to  $\frac{K-1}{K}$  and  $a'$  with probability close to  $\frac{1}{K}$  after observing  $a^*$  occurred  $K - 1$  times and  $a'$  occurred once in the last  $K$  periods.

Using similar ideas, one can show that for every  $m \in \{1, 2, \dots, K - 1\}$  and  $a' \neq a^*$  such that  $b^*$  is a strict best reply to  $\frac{K-m}{K}a^* + \frac{m}{K}a'$ , there exists an equilibrium under a sufficiently large  $\delta$  where in every  $K$  consecutive periods, player 1 plays  $a'$  in  $m$  periods and plays  $a^*$  in  $K - m$  periods, and player 2 plays  $b^*$  when she observes  $a'$  being played  $m$  times and  $a^*$  being played  $K - m$  times in the last  $K$  periods.

My main technical contribution is in the proof of the first part. My technique is potable to analyze other repeated games with limited record-keeping (e.g., games with anonymous random matching between a population of long-run players) and to deliver predictions on payoffs and behaviors that apply to *all* equilibria.

A major challenge is that characterizing *all* equilibria in an infinitely repeated game is not tractable in general. Moreover, ruling out the type of equilibria constructed in the proof for the second part is *insufficient* to show that player 1 will play  $a^*$  with frequency close to 1 in *all* equilibria. One of the reasons is that a Bayesian short-run player's expectation of the long-run player's current-period action *may not* equal the frequency with which the long-run player played each of his actions in the last  $K$  periods. For example, the seller may play  $H$  with high probability even when he played  $L$  in all of the last  $K$  periods, and he may play  $L$  with probability bounded away from  $\frac{1}{K}$  when  $L$  was played once in the last  $K$  periods.

I explain the ideas behind the proof using the product choice game when  $K = 2$  and  $x > \frac{1}{2}$ , so that  $K < \frac{1}{1-x}$ . Readers who are not interested in the proof can skip this part and jump to Section 3.4. In every period  $t$ , the seller's continuation value and incentive as well as consumer  $t$ 's action depend only on  $(a_{t-2}, a_{t-1})$ . I call  $(a_{t-2}, a_{t-1})$  the *state* in period  $t$ . Let  $S \equiv \{LL, LH, HL, HH\}$  be the *state space*, with  $s \in S$  a typical element. For example, the state in period  $t$  is  $LH$  if and only if  $a_{t-2} = L$  and  $a_{t-1} = H$ .

Fix any equilibrium  $\sigma$ . For every  $s \in S$ , let  $\mu(s)$  be the probability that the current state is  $s$  conditional on the seller being the strategic type and calendar time being at least 2 (in the general case, I replace 2 with  $K$ ). For every  $s, s' \in S$ , let  $Q(s \rightarrow s')$  be the probability that the state in the next period is  $s'$  conditional on the state in the current period being  $s$ , the seller being the strategic type, and the calendar time being at least 2. Let  $p(s)$  be the probability that the state is  $s$  in period 2 conditional on the seller being the strategic type.

**Step 1: Decompose the Discounted Frequency of States** My proof hinges on the following equation, which I establish later as Lemma B.1:

$$\mu(s^*) = (1 - \delta)p(s^*) + \delta \sum_{s \in S} \mu(s)Q(s \rightarrow s^*) \text{ for every } s^* \in S. \quad (3.7)$$



The intuition behind (3.7) is that consumers do not directly observe calendar time, so  $\mu(s^*)$  is a convex combination of the probability of state  $s^*$  in period 2, which equals  $p(s^*)$ , and the average probability with which play reaches state  $s^*$  after period 2, which equals  $\sum_{s \in S} \mu(s)Q(s \rightarrow s^*)$ . Since consumers' prior belief assigns probability  $(1 - \delta)\delta^t$  to the calendar time being  $t$ , the weight on the first term equals  $1 - \delta$ .

**Step 2: The Inflow-Outflow Lemma** For every non-empty subset of states  $S' \subset S$ , let

$$\mathcal{I}(S') \equiv \sum_{s \notin S'} \sum_{s^* \in S'} \mu(s)Q(s \rightarrow s^*)$$

be the *inflow* from other states to states in  $S'$ , and let

$$\mathcal{O}(S') \equiv \sum_{s \in S'} \sum_{s^* \notin S'} \mu(s)Q(s \rightarrow s^*)$$

be the *outflow* from states in  $S'$  to other states. Since  $\sum_{s \notin S'} Q(s^* \rightarrow s) + \sum_{s \in S'} Q(s^* \rightarrow s) = 1$  for every  $s^* \in S$  and  $S' \subset S$ , we have:

$$\sum_{s^* \in S'} \mu(s^*) = \sum_{s^* \in S'} \sum_{s \notin S'} \mu(s^*)Q(s^* \rightarrow s) + \sum_{s^* \in S'} \sum_{s \in S'} \mu(s^*)Q(s^* \rightarrow s) = \mathcal{O}(S') + \sum_{s^* \in S'} \sum_{s \in S'} \mu(s^*)Q(s^* \rightarrow s). \quad (3.8)$$

According to (3.7), we have

$$\sum_{s^* \in S'} \mu(s^*) = (1 - \delta) \sum_{s^* \in S'} p(s^*) + \delta \mathcal{I}(S') + \delta \sum_{s^* \in S'} \sum_{s \in S'} \mu(s^*)Q(s^* \rightarrow s),$$

or equivalently,

$$\sum_{s^* \in S'} \mu(s^*) = \mathcal{I}(S') + \sum_{s^* \in S'} \sum_{s \in S'} \mu(s^*)Q(s^* \rightarrow s) + \frac{1 - \delta}{\delta} \sum_{s^* \in S'} (p(s^*) - \mu(s^*)). \quad (3.9)$$

Equations (3.8) and (3.9) imply that

$$\left| \mathcal{I}(S') - \mathcal{O}(S') \right| = \left| \frac{1 - \delta}{\delta} \sum_{s^* \in S'} (p(s^*) - \mu(s^*)) \right| \leq \frac{1 - \delta}{\delta}. \quad (3.10)$$

That is to say, the difference between the inflow to and the outflow from any subset of states  $S'$  is no more than a linear function of  $1 - \delta$ .

**Step 3: Inflow to State  $HH$  and Outflow from State  $HH$  Must Be Small** The *no-back-loop* lemma implies that *either* (i) the strategic-type seller never plays  $L$  when the state is  $HH$ , *or* (ii) the strategic-type seller never returns to state  $HH$  in the future after playing  $L$  in state  $HH$ .

In the first case, the outflow from state  $HH$  to other states is 0, and (3.10) implies that the inflow from other states to state  $HH$  is no more than  $\frac{1-\delta}{\delta}$ . In the second case, the inflow to  $S'$  is zero, where

$$S' \equiv \left\{ s' \neq HH \text{ s.t. the seller can reach } HH \text{ from } s' \text{ under one of his best replies to player 2's equilibrium strategy} \right\}$$

Inequality (3.10) implies that the outflow from  $S'$  is no more than  $\frac{1-\delta}{\delta}$ , and the definition of  $S'$  implies that the inflow to state  $HH$  is no more than the outflow from  $S'$ , which is also no more than  $\frac{1-\delta}{\delta}$ . Applying inequality (3.10) again, we obtain that the outflow from state  $HH$  to other states is no more than  $\frac{2(1-\delta)}{\delta}$ . Summarizing these two cases, we have:

$$\max\{\mathcal{I}(\{HH\}), \mathcal{O}(\{HH\})\} \leq \frac{2(1-\delta)}{\delta}. \quad (3.11)$$

**Step 4: Consumers' Incentive After the Seller Loses His Reputation** I use inequalities (3.7), (3.10), and (3.11) to show that when  $K = 2$  and  $x > \frac{1}{2}$ , either  $\mu(HL) + \mu(LH)$  is bounded above by some linear function of  $1 - \delta$ , i.e.,  $\mu(HL) + \mu(LH) \rightarrow 0$  as  $\delta \rightarrow 1$ , or consumers have a strict incentive to play  $N$  after they observe  $L$  being played exactly once in the last two periods.

This is because when consumers observe one  $L$  and one  $H$  in the last 2 periods, i.e., they know that the state is either  $HL$  or  $LH$ , they believe that the seller will play  $H$  in the current period with probability

$$\frac{\mu(HL)Q(HL \rightarrow LH) + \mu(LH)Q(LH \rightarrow HH)}{\mu(HL) + \mu(LH)},$$

and will play  $L$  in the current period with probability

$$\frac{\mu(HL)Q(HL \rightarrow LL) + \mu(LH)Q(LH \rightarrow HL)}{\mu(HL) + \mu(LH)}.$$

Since  $x > \frac{1}{2}$ , consumers strictly prefer to play  $N$  after observing one  $L$  and one  $H$  if

$$\mu(HL)Q(HL \rightarrow LH) + \mu(LH)Q(LH \rightarrow HH) < \mu(HL)Q(HL \rightarrow LL) + \mu(LH)Q(LH \rightarrow HL).$$

By definition,  $\mathcal{I}(\{HH\}) = \mu(LH)Q(LH \rightarrow HH)$  and  $\mathcal{O}(\{HH\}) = \mu(HH)Q(HH \rightarrow HL)$ . Inequality (3.11) implies that  $\mu(LH)Q(LH \rightarrow HH)$  and  $\mu(HH)Q(HH \rightarrow HL)$  are close to 0 when  $\delta \rightarrow 1$ . Hence,

$\mu(HL)Q(HL \rightarrow LH) + \mu(LH)Q(LH \rightarrow HH)$  is close to  $\mu(HL)Q(HL \rightarrow LH)$ , and  $\mu(HL)Q(HL \rightarrow LL) + \mu(LH)Q(LH \rightarrow HL)$  is close to  $\mu(HL)Q(HL \rightarrow LL) + \mu(LH)$ . According to (3.7),

$$\mu(HL) \approx \underbrace{\mu(HH)Q(HH \rightarrow HL) + \mu(LH)Q(LH \rightarrow HL)}_{\equiv \mathcal{O}(\{HH\}) \approx 0} \approx \mu(LH)Q(LH \rightarrow HL),$$

we know that either  $\mu(HL)Q(HL \rightarrow LH) \approx \mu(LH)Q(LH \rightarrow HL)Q(HL \rightarrow LH)$ , with the right-hand-side being no more than  $\mu(LH)$ , is smaller than  $\mu(HL)Q(HL \rightarrow LL) + \mu(LH)$ , in which case consumers prefer to play  $N$  upon observing  $\{LH, HL\}$ , or  $\mu(LH) + \mu(HL)$  is small in the sense that it is bounded above by a linear function of  $1 - \delta$ .

**Step 5: States where Consumers Play  $N$  Occur with Frequency Close to Zero** I use inequality (3.11) to show that if consumers have a strict incentive to play  $N$  upon observing one  $H$  and one  $L$ , then  $\mu(HL) + \mu(LH)$  is bounded above by a linear function of  $1 - \delta$ . Intuitively, the seller's continuation value in state  $HH$  is 1. Because he can reach state  $HH$  by playing  $H$  in two consecutive periods, his continuation value in any other state is at least  $-(1 - \delta^2)c_N + \delta^2$ . Under the hypothesis that consumers play  $N$  when  $(a_{t-2}, a_{t-1}) \in \{HL, LH\}$ , the state cannot remain in  $\{HL, LH\}$  for too long. This is because otherwise, the seller cannot obtain payoff weakly greater than  $-(1 - \delta^2)c_N + \delta^2$ . In another word, the state must reach  $HH$  or  $LL$  within a bounded number of periods. Since the inflow to  $HH$  is small, we know that when  $\mu(HL) + \mu(LH)$  is large, there must be a large inflow to  $LL$  from  $\{HL, LH\}$ . Given the relationship between inflows and outflows in (3.10), there must also be a large outflow from  $LL$  to  $\{HL, LH\}$ . Hence, there exists a best reply of the seller that plays  $H$  at  $LL$ , plays  $L$  at  $LH$ , and plays  $L$  at  $HL$ . However, the seller's payoff from such a behavior is strictly less than his payoff from playing  $L$  in every period, since consumers have no incentive to play  $T$  when the state is either  $HL$  or  $LH$ . This leads to a contradiction.

Similarly, one can show that  $\mu(LL)$  is also close to 0 in all equilibria given that  $\mu(HL) + \mu(LH)$ ,  $\mathcal{I}(\{HH\})$ , and  $\mathcal{O}(\{HH\})$  are all bounded above by some linear function of  $1 - \delta$ . Suppose by way of contradiction that the ratio between  $\mu(LL)$  and  $\max\{\mu(HL) + \mu(LH), \mathcal{I}(\{HH\}), \mathcal{O}(\{HH\})\}$  can be arbitrarily large in some equilibria. Since  $\mu(HL) + \mu(LH)$  is small, the inflow to  $\{HL, LH\}$  must be small, which implies that  $\mu(LL)Q(LL \rightarrow LH)$  is small. Since  $\mu(LL)$  is large relative to  $\mu(HL) + \mu(LH)$ , we know that  $Q(LL \rightarrow LH)$  must be close to zero, which implies that the strategic-type seller plays  $L$  with positive probability in state  $LL$ . Therefore, playing  $L$  is optimal for the seller in state  $LL$ , which implies that it is optimal for the seller to remain in state  $LL$  after he reaches there. Since the seller's continuation value is no less than  $-(1 - \delta^2)c_N + \delta^2$  at every history, consumers must have an incentive

to play  $T$  in state  $LL$  in order for the seller to receive such a payoff by remaining in state  $LL$ . Hence, the probability with which the seller plays  $H$  at  $LL$  is at least  $x$ , which implies that  $Q(LL \rightarrow LL) \leq 1 - x$ . However, (3.7) implies that  $\mu(LL)$  is close to  $\mu(LL)Q(LL \rightarrow LL) + \mu(HL)Q(HL \rightarrow LL)$ , and under the hypothesis that  $\mu(HL)$  is close to 0, we know that  $\mu(LL)$  is close to  $\mu(LL)Q(LL \rightarrow LL)$ . This suggests that  $\mu(LL)(1 - Q(LL \rightarrow LL))$  must be close to 0. Since  $\mu(LL)(1 - Q(LL \rightarrow LL)) \geq x\mu(LL)$  and  $x > 0$ , we know that  $\mu(LL) \rightarrow 0$  as  $\mu(LL)(1 - Q(LL \rightarrow LL)) \rightarrow 0$ .

**Remark:** Appendix B generalizes the above argument to  $K \geq 2$  as well as to stage games where players have more than two actions. First, I derive the analogs of (3.7) and (3.10) under a general  $K$ . I define each *state* as a sequence of player 1's actions with length  $K$ . For every  $0 \leq k \leq K$ , let  $S_k$  be the set of states where  $k$  of the last  $K$  actions are not  $a^*$ . The no-back-loop lemma and (3.10) imply that the inflow to  $S_0$  from other states and the outflow from  $S_0$  to other states must be bounded above by  $\frac{2(1-\delta)}{\delta}$  (Lemma B.3).

The rest of the proof uses an induction argument. For any two subsets of states  $S' \cap S'' = \emptyset$ , let

$$Q(S' \rightarrow S'') \equiv \sum_{s' \in S'} \sum_{s'' \in S''} \mu(s')Q(s' \rightarrow s'') \quad (3.12)$$

be the *flow* from  $S'$  to  $S''$ . I show that for every  $k \geq 1$ , if the flow from  $S_{k-1}$  to  $S_k$  and the flow from  $S_{k-1}$  to  $S_k$  are both bounded above by a linear function of  $1 - \delta$ , then the total discounted frequency of states in  $S_k$  is also bounded above by a linear function of  $1 - \delta$ . Iterate the above argument, we obtain that the flow from  $S_k$  to  $S_{k+1}$  and the flow from  $S_k$  to  $S_{k+1}$  are also bounded above by a linear function of  $1 - \delta$ .

The case in which  $k = 1$  is shown in Lemma B.3, i.e., the flow from  $S_1$  to  $S_0$  and the flow from  $S_0$  to  $S_1$  are both close to zero. Establishing the inductive step proceeds as follows. For every  $k \in \{0, \dots, K\}$ , I partition  $S_k \equiv \bigcup_k S_{j,k}$  such that player 2 only observes which partition element the state belongs to. First, I show in Lemma B.4 that for every  $S_{j,k}$  with  $k \geq 1$ , if the flow from  $S_{k-1}$  to  $S_{j,k}$  and the flow from  $S_{j,k}$  to  $S_{k-1}$  are both bounded above by some linear function of  $1 - \delta$ , then *either* the total discounted frequency of states in  $S_{j,k}$  is also bounded above by some linear function of  $1 - \delta$ , *or* player 2 has no incentive to play  $b^*$  when the state belongs to  $S_{j,k}$ . Second, I show in Lemma B.5 that for every  $S_{j,k}$  with  $k \geq 1$ , if the flow from  $S_{k-1}$  to  $S_{j,k}$  and the flow from  $S_{j,k}$  to  $S_{k-1}$  are both bounded above by some linear function of  $1 - \delta$ , and player 2 has no incentive to play  $b^*$  when the state belongs to  $S_{j,k}$ , then under Assumption 3, the total discounted frequency of states in  $S_{j,k}$  must be bounded above by some linear function of  $1 - \delta$ .<sup>15</sup>

<sup>15</sup>Since player 1 receives  $u_1(a^*, b^*)$  after playing  $a^*$  for  $K$  periods, he cannot remain in  $S_{j,k}$  for a long time given that his payoff is less than  $u_1(a^*, b^*)$  when the state belongs to  $S_{j,k}$ . If the flow from  $S_{j,k}$  to  $S_{k-1}$  is small, and every  $S_{i,k}$  occurs with discounted frequency close to 0 when player 2 has an incentive to play  $b^*$  at  $S_{i,k}$ , then there must exist a best reply for player 1 under which he reaches  $S_{k+1}$  starting from  $S_{j,k}$ . If he never returns to  $S_{j,k}$  after reaching  $S_{k+1}$ , then states in  $S_{j,k}$  are transitory and therefore,

### 3.4 Implications of Theorem 2

**The Short-Run Players' Welfare:** In the equilibrium I constructed in order to show the second statement of Theorem 2, player 1 plays some  $a'$  ( $\neq a^*$ ) once in every  $K$  consecutive periods, and player 2 plays  $b^*$  either when player 1 has a positive reputation or when player 1 has played  $a^*$  in  $K - 1$  of the last  $K$  periods, and has played  $a'$  once in the last  $K$  periods. Player 1's payoff in this equilibrium is approximately  $\frac{K-1}{K}u_1(a^*, b^*) + \frac{1}{K}u_1(a', b^*)$ , which is strictly greater than his commitment payoff  $u_1(a^*, b^*)$ .

Hence, Theorem 2 implies that player 1 can obtain a payoff that is strictly greater than his optimal commitment payoff  $u_1(a^*, b^*)$  in some equilibria *if and only if*  $K$  is above some cutoff  $K^*$ . Under an additional assumption that  $u_2(a, b)$  is strictly increasing in  $a$ , i.e., consumer's payoff is strictly increasing in the seller's effort, Theorem 2 implies that player 2's discounted average payoff is arbitrarily close to their first best payoff  $u_2(a^*, b^*)$  in all equilibria *if and only if* their memory length  $K$  is below the cutoff  $K^*$ .

**Corollary 1.** *Suppose  $(u_1, u_2)$  satisfies Assumptions 1, 2, and 3.*

1. *Suppose  $b^*$  does not best reply to mixed action  $\frac{K-1}{K}a^* + \frac{1}{K}a'$  for every  $a' \neq a^*$ . For every  $\varepsilon > 0$ , there exists  $\underline{\delta} \in (0, 1)$  such that for every  $\delta > \underline{\delta}$ , and in every equilibrium strategy profile  $\sigma$  under  $\delta$ , player 1's payoff is no more than  $u_1(a^*, b^*) + \varepsilon$  and player 2's welfare satisfies*

$$U_2^\sigma \equiv \mathbb{E}^\sigma \left[ \sum_{t=0}^{+\infty} (1 - \delta)\delta^t u_2(a_t, b_t) \right] \geq u_2(a^*, b^*) - \varepsilon. \quad (3.13)$$

2. *Suppose  $b^*$  is a strict best reply to  $\frac{K-1}{K}a^* + \frac{1}{K}a'$  for some  $a' \neq a^*$ , and  $u_2(a, b)$  is strictly increasing in  $a$ . There exist  $\eta > 0$  and  $\underline{\delta} \in (0, 1)$ , such that for every  $\delta > \underline{\delta}$ , there exists an equilibrium  $\sigma$  such that player 1's payoff is more than  $u_1(a^*, b^*) + \eta$  and player 2's welfare is less than  $u_2(a^*, b^*) - \eta$ .*

The proof directly follows from that of Theorem 2, and is omitted in order to avoid repetition. Corollary 1 implies that longer memories enable the seller to obtain a higher payoff at the expense of consumers' welfare. This happens in equilibria where the seller shirks occasionally and consumers trust the seller as long as the seller did not shirk too many times in the last  $K$  periods. By contrast, when  $K$  is small, the seller receives his commitment payoff and consumers receive their first best payoff in *all* equilibria.

**More on Player 1's Behavior:** The next corollary focuses on the case where the discounted frequency of  $(a^*, b^*)$  is arbitrarily close to 1 in all equilibria and examines the implication of Theorem 2 on player

---

must occur with frequency close to 0. If he returns to  $S_{j,k}$  after reaching  $S_{k+1}$ , he needs to play  $a^*$  in  $S_{k+1}$  which is the most costly action, after which the state reaches  $S_{j,k}$  and he receives a low payoff. This contradicts his incentive to play  $a^*$  at  $S_{k+1}$ .

1's behavior in each period. It shows that in every equilibrium, the strategic-type patient player will have a positive reputation with probability close to 1 after the initial few periods, after which he will take the highest action  $a^*$  with probability close to 1 in every period until  $t$  is large enough such that  $\delta^t$  is close to 0.

**Corollary 2.** *Suppose  $(u_1, u_2)$  satisfies Assumptions 1, 2, and 3, and  $b^*$  does not best reply to mixed action  $\frac{K-1}{K}a^* + \frac{1}{K}a'$  for every  $a' \neq a^*$ . For every  $\varepsilon > 0$ , there exist  $C_\varepsilon \in \mathbb{R}_+$  and  $\underline{\delta} \in (0, 1)$  such that for every  $\delta > \underline{\delta}$ , every equilibrium under  $\delta$ , and every  $t \in \mathbb{N}$  that satisfies  $\delta^t \in (\varepsilon, 1 - \varepsilon)$ :*

1. *The probability that  $h^t \in \mathcal{H}_1^*$  is at least  $1 - (1 - \delta)C_\varepsilon$ .*
2. *The strategic-type player 1 plays  $a^*$  with probability at least  $1 - (1 - \delta)C_\varepsilon$  in period  $t$ .*

The proof of Corollary 2 is in Online Appendix D, which uses Theorem 2 and the no-back-loop lemma.

**Comparison with the Behavioral Predictions in Fudenberg and Levine (1989):** I compare the frequency with which player 1 plays his commitment action in any equilibrium under any finite  $K$  to the *equilibrium frequencies* with which player 1 plays his commitment action in Fudenberg and Levine (1989).

In Fudenberg and Levine (1989), every short-run player observes the entire sequence of the long-run player's past actions. When players' payoffs satisfy Assumptions 1, 2, and 3, 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

$$G^*(u_1, u_2) \equiv \min_{(\alpha_1, \alpha_2, b_1, b_2, q) \in \Delta(A) \times \Delta(A) \times B \times B \times [0, 1]} \left\{ q\alpha_1(a^*) + (1 - q)\alpha_2(a^*) \right\}, \quad (3.14)$$

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

$$qu_1(\alpha_1, b_1) + (1 - q)u_1(\alpha_2, b_2) \geq u_1(a^*, b^*). \quad (3.15)$$

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  $u_1(a^*, b^*)$ . In the product choice game,  $G^*(u_1, u_2) = \frac{x}{1 + (1-x)c_T}$ . Corollary 3 shows that in my model, the discounted frequency of  $a^*$  is strictly bounded above  $G^*(u_1, u_2)$  for every finite  $K$ .

**Corollary 3.** *Suppose  $(u_1, u_2)$  satisfies Assumptions 1, 2, and 3. 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  $\sigma$  under  $\delta$ , we have:

$$\underbrace{\sum_{b \in B} F^\sigma(a^*, b)}_{\text{the discounted frequency of } a^* \text{ under } \sigma} > G^*(u_1, u_2) + \psi. \quad (3.16)$$

The proof is in Online Appendix E. Compared to Fudenberg and Levine (1989)'s model, the short-run players in my model receive coarser information about the patient player's history. Corollary 3 implies that the patient player plays his commitment action with strictly higher frequency in every equilibrium of my model compared to the worst equilibrium in terms of the frequency with which the patient player plays  $a^*$  in Fudenberg and Levine (1989)'s model. This suggests that for every finite  $K$ , my model delivers a sharper prediction on the patient player's equilibrium behavior relative to Fudenberg and Levine (1989)'s.<sup>16</sup>

This is because in Fudenberg and Levine (1989)'s model, in order to sustain an equilibrium where the patient player plays his commitment action with frequency bounded away from 1, the short-run players need to fine-tune their punishments based on the details of the patient player's history. For example, in the product choice game, an equilibrium where the discounted frequency of  $H$  is approximately  $G^*(u_1, u_2)$  is given by:

- In period 0, the strategic-type player 1 plays  $L$  and player 2 plays  $N$ .
- Starting from period 1, (i) if player 1 played  $H$  in period 0, then player 1 plays  $H$  and player 2 plays  $T$ , and player 2 plays  $N$  in all subsequent periods after observing player 1 played  $L$ ; (ii) if player 1 played  $L$  in period 0, then player 1 plays  $H$  with probability  $x$  and player 2 plays  $T$  until period  $\tilde{t}$ , where  $\tilde{t}$  is the smallest  $t \in \mathbb{N}$  that satisfies  $\sum_{s=1}^t (1 - \delta)\delta^{s-1}u_1(a_s, b_s) + (\delta^t - \delta^{t+1})(1 + c_T) > 1$ , after which the continuation play is either  $(H, T)$  or  $(L, N)$ . The discounted frequencies of  $(H, T)$  and  $(L, N)$  are such that player 1's continuation value in period 1 is 1. After period  $\tilde{t}$ , if player 1 plays  $L$  in any period where he is supposed to play  $H$ , then player 2 plays  $N$  in all subsequent periods.

A notable feature of the above equilibrium is that player 1 faces different punishments for shirking at different histories. If player 1 played  $H$  in period 0, then he will face a grim-trigger punishment when he plays  $L$  in any subsequent period, in which case he strictly prefers  $H$  to  $L$  starting from period 1. If player 1 played  $L$  in period 0, then he will be indifferent between  $H$  and  $L$  starting from period 1 until period  $\tilde{t}$  since he only faces a mild punishment for shirking in each of these periods. Moreover, the calendar time  $\tilde{t}$  starting from which shirking faces a grim-trigger punishment depends on the *discounted frequency* with

<sup>16</sup>Nevertheless, it is *not* the case that every equilibrium of my model remains an equilibrium in Fudenberg and Levine (1989)'s model. This is because the short-run players' incentive constraints can be violated once they receive more information, for example, after they observe the long-run player's actions in the distant past and/or the order of the long-run player's actions.

which player 1 has played  $H$  from period 1 to  $\tilde{t}$ . That is to say, player 2 needs to know the order of player 1's past actions in order to compute this discounted frequency. Therefore, such a strategy is infeasible when player 2 can only observe the summary statistics of player 1's recent play.

By contrast, the short-run players in my model can only observe the summary statistics of the long-run player's recent actions. Their lack of information limits their abilities to use different punishments at different histories. Suppose there exists an equilibrium where player 1 plays  $a^*$  with frequency less than 1. The strategic type must have an incentive to separate from the commitment type at some histories, so his continuation value after separation at each of these histories must be at least  $u_1(a^*, b^*)$ . In order to attain a continuation value of at least  $u_1(a^*, b^*)$ , the punishments for shirking at some subsequent histories need to be harsh enough so that player 1 strictly prefers to play  $H$  at those histories. Since player 2 receives coarse information about player 1's history, she must use the same harsh punishment for shirking at all histories that are indistinguishable from the ones at which she needs to punish player 1. Therefore, player 1 will face a harsh punishment for shirking at a larger set of histories when player 2 receives coarser information, so he has a strict incentive to play his commitment action  $a^*$  at a larger set of histories. This increases the frequency with which he plays his commitment action compared to that in Fudenberg and Levine (1989).

### 3.5 Learning from Coarse Summary Statistics

In this section, I focus on games that satisfy Assumptions 1, 2, and 3, and analyze the equilibrium outcomes when consumers can only observe *coarse summary statistics* about the seller's recent behaviors.

Following Acemoglu, Makhdoum, Malekian and Ozdaglar (2020), every coarse summary statistics is characterized by a partition of  $A$ , denoted by  $\{A_1, \dots, A_n\}$ , so that for every  $t \in \mathbb{N}$ , consumer  $t$  only observes the number of times the seller's action belongs to each  $A_i$  in the last  $\min\{t, K\}$  periods.

This extension fits when consumers cannot communicate precise information about the seller's action to future consumers. Instead, they can only describe which of the several broad categories the seller's action belongs to. For example, she can only tell whether she had a good experience or a bad experience with the seller, but she finds it too time-consuming to precisely describe exactly how good or how bad.

The summary statistics in my baseline model correspond to the finest partition of  $A$ . Under the coarsest partition of  $A$ , consumers receive no information about the seller's past actions.

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 player 1 will never choose action  $a \in A_i$  if there exists  $a' \in A_i$  that satisfies  $a \succ_A a'$ . Hence, analyzing the game under an  $n$ -partition  $\{A_1, \dots, A_n\}$  of  $A$  is equivalent to analyzing a game where player 1's action set only contains the following  $n$  actions:  $\{\min A_1, \dots, \min A_n\}$ .



When players' stage-game payoffs satisfy Assumptions 1, 2, and 3, player 2 has no incentive to play  $b^*$  unless player 1 plays  $a^*$  with positive probability. When the prior probability of commitment type  $\pi_0$  is small enough such that player 2 has no incentive to play  $b^*$  when player 1 plays  $a^*$  with probability no more than  $\pi_0$ , 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  $A = \{a^*\} \cup (A \setminus \{a^*\})$ , i.e., consumers only observe the number of times the seller chose  $a^*$  in the last  $K$  periods but cannot distinguish other actions, then consumers may receive a higher welfare under some intermediate  $K$ . Intuitively, such a partition helps the seller 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 consumers a stronger incentive to punish the seller after the seller loses his reputation, since consumers know that the seller will take the lowest action as long as he does not take the highest action. Such an effect motivates the seller to play  $a^*$  in every period.

**Corollary 4.** *Suppose  $(u_1, u_2)$  satisfies Assumptions 1, 2, and 3.*

1. *Suppose the partition element that contains  $a^*$  is not a singleton, then the strategic-type player 1 never plays  $a^*$  at any on-path history of any equilibrium.*
2. *Suppose the partition element that contains  $a^*$  is a singleton (without loss, let  $A_1 \equiv \{a^*\}$ ), then when  $\delta > \underline{\delta}(\pi_0)$ , the strategic-type player 1's payoff is at least (3.2) in every equilibrium. Furthermore,*
  - (i) *Suppose  $b^*$  does not best reply to mixed action  $\frac{K-1}{K}a^* + \frac{1}{K} \min_{j \in \{2, \dots, n\}} \{\min A_j\}$ . There exists  $C \in \mathbb{R}_+$  such that  $F^\sigma(a^*, b^*) \geq 1 - (1 - \delta)C$  and  $H^\sigma \geq 1 - (1 - \delta)C$  for every equilibrium strategy profile  $\sigma$  under  $\delta$ .*
  - (ii) *Suppose  $b^*$  is a strict best reply to mixed action  $\frac{K-1}{K}a^* + \frac{1}{K} \min_{j \in \{2, \dots, n\}} \{\min A_j\}$ . There exist  $\underline{\delta} \in (0, 1)$  and  $\eta > 0$  such that for every  $\delta > \underline{\delta}$ , there exists a PBE strategy profile  $\sigma$  such that  $\sum_{b \in B} F^\sigma(a^*, b) < 1 - \eta$  and  $H^\sigma = 0$ , i.e.,  $\pi_t = 0$  with probability 1 for every  $t \geq 1$  conditional on player 1 being the strategic type.*

Corollary 4 directly follows from Theorems 1 and 2. I omit the proof in order to avoid repetition. This corollary implies that fixing any  $(u_1, u_2)$  and  $K \in \mathbb{N}$ , if there exists a partition of  $A$  under which player 1 plays  $a^*$  with frequency arbitrarily close to one in all equilibria, then player 1 plays  $a^*$  with frequency arbitrarily close to one in all equilibria under partition  $A = \{a^*\} \cup (A \setminus \{a^*\})$ , i.e., each consumer only knows the number of times with which the seller chose the highest action in the last  $K$  periods but she cannot distinguish other actions. It also implies that coarsening the summary statistics *cannot* improve consumers'

welfare when the seller's action choice is binary, but *can* improve consumers' welfare when the seller has at least three actions and the memory length  $K$  satisfies:

1.  $b^*$  does not best reply to  $\frac{K-1}{K}a^* + \frac{1}{K}a$ ,
2.  $b^*$  is a strict best reply to  $\frac{K-1}{K}a^* + \frac{1}{K}a'$  for some  $a' \notin \{a^*, a\}$ .

That is, coarsening the summary statistics observed improves consumers' welfare when  $K$  is intermediate.

## 4 Discussions

### 4.1 Observing the Order of Player 1's Actions & Submodular Stage-Game Payoffs

My analysis relies on the *no-back-loop property*, which hinges on players' stage-game payoffs being *supermodular* and the short-run players *cannot* observe the order of the patient player's past actions. I explain why these conditions are not redundant for my results using the product choice game example.

**Observing the Order of Player 1's Past Actions:** Suppose every consumer can observe the seller's last  $K$  actions including the order of these actions. I maintain the supermodularity assumption that  $c_N > c_T > 0$ .

When  $K = 1$ , whether consumers can observe the order of the seller's last  $K$  actions is irrelevant. Theorems 1 and 2 imply that in all equilibria, the patient seller can secure his commitment payoff and will play  $H$  with frequency arbitrarily close to one. Proposition 1 shows that there exists some cutoff  $\underline{K} \geq 2$  such that when  $K \geq \underline{K}$ , there exist equilibria that violate the no-back-loop property. In those equilibria, the patient seller plays his commitment action with frequency bounded away from one and his discounted average payoff is bounded below his commitment payoff.

**Proposition 1.** *Suppose consumer in period  $t$  observes  $(a_{\max\{0, t-K\}}, \dots, a_{t-1})$  for every  $t \in \mathbb{N}$ . For every  $x \in (0, 1)$  and  $c_N > c_T > 0$ , there exist  $\underline{K} \geq 2$ ,  $\bar{\pi} \in (0, 1)$ , and  $\eta > 0$  such that when  $\pi_0 \in (0, \bar{\pi})$  and  $K \geq \underline{K}$ , for every  $\delta$  close enough to 1, there exists an equilibrium  $\sigma$  that violates the no-back-loop property. In this equilibrium,  $F^\sigma(H, T) < 1 - \eta$  and the seller's payoff is no more than  $1 - \eta$ .*

The proof is in Online Appendix F. I illustrate the ideas behind my proof using an example where  $K = 2$ . I construct an equilibrium in which the seller plays  $H$  when  $(a_{t-2}, a_{t-1}) = (L, H)$ , plays  $L$  when  $(a_{t-2}, a_{t-1}) = (H, L)$ , and mixes between  $H$  and  $L$  when  $(a_{t-2}, a_{t-1})$  is either  $(L, L)$  or  $(H, H)$ . Consumers play  $T$  when  $(a_{t-2}, a_{t-1}) = (L, H)$ , play  $N$  when  $(a_{t-2}, a_{t-1}) = (H, L)$ , and mix between  $T$  and  $N$  when  $(a_{t-2}, a_{t-1}) \in \{(L, L), (H, H)\}$ . Consumers' strategy in such an equilibrium is not feasible

when they cannot observe the order of the seller's actions. The existence of such an equilibrium also implies that there is no guarantee that consumers play  $T$  with higher probability when the seller exerted high effort more frequently in the last  $K$  periods. In the above equilibrium, consumers play  $T$  with higher probability when  $(a_{t-2}, a_{t-1}) = (L, H)$  compared to  $(a_{t-2}, a_{t-1}) = (H, H)$ , and they play  $T$  with higher probability when  $(a_{t-2}, a_{t-1}) = (L, L)$  compared to  $(a_{t-2}, a_{t-1}) = (H, L)$ .

The comparison between Proposition 1 and Theorem 1 suggests that allowing the uninformed players to observe the order of the informed player's actions changes the set of equilibrium payoffs. This stands in contrast to repeated Bayesian games where the uninformed player is long-lived, e.g., Renault, Solan and Vieille (2013), in which it is sufficient for the uninformed player to check the frequency with which the informed player played each of his actions. This is because in my model, the uninformed players are short-lived, so their incentives are sensitive to their beliefs about the informed player's current-period action. By contrast, the equilibrium constructed in Renault, Solan, and Vieille (2013) requires the uninformed player to have intertemporal incentives, since she does not play her myopic best reply at some on-path histories.

Proposition 1, Theorem 1, the results in Fudenberg and Levine (1989) and Pei (2022) together imply that the patient player's lowest equilibrium payoff is *non-monotone* with respect to the quality of the short-run players' information, measured in the sense of Blackwell. My discussion focuses on situations where players' stage-game payoffs satisfy Assumptions 1 and 2. My Theorem 1 and Fudenberg and Levine (1989)'s result imply that the patient player can secure his Stackelberg payoff in all equilibria *either* when the short-run players can observe the entire history of his actions, *or* when the short-run players can observe the summary statistics of the long-run player's actions (my baseline model with  $K = +\infty$ ),<sup>17</sup> *or* when the short-run players can only observe the summary statistics of the long-run player's last  $K$  actions where  $K$  is a strictly positive and finite integer. The patient player's lowest equilibrium payoff is bounded below his Stackelberg payoff *either* when the short-run players can observe his actions in the last  $K(\geq \underline{K})$  periods including the order of these actions, *or* when every short-run player can observe the patient player's actions in the last  $K(\geq 1)$  periods and at least one previous short-run player's action. One can also show that when the short-run players can observe the patient player's last  $K$  actions including the order of these actions, the patient player's *lowest* equilibrium payoff is weakly decreasing in  $K$  when  $K$  is strictly positive and finite.

**Submodular Stage-Game Payoff:** Suppose  $K = 1$  and players' payoffs are submodular, i.e.,  $c_T > c_N > 0$ . I show that there exists an equilibrium that violates the no-back-loop property, the patient seller receives his minmax payoff, and his frequency of exerting high effort is bounded away from 1.

---

<sup>17</sup>Although this is not a direct implication of Fudenberg and Levine (1989)'s result, one can establish this conclusion using the argument in Fudenberg and Levine (1989)'s proof. The details are available upon request.

**Proposition 2.** *Suppose  $K = 1$ ,  $c_T > c_N > 0$ , and  $\pi_0$  is small enough such that  $\frac{\pi_0}{\delta(1-\pi_0)-\pi_0} \leq \frac{1}{2}$ . There exists  $\underline{\delta} \in (0, 1)$  such that for every  $\delta > \underline{\delta}$ , there is an equilibrium where the seller's payoff is 0.*

The proof is in Online Appendix G. I provide intuition for why submodular stage-game payoffs differ from supermodular stage-game payoffs. When payoffs are submodular, i.e.,  $c_T > c_N > 0$ , it is still true that consumer  $t$  plays  $T$  with strictly higher probability when  $a_{t-1} = H$ . However, the seller has a stronger incentive to exert high effort when  $a_{t-1} = L$ . Unlike the case with supermodular payoffs, the seller can have incentives both to milk his reputation when  $a_{t-1} = H$  and to rebuild his reputation when  $a_{t-1} = L$ . This can provide consumers a rationale for not trusting the seller even after they observe high effort in the period before. My proof constructs an equilibrium where the seller exerts low effort when  $a_{t-1} = H$  and mixes between high and low effort when  $a_{t-1} = L$ . Consumer  $t$  plays  $N$  when  $a_{t-1} = L$  and plays  $T$  with probability between 0 and 1 when  $a_{t-1} = H$ .

## 4.2 Disentangle Player 1's Discount Factor & Player 2's Prior Belief about Calendar Time

The parameter  $\delta$  plays a dual role in my baseline model: It is both the long-run player's discount factor and the probability with which the game continues after each period. The latter affects the short-run players' incentives through their prior beliefs about calendar time.

My theorems can be extended to situations where the long-run player's discount factor does not coincide with the game's continuation probability. For example, suppose the short-run players' prior belief assigns probability  $(1 - \bar{\delta})\bar{\delta}^t$  to calendar time being  $t \in \mathbb{N}$ , and the long-run player's discount factor is  $\delta$ , which I assume is no more than  $\bar{\delta}$ . This model describes situations where the long-run player discounts future payoffs for two reasons. First, he is indifferent between receiving one unit of utility in period  $t$  and receiving  $\delta/\bar{\delta}$  unit of utility in period  $t - 1$ . Second, the game ends with probability  $1 - \bar{\delta}$  after each period.

My *no-back-loop lemma* generalizes to this setting, since the conclusion is independent of player 1's discount rate and the game's continuation probability. The statement of Theorem 1 is modified as follows: Suppose  $(u_1, u_2)$  satisfies Assumptions 1 and 2, and the continuation probability  $\bar{\delta}$  is large enough such that all of player 2's best replies to the mixed action

$$\left\{ 1 - \frac{(1 - \bar{\delta})(1 - \pi_0)}{\pi_0} \right\} a^* + \frac{(1 - \bar{\delta})(1 - \pi_0)}{\pi_0} \underline{a}$$

are no less than  $b^*$ , then the strategic-type player 1's discounted average payoff at every on-path history of every equilibrium is at least

$$(1 - \delta^K) \min_{b \in B} u_1(a^*, b) + \delta^K u_1(a^*, b^*).$$

Hence, my payoff lower bound applies to *any* discount factor of the long-run player, as long as the probability with which the game continues after each period is above some cutoff, and that cutoff depends only on the prior probability of commitment type  $\pi_0$  and player 2's stage-game payoff function  $u_2$ . Intuitively, since the long-run player's equilibrium strategy satisfies the no-back-loop property, there is at most one period over the infinite horizon in which he has a positive reputation yet he plays an action other than  $a^*$ . Therefore, every short-run player has a strict incentive to play  $b^*$  after observing  $a^*$  in the last  $K$  periods when her prior belief assigns a low enough probability to each calendar time. The latter is the case when  $\bar{\delta}$  is large.

For Theorem 2, let us redefine  $F^\sigma(a, b)$  and  $H^\sigma(a, b)$  based on the game's continuation probability  $\bar{\delta}$ :

$$F^\sigma(a, b) \equiv \mathbb{E}^\sigma \left[ \sum_{t=0}^{+\infty} (1 - \bar{\delta}) \bar{\delta}^t \mathbf{1}\{a_t = a, b_t = b\} \right],$$

and

$$H^\sigma \equiv \mathbb{E}^\sigma \left[ \sum_{t=1}^{+\infty} (1 - \bar{\delta}) \bar{\delta}^{t-1} \mathbf{1}\{\pi_t > 0\} \right].$$

The motivation is that a social planner who wants to maximize the expected sum of the short-run players' welfare only cares about  $F^\sigma(\cdot, \cdot)$ . The statement of Theorem 2 should be modified as follows. Suppose  $(u_1, u_2)$  satisfies Assumptions 1, 2, and 3, and the discount factor  $\delta$  is larger than some cutoff  $\underline{\delta} \in (0, 1)$ ,

1. Suppose  $b^*$  does not best reply to mixed action  $\frac{K-1}{K}a^* + \frac{1}{K}a'$  for every  $a' \neq a^*$ . Then there exists a constant  $C \in \mathbb{R}_+$  that is independent of  $\delta$  and  $\bar{\delta}$  such that  $F^\sigma(a^*, b^*) \geq 1 - (1 - \bar{\delta})C$  and  $H^\sigma \geq 1 - (1 - \bar{\delta})C$  for every equilibrium strategy profile  $\sigma$  under discount factor  $\delta$ .
2. Suppose  $b^*$  is a strict best reply to  $\frac{K-1}{K}a^* + \frac{1}{K}a'$  for some  $a' \neq a^*$ . There exist  $\underline{\delta} \in (0, 1)$  and  $\eta > 0$  such that for every  $\delta > \underline{\delta}$ , there is a PBE strategy profile  $\sigma$  such that  $\sum_{b \in B} F^\sigma(a^*, b) < 1 - \eta$  and  $H^\sigma = 0$ , i.e.,  $\pi_t = 0$  with probability 1 for every  $t \geq 1$  when player 1 is the strategic type.

Hence, the occupation measure of  $(a^*, b^*)$  is arbitrarily close to 1 in all equilibria even when player 1's discount factor  $\delta$  is bounded away from 1. The important parameter is  $\bar{\delta}$ , the game's continuation probability, which affects player 2's belief about calendar time. When  $K$  is small, the total occupation measure of action profiles other than  $(a^*, b^*)$  is bounded above by some linear function of  $1 - \bar{\delta}$ , i.e., it vanishes to zero as long as the game continues after each period with probability arbitrarily close to 1. Intuitively,  $\bar{\delta}$  being close to 1 implies that the short-run players' prior belief assigns similar probabilities to adjacent calendar times. This is also reflected in the proof, where one needs to replace  $\delta$  with  $\bar{\delta}$  in the first four steps in the proof of Theorem 2 (see Section 3.3). The fifth step only requires player 1's discount factor  $\delta$  to be above some cutoff, in which case histories where his stage-game payoffs are bounded below  $u_1(a^*, b^*)$  are transitory.

## 5 Conclusion & Related Literature

I study a reputation game in which the short-run players face a novel lack-of-identification problem: Each of them can only observe the summary statistics of the long-run player's recent actions. I show that when players' stage-game payoffs are monotone-supermodular, each of the long-run player's best reply in the repeated game must satisfy a *no-back-loop property*. This property implies that either the long-run player has no incentive to abandon his reputation when it is strictly positive, or he has no incentive to restore his reputation to a positive level after abandoning his reputation.

When the short-run players know that the long-run player's strategy satisfies the no-back-loop property, they prefer to play their myopic best reply to the long-run player's commitment action as long as the long-run player has a positive reputation and the game has lasted for at least  $K$  periods. This implies that the long-run player can secure his commitment payoff by playing his commitment action in every period.

I also show that the long-run player plays his commitment action and has a positive reputation with frequency close to one in all equilibria *if and only if* the short-run players' memory length  $K$  is below some cutoff. This is because when the short-run players have long memories, they have an incentive to trust the long-run player even after the long-run player has lost his reputation. This encourages the long-run player to shirk occasionally instead of sustaining his reputation for exerting high effort. When the short-run players have short memories, the long-run player has a strong incentive to sustain his reputation since he is guaranteed to be punished after he loses his reputation. The welfare implication of my result is that longer memories may increase the seller's equilibrium payoff at the expense of consumers' welfare.

**Reputation with Limited Memories:** The short-run players in my model can only observe the summary statistics of the long-run player's actions in the last  $K$  periods. Similar to existing models with limited memories, such as Liu (2011), Liu and Skrzypacz (2014), and Pei (2022), the long-run player needs to sustain his reputation since the short-run players forget what happened in the distant past.<sup>18</sup> In contrast to my model, those models assume that the uninformed players *can* observe the order of the informed player's actions and the patient player's best reply may not satisfy the *no-back-loop property*.

Kaya and Roy (2022) study a model with limited memories in which a seller has persistent private information about his quality and decides whether to sell in each period. Consumers' willingness to pay depends only on the seller's type, and each of them observes whether the seller sold his product in each of the last  $K$  periods. Longer memories have an ambiguous effect on welfare, since on the one hand, they

---

<sup>18</sup>The long-run player also needs to sustain his reputation when he is replaced with positive probability after each period, in which case his opponents put a low weight on his past behaviors. See for example, Mailath and Samuelson (2001).

encourage low-quality sellers to imitate high-quality sellers, which makes screening harder; and on the other hand, they increase the number of consumers who can benefit from the information obtained from screening.

By contrast, I show that longer memories *undermine* the strategic seller's incentive to imitate the commitment type, instead of *encouraging* him to imitate the high-quality type. My model differs from Kaya and Roy (2022)'s model since the short-run player's best reply depends only on the long-run player's action.

Pei (2022) constructs an equilibrium in which the patient seller receives his minmax payoff when every consumer observes *at least one previous consumer's action* as well as a bounded number of the seller's past actions. By contrast, the current paper assumes that consumers *cannot* observe previous consumers' actions and *cannot* observe the order with which the seller took his actions. I show that the patient seller can secure his commitment payoff in all equilibria irrespective of consumers' memory length. I also provide conditions under which the patient seller plays his commitment action with frequency close to one in *all* equilibria.

**Sustaining Cooperation under Limited Information:** My paper contributes to the discussions on the sustainability of cooperation when players have limited information about their opponents' past behaviors. This question has been asked in the literature on repeated games with anonymous random matching such as Kandori (1992), Ellison (1994), Takahashi (2010), Deb (2020), Clark, Fudenberg and Wolitzky (2021), and Sugaya and Wolitzky (2020, 2021). These papers study repeated games where all players are long-lived and provide conditions on the monitoring technology under which players can cooperate in some equilibria.

More closely related is the work of Bhaskar and Thomas (2019), which I have already discussed in the introduction, as well as the works of Ekmekci (2011) and Heller and Mohlin (2018). Ekmekci (2011) studies a repeated product choice game where the seller's cost of effort is independent of the consumer's action. He constructs a rating system under which there exist equilibria where the patient seller obtains his optimal commitment payoff, and another rating system where the consumers receive a high payoff. By contrast, I focus on stage games where payoffs are *strictly supermodular*. I examine the effects of memory length on the patient player's incentives to take his commitment action. I provide conditions under which the patient player takes his commitment action with frequency close to one in *all* equilibria.

Heller and Mohlin (2018) study repeated prisoner's dilemma game with anonymous random matching where every player is committed with positive probability and can sample a finite number of his opponent's past actions. They show that players can cooperate in *some* equilibria when their actions are strategic complements but cannot cooperate in any equilibrium when their actions are strategic substitutes.

By contrast, I study a reputation game between an informed long-run player and a sequence of uninformed short-run players. I focus on the effects of the short-run players' *memory length* on the long-run

player's returns from building reputations as well as his incentives to sustain his reputation. My results examine whether the long-run player will sustain his reputation in *all* equilibria, instead of examining whether cooperation is feasible in *some* equilibria. Focusing on games where players' actions are strategic complements, I show that the longer memories undermine the long-run player's incentive to sustain his reputation since they undermine the short-run players' incentives to punish the strategic-type long-run player.

**Reputation Sustainability:** My paper contributes to the discussions on the sustainability of reputations.

In contrast to Cripps, Mailath and Samuelson (2004)'s results that focus on the patient player's reputation and behavior as  $t \rightarrow +\infty$  and examine whether reputations can be sustained in *some* equilibria, my analysis focuses on the *discounted frequency of players' actions* and my result provides conditions under which the patient player plays his commitment action with frequency close to one in *all* equilibria. Compared to the existing results that focus on the patient player's reputation and behavior in the  $t \rightarrow +\infty$  limit, my result on the *discounted* action frequency has a comparative advantage in evaluating consumers' welfare.

Pei (2020) and Ekmekci and Maestri (2022) study reputation models with unbounded memories and interdependent values, that is, the patient player's type directly affects his opponents' payoffs. They provide sufficient conditions under which the patient player sustains his reputation in all equilibria. Similar to my model, the patient player is guaranteed to receive a low payoff after he abandons his reputation.

The mechanisms behind their results are different from mine. The patient player is guaranteed to be punished in their interdependent value settings since deviating from the commitment action signals negative information about the payoff-relevant state. By contrast, the current paper studies a private-value reputation model in which the uninformed players receive coarse information about the patient player's past records. The patient player is guaranteed to be punished after he milks his reputation since the uninformed players *cannot* fine-tune their punishments based on the game's history. As a result, the punishments needed to sustain cooperation inevitably punishes the patient player at other histories, and harsh punishments at a larger set of histories encourage the patient player to sustain his reputation.

**Behavioral Reputation Models:** Jehiel and Samuelson (2012) study a reputation model where the short-run players have *infinite memories*, but they mistakenly believe that all types of the patient player use stationary strategies. Under such a misspecified belief, each short-run player's belief about the patient player's current-period action depends only on the frequencies with which the patient player took his actions. This feature resembles my model where the short-run players can only observe the summary statistics of the patient player's recent play, in which case their behaviors must be measurable with respect to this information.



In contrast to Jehiel and Samuelson (2012)'s model, the short-run players in my model understand that the patient player's behavior may depend on the game's history, e.g., his behavior when he has a positive reputation may differ from his behavior after he has lost his reputation. This assumption on the short-run players' sophistication yields different predictions in terms of the long-run player's behaviors. For example, in the monotone-supermodular product choice game, the patient player exerts high effort with frequency arbitrarily close to one in all equilibria when  $K$  is small. By contrast, in Jehiel and Samuelson (2012)'s model, the seller exerts high effort with frequency close to  $x \in (0, 1)$  in order to exploit the short-run players' misspecified beliefs. Therefore, my results imply that the short-run players' understanding of the game's non-stationarity can provide strong incentives for the long-run player to sustain reputations.

One potential concern of my model is that consumers need to make Bayesian inferences about the seller's current-period action based on the summary statistics, which can be quite complicated in some equilibria. However, my proofs suggest that my main results apply even when we focus on equilibria where consumers' inference problems are simple and intuitive on the equilibrium path. For example, in the equilibria I constructed in the proof of Theorem 2, consumers believe that the distribution of the seller's current-period action is close to his action frequencies in the last  $K$  periods.

## A Proof of the No-Back-Loop Lemma

For every  $t \geq K$ , player 2's incentive depends only on the number of times that player 1 takes each action in the last  $K$  periods. Therefore, player 1's continuation value and incentive in period  $t$  depend only on  $(a_{t-K}, \dots, a_{t-1})$ . Although the order of actions in the vector  $(a_{t-K}, \dots, a_{t-1})$  does not affect player 2's action, it can affect player 1's incentives. Moreover, player 1's action in period  $t$  may depend on variables other than  $(a_{t-K}, \dots, a_{t-1})$ , such as his actions more than  $K$  periods ago and previous player 2's actions.

Fix  $\sigma_2 : \mathcal{H}_2 \rightarrow \mathcal{B}^*$ . Let  $V(a_{t-K}, \dots, a_{t-1})$  be player 1's continuation value in period  $t$ . Let  $\beta^* \in \Delta(B)$  be player 2's action at histories that belong to  $\mathcal{H}_1^*$  under  $\sigma_2$ . For every  $a \neq a^*$ , let  $\beta(a)$  be player 2's action under  $\sigma_2$  when exactly one of player 1's last  $K$  actions was  $a$  and the other  $K - 1$  actions were  $a^*$ . A pure strategy  $\hat{\sigma}_1$  is *canonical* if it depends only on the last  $K$  actions of player 1's. For every strategy profile  $(\sigma_1, \sigma_2)$  and  $h^t \in \mathcal{H}_1$ , let  $\mathcal{H}_1(\sigma_1, \sigma_2 | h^t)$  be the set of histories  $h^s$  satisfying  $h^s \succ h^t$  and  $h^s$  occurring with positive probability when the game starts from history  $h^t$  and players use strategies  $(\sigma_1, \sigma_2)$ . If  $\hat{\sigma}_1$  is canonical, then  $\mathcal{H}_1(\hat{\sigma}_1, \sigma_2 | h^t) = \mathcal{H}_1(\hat{\sigma}_1, \sigma_2' | h^t)$  for every  $\sigma_2, \sigma_2'$ , and  $h^t$ .

Since player 2's action depends only on player 1's actions in the last  $K$  periods, for every  $\sigma_2$ , there exists a canonical pure strategy  $\hat{\sigma}_1$  that best replies to  $\sigma_2$ . Therefore, as long as there exists a pure strategy that

best replies to  $\sigma_2$  and violates the no-back-loop property with respect to  $\sigma_2$ , there also exists a canonical pure strategy the best replies to  $\sigma_2$  and violates the no-back-loop property with respect to  $\sigma_2$ . Hence, the no-back-loop lemma is implied by the following *no-back-loop lemma\**, which I show next.

**No-Back-Loop Lemma\*.** *For any  $\sigma_2 : \mathcal{H}_2 \rightarrow \mathcal{B}^*$  and any canonical pure strategy  $\hat{\sigma}_1$  that best replies to  $\sigma_2$ . If there exists  $h^t \in \mathcal{H}_1^* \cap \mathcal{H}_1(\hat{\sigma}_1, \sigma_2)$  such that  $\hat{\sigma}_1(h^t) \neq a^*$ , then  $\mathcal{H}_1(\hat{\sigma}_1, \sigma_2|h^t) \cap \mathcal{H}_1^* = \emptyset$ .*

Suppose by way of contradiction that there exists a canonical pure strategy  $\hat{\sigma}_1$  that best replies to  $\sigma_2$  such that there exist two histories  $h^t, h^s \in \mathcal{H}_1^* \cap \mathcal{H}_1(\hat{\sigma}_1, \sigma_2)$  that satisfy  $h^s \in \mathcal{H}_1(\hat{\sigma}_1, \sigma_2|h^t)$ , and  $\hat{\sigma}_1(h^t) = a'$  for some  $a' \neq a^*$ . Without loss of generality, let  $h^s$  be the first history in  $\mathcal{H}_1^*$  that succeeds  $h^t$  when player 1 behaves according to  $\hat{\sigma}_1$ . Let  $h^{s-1} \equiv (a_0, \dots, a_{s-2}) \in \mathcal{H}_1(\hat{\sigma}_1, \sigma_2|h^t)$ . Since  $h^s$  is the first history in  $\mathcal{H}_1^*$  that succeeds  $h^t$ , it must be the case that  $h^{s-1} \notin \mathcal{H}_1^*$ , so  $(a_{s-K-1}, \dots, a_{s-2}) = (a'', a^*, \dots, a^*)$  for some  $a'' \neq a^*$ . Since  $h^s \in \mathcal{H}_1^*$ , player 1 plays  $a^*$  at  $h^{s-1}$  when he uses strategy  $\hat{\sigma}_1$ . This implies that

$$(1 - \delta)u_1(a^*, \beta(a'')) + \delta V(a^*, a^*, \dots, a^*, a^*) \geq (1 - \delta)u_1(a', \beta(a'')) + \delta V(a^*, a^*, \dots, a^*, a'). \quad (\text{A.1})$$

Since  $\hat{\sigma}_1(h^t) = a'$ , player 1 weakly prefers  $a'$  to  $a^*$  at histories in  $\mathcal{H}_1^*$ , we have:

$$(1 - \delta)u_1(a^*, \beta^*) + \delta V(a^*, a^*, \dots, a^*, a^*) \leq (1 - \delta)u_1(a', \beta^*) + \delta V(a^*, a^*, \dots, a^*, a'). \quad (\text{A.2})$$

Since the seller's stage-game payoff function is strictly supermodular, and  $\beta^*$  and  $\beta(a'')$  can be ranked according to FOSD under Assumption 2, inequalities (A.1) and (A.2) imply that  $\beta^* \preceq_{FOSD} \beta(a'')$ . Let

$$U \equiv \frac{\sum_{\tau=t+1}^{s-2} \delta^{\tau-t} u_1(\hat{\sigma}_1(h^\tau), \sigma_2(h^\tau))}{\sum_{\tau=t+1}^{s-2} \delta^{\tau-t}}. \quad (\text{A.3})$$

be player 1's discounted average payoff from period  $t + 1$  to period  $s - 2$  when his period  $t$  history is  $h^t$  and players play according to  $(\hat{\sigma}_1, \sigma_2)$ . Since the strategic-type player 1's incentive depends only on his actions in the last  $K$  periods, when  $(a_{s-K-1}, \dots, a_{s-2}) = (a'', a^*, \dots, a^*)$ , the following strategy is optimal for him:

- **Strategy \*:** Play  $a^*$  in period  $s - 1$ , play  $a'$  in period  $s$ , play  $\hat{\sigma}_1(h^\tau)$  in period  $\tau + (s - t)$  for every  $\tau \in \{t + 1, \dots, s - 2\}$ , and play the same action that he has played  $s - t$  periods ago in every period after period  $2s - t - 1$ .

Since Strategy \* is optimal for player 1, it must yield a weakly greater payoff compared to any of the following two deviations starting from a period  $s - 1$  history where  $(a_{s-K-1}, \dots, a_{s-2}) = (a'', a^*, \dots, a^*)$ :

- **Deviation A:** Play  $a'$  in period  $s - 1$ ,  $\hat{\sigma}_1(h^\tau)$  in period  $\tau + (s - t - 1)$  for every  $\tau \in \{t + 1, \dots, s - 2\}$ , and play the same action that he has played  $s - t - 1$  periods ago in every period after  $2s - t - 2$ .
- **Deviation B:** Play  $a''$  in period  $s - 1$ , play  $a^*$  from period  $s$  to  $s + K - 2$ , and play the same action that he has played  $K$  periods ago in every period after  $s + K - 1$ .

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

$$\frac{(1 - \delta)u_1(a', \beta(a'')) + (\delta - \delta^{s-t-2})U}{1 - \delta^{s-t-2}} \leq \frac{(1 - \delta)u_1(a^*, \beta(a'')) + (1 - \delta)\delta u_1(a', \beta^*) + (\delta^2 - \delta^{s-t-1})U}{1 - \delta^{s-t-1}}$$

This leads to the following upper bound on  $U$ , defined in (A.3):

$$(\delta - \delta^{s-t-2})U \leq (1 - \delta^{s-t-2})u_1(a^*, \beta(a'')) + \delta(1 - \delta^{s-t-2})u_1(a', \beta^*) - (1 - \delta^{s-t-1})u_1(a', \beta(a'')). \quad (\text{A.4})$$

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

$$\frac{(1 - \delta)u_1(a'', \beta(a'')) + (\delta - \delta^K)u_1(a^*, \beta(a''))}{1 - \delta^K} \leq \frac{(1 - \delta)u_1(a^*, \beta(a'')) + (1 - \delta)\delta u_1(a', \beta^*) + (\delta^2 - \delta^{s-t-1})U}{1 - \delta^{s-t-1}}. \quad (\text{A.5})$$

This leads to a lower bound on  $U$ . The left-hand-side of (A.5) equals

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

and inequality (A.4) implies that the right-hand-side of (A.5) is no more than:

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

Since  $u_1(a, b)$  is strictly increasing in  $b$  and is strictly decreasing in  $a$ ,  $a^* \succ a''$ , and  $\beta(a'') \succeq \beta^*$ , inequality (A.5) cannot be true. This leads to a contradiction and implies the no-back-loop lemma.

## B Proof of Theorem 2: Part 1

Player 1's continuation value and incentive after period  $K$  depend only on his actions in the last  $K$  periods, including the order of these actions. We call  $S \equiv A^K$  the *state space*, with  $s \in S$  a typical *state*.

Fix any strategy profile  $\sigma$ . For every  $s \in S$ , let  $\mu(s)$  be the probability that the current-period state is

$s$  conditional on the event that player 1 is the strategic type and calendar time is at least  $K$ . For every pair of states  $s, s' \in S$ , let  $Q(s \rightarrow s')$  be the probability that the state in the next period is  $s'$  conditional on the state in the current period is  $s$ , player 1 is the strategic type, and the calendar time is at least  $K$ . Let  $p(s)$  be the probability that the state is  $s$  conditional on calendar time being  $K$  and player 1 is the strategic type.

**Lemma B.1.** *For any  $\delta \in (0, 1)$  and any equilibrium under  $\delta$ , we have*

$$\sum_{s \in S} \mu(s)Q(s \rightarrow s^*) = \frac{1}{\delta} \left\{ \mu(s^*) - (1 - \delta)p(s^*) \right\} \text{ for every } s^* \in S. \quad (\text{B.1})$$

*Proof.* For every  $t \in \mathbb{N}$ , let  $p_t(s)$  be the probability that the state is  $s$  in period  $K + t$  conditional on player 1 being the strategic type, and let  $q_t(s \rightarrow s')$  be the probability that the state in period  $t + K + 1$  is  $s'$  conditional on the state being  $s$  in period  $t + K$  and player 1 being the strategic type. By definition,  $p_0(s) = p(s)$  and  $p_{t+1}(s) = \sum_{s' \in S} p_t(s')Q(s' \rightarrow s)$ . According to Bayes rule, we have

$$\mu(s) = \sum_{t=0}^{+\infty} (1 - \delta)\delta^t p_t(s) \quad \text{and} \quad Q(s \rightarrow s^*) = \frac{\sum_{t=0}^{+\infty} (1 - \delta)\delta^t p_t(s)q_t(s \rightarrow s^*)}{\sum_{t=0}^{+\infty} (1 - \delta)\delta^t p_t(s)}.$$

This implies that

$$\sum_{s \in S} \mu(s)Q(s \rightarrow s^*) = \sum_{s \in S} \sum_{t=0}^{+\infty} (1 - \delta)\delta^t p_t(s)q_t(s \rightarrow s^*) = \sum_{t=0}^{+\infty} (1 - \delta)\delta^t \sum_{s \in S} p_t(s)q_t(s \rightarrow s^*) = \sum_{t=0}^{+\infty} (1 - \delta)\delta^t p_{t+1}(s^*)$$

and

$$\frac{1}{\delta} \left\{ \mu(s^*) - (1 - \delta)p(s^*) \right\} = \frac{1}{\delta} \left\{ \sum_{t=0}^{+\infty} (1 - \delta)\delta^t p_t(s^*) - (1 - \delta)p(s^*) \right\} = \sum_{t=0}^{+\infty} (1 - \delta)\delta^t p_{t+1}(s^*).$$

These two equations together imply (B.1). □

For any non-empty  $S' \subset S$ , let

$$\mathcal{I}(S') \equiv \sum_{s' \in S'} \sum_{s \notin S'} \mu(s)Q(s \rightarrow s') \quad (\text{B.2})$$

be the *inflow* to  $S'$  from states that do not belong to  $S'$ , and let

$$\mathcal{O}(S') \equiv \sum_{s' \in S'} \sum_{s \notin S'} \mu(s')Q(s' \rightarrow s) \quad (\text{B.3})$$

be the *outflow* from  $S'$  to states that do not belong to  $S'$ . Notice that  $\mathcal{I}(S')$  and  $\mathcal{O}(S')$  depend on the equilibrium strategy profile, which I omit in order to avoid cumbersome notation. By definition,

$$\begin{aligned} \sum_{s' \in S'} \mu(s') &= \sum_{s' \in S'} \mu(s') \left( \underbrace{\sum_{s \in S'} Q(s' \rightarrow s) + \sum_{s \notin S'} Q(s' \rightarrow s)}_{=1} \right) \\ &= \sum_{s' \in S'} \sum_{s \in S'} \mu(s') Q(s' \rightarrow s) + \underbrace{\sum_{s' \in S'} \sum_{s \notin S'} \mu(s') Q(s' \rightarrow s)}_{\equiv \mathcal{O}(S')}. \end{aligned}$$

Applying equation (B.1) to any non-empty subset of states  $S' \subset S$ , we have

$$\sum_{s' \in S'} \sum_{s \in S'} \mu(s') Q(s' \rightarrow s) + \underbrace{\sum_{s' \in S'} \sum_{s \notin S'} \mu(s) Q(s \rightarrow s')}_{\equiv \mathcal{I}(S')} = \sum_{s' \in S'} \mu(s') + \sum_{s' \in S'} \frac{1-\delta}{\delta} \{ \mu(s') - p(s') \}.$$

**Lemma B.2.** *For every non-empty subset  $S' \subset S$ , we have:*

$$|\mathcal{I}(S') - \mathcal{O}(S')| = \left| \sum_{s' \in S'} \frac{1-\delta}{\delta} (\mu(s') - p(s')) \right| \leq \frac{1-\delta}{\delta}. \quad (\text{B.4})$$

Let  $s^* \in S$  be the state where all of the last  $K$  actions were  $a^*$ , i.e., the only state where player 1 has a positive reputation. Lemma B.2 and the no-back-loop lemma imply the following lemma:

**Lemma B.3.** *For every  $\delta$  and in every equilibrium under  $\delta$ ,  $\mathcal{O}(\{s^*\}) \leq \frac{2(1-\delta)}{\delta}$  and  $\mathcal{I}(\{s^*\}) \leq \frac{1-\delta}{\delta}$ .*

*Proof.* Let  $S'$  be a subset of states such that  $s \in S'$  if and only if (i)  $s \neq s^*$ , and (ii) there exists a pure-strategy best reply  $\hat{\sigma}_1$  such that  $s^*$  is reached within a finite number of periods when the initial state is  $s$  and player 1 uses strategy  $\hat{\sigma}_1$ . The no-back-loop lemma implies that at least one of the two statements is true:

1. Player 1 has no incentive to play actions other than  $a^*$  at  $s^*$ .
2. Player 1 has an incentive to play actions other than  $a^*$  at  $s^*$ , and as long as player 1 plays any such best reply, the state never reaches  $S'$  when the initial state is  $s^*$ .

In the first case,  $\mathcal{O}(\{s^*\}) = 0$ , and Lemma B.2 implies that  $\mathcal{I}(\{s^*\}) \leq \frac{1-\delta}{\delta}$ . In the second case, the definition of  $S'$  implies that  $\mathcal{I}(S') = 0$ . According to Lemma B.2,  $\mathcal{O}(S') \leq \frac{1-\delta}{\delta}$ . The definition of  $S'$  implies that  $\mathcal{I}(\{s^*\}) \leq \mathcal{O}(S')$ , so  $\mathcal{I}(\{s^*\}) \leq \frac{1-\delta}{\delta}$ . According to Lemma B.2,  $\mathcal{O}(\{s^*\}) \leq \frac{2(1-\delta)}{\delta}$ .  $\square$

For every  $k \in \{0, 1, \dots, K\}$ , let  $S_k \subset S$  be the subset of states such that  $k$  of the last  $K$  actions were *not*  $a^*$ . By definition,  $S = \bigcup_{k=0}^K S_k$  and  $S_0 = \{s^*\}$ . I partition every  $S_k$  into  $S_k = S_{1,k} \cup \dots \cup S_{j(k,|A|),k}$  such that  $s, s'$  belong to the same partition element *if and only if* player 2 cannot distinguish between  $s$  and  $s'$ . The number of elements in the partition of  $S_k$  depends on  $k$  and the cardinality of player 1's action set  $A$ . I use  $S_{j,k}$  to denote a typical partition element. For every state that belongs to  $S_k$ , exactly one of the following two statements is true, depending on whether player 1's action  $K$  periods before was  $a^*$ :

1. The state in the next period belongs to  $S_{k-1}$  or  $S_k$ , depending on player 1's current-period action.
2. The state in the next period belongs to  $S_k$  or  $S_{k+1}$ , depending on player 1's current-period action.

Therefore, I partition  $S_{j,k}$  into  $S_{j,k}^*$  and  $S'_{j,k}$ , such that for every  $s \in S_{j,k}$ ,  $s \in S_{j,k}^*$  if and only if player 1's action  $K$  periods before was  $a^*$ , and  $s \in S'_{j,k}$  otherwise. For any  $S', S'' \subset S$  that satisfy  $S' \cap S'' = \emptyset$ , let

$$\mathcal{Q}(S' \rightarrow S'') \equiv \sum_{s' \in S'} \sum_{s'' \in S''} \mu(s') \mathcal{Q}(s' \rightarrow s'') \quad (\text{B.5})$$

be the expected flow from  $S'$  to  $S''$ . By definition,  $\mathcal{Q}(S_{j,k}^* \rightarrow S_i) = 0$  for every  $i \leq k-1$  and every  $i \geq k+2$ , and  $\mathcal{Q}(S'_{j,k} \rightarrow S_i) = 0$  for every  $i \geq k+1$  and  $i \leq k-2$ . The operator  $\mathcal{Q}$  is additive in the sense that for any  $S_1, S_2, S_3 \subset S$  such that  $S_1 \cap S_2, S_1 \cap S_3, S_2 \cap S_3 = \emptyset$ , we have

$$\mathcal{Q}(S_1 \rightarrow S_3) + \mathcal{Q}(S_2 \rightarrow S_3) = \mathcal{Q}(S_1 \cup S_2 \rightarrow S_3) \quad (\text{B.6})$$

$$\mathcal{Q}(S_1 \rightarrow S_2) + \mathcal{Q}(S_1 \rightarrow S_3) = \mathcal{Q}(S_1 \rightarrow S_2 \cup S_3). \quad (\text{B.7})$$

Upon observing a state that belongs to  $S_{j,k}$ , player 2 believes that player 1's action is  $a^*$  with probability:

$$\mathcal{Q}(S_{j,k} \rightarrow S_{k-1}) + \sum_{s \in S_{j,k}^*} \sum_{s' \in S_{j,k}} \mu(s) \mathcal{Q}(s \rightarrow s'), \quad (\text{B.8})$$

and player 1's action is not  $a^*$  with probability:

$$\mathcal{Q}(S_{j,k} \rightarrow S_{k+1}) + \sum_{s \in S'_{j,k}} \sum_{s' \in S_k} \mu(s) \mathcal{Q}(s \rightarrow s'). \quad (\text{B.9})$$

I show that when  $\mathcal{Q}(S_{k-1} \rightarrow S_{j,k})$  and  $\mathcal{Q}(S_{j,k} \rightarrow S_{k-1})$  are bounded above by a linear function of  $1 - \delta$ , either  $\sum_{s \in S_{j,k}} \mu(s)$  is also bounded above by some linear function of  $1 - \delta$ , or the probability that player 1 plays  $a^*$  at  $S_{k,j}$  is no more than  $\frac{K-1}{K}$ . In the second case, player 2 has no incentive to play  $b^*$  at  $S_{j,k}$ .

**Lemma B.4.** For every  $z \in \mathbb{R}_+$ , there exist  $y \in \mathbb{R}_+$  and  $\underline{\delta} \in (0, 1)$  such that for every equilibrium under  $\delta > \underline{\delta}$  and every  $S_{j,k}$  with  $k \geq 1$ . If  $\max\{\mathcal{Q}(S_{j,k} \rightarrow S_{k-1}), \mathcal{Q}(S_{k-1} \rightarrow S_{j,k})\} \leq z(1 - \delta)$ , then either

$$\sum_{s \in S_{j,k}} \mu(s) \leq y(1 - \delta) \quad (\text{B.10})$$

or

$$\frac{\mathcal{Q}(S_{j,k} \rightarrow S_{k-1}) + \sum_{s \in S_{j,k}^*} \sum_{s' \in S_{j,k}} \mu(s) \mathcal{Q}(s \rightarrow s')}{\mathcal{Q}(S_{j,k} \rightarrow S_{k+1}) + \sum_{s \in S'_{j,k}} \sum_{s' \in S_k} \mu(s) \mathcal{Q}(s \rightarrow s')} < K - 1. \quad (\text{B.11})$$

*Proof.* Since  $\mathcal{Q}(S_{j,k} \rightarrow S_{k-1}) = \mathcal{Q}(S'_{j,k} \rightarrow S_{k-1})$ ,  $\mathcal{Q}(S_{j,k} \rightarrow S_{k+1}) = \mathcal{Q}(S_{j,k}^* \rightarrow S_{k+1})$ , and under the hypothesis that  $\mathcal{Q}(S_{j,k} \rightarrow S_{k-1}) \leq z(1 - \delta)$  and  $\mathcal{Q}(S_{k-1} \rightarrow S_{j,k}) \leq z(1 - \delta)$ , we have:

$$\underbrace{\sum_{s \in S_{j,k}^*} \sum_{s' \in S_{j,k}} \mu(s) \mathcal{Q}(s \rightarrow s')}_{=\sum_{s \in S_{j,k}^*} \mu(s) - \mathcal{Q}(S_{j,k}^* \rightarrow S_{k+1})} + \underbrace{\mathcal{Q}(S_{j,k} \rightarrow S_{k-1})}_{\leq z(1-\delta)} \leq \sum_{s \in S_{j,k}^*} \mu(s) - \mathcal{Q}(S_{j,k}^* \rightarrow S_{k+1}) + z(1 - \delta),$$

$$\underbrace{\sum_{s \in S'_{j,k}} \sum_{s' \in S_k} \mu(s) \mathcal{Q}(s \rightarrow s')}_{=\sum_{s \in S'_{j,k}} \mu(s) - \mathcal{Q}(S'_{j,k} \rightarrow S_{k-1})} + \underbrace{\mathcal{Q}(S_{j,k} \rightarrow S_{k+1})}_{=\mathcal{Q}(S_{j,k}^* \rightarrow S_{k+1})} \geq \sum_{s \in S'_{j,k}} \mu(s) + \mathcal{Q}(S_{j,k}^* \rightarrow S_{k+1}) - z(1 - \delta).$$

Suppose there exists no such  $y \in \mathbb{R}_+$ , i.e.,  $\frac{\sum_{s \in S_{j,k}^*} \mu(s)}{z(1-\delta)}$  can be arbitrarily large as  $\delta \rightarrow 1$ . Since the sum of  $\sum_{s \in S_{j,k}^*} \mu(s) - \mathcal{Q}(S_{j,k}^* \rightarrow S_{k+1}) + z(1 - \delta)$  and  $\sum_{s \in S'_{j,k}} \mu(s) + \mathcal{Q}(S_{j,k}^* \rightarrow S_{k+1}) - z(1 - \delta)$  equals  $\sum_{s \in S_{j,k}} \mu(s)$ , we know that when  $\delta$  is close to 1, (B.11) is implied by:

$$\frac{\sum_{s \in S_{j,k}^*} \mu(s) - \mathcal{Q}(S_{j,k}^* \rightarrow S_{k+1})}{\sum_{s \in S'_{j,k}} \mu(s) + \mathcal{Q}(S_{j,k}^* \rightarrow S_{k+1})} < K - 1,$$

or equivalently,

$$\sum_{s \in S_{j,k}^*} \mu(s) < (K - 1) \sum_{s \in S'_{j,k}} \mu(s) + K \mathcal{Q}(S_{j,k}^* \rightarrow S_{k+1}). \quad (\text{B.12})$$

I derive a lower bound for  $\mathcal{Q}(S_{j,k}^* \rightarrow S_{k+1})$ . Since  $\mathcal{O}(S_{j,k}) = \mathcal{Q}(S_{j,k}^* \rightarrow S_{k+1}) + \mathcal{Q}(S'_{j,k} \rightarrow S_{k-1}) + \mathcal{Q}(S_{j,k} \rightarrow S_k \setminus S_{j,k})$ , under the hypothesis that  $\max\{\mathcal{Q}(S_{j,k} \rightarrow S_{k-1}), \mathcal{Q}(S_{k-1} \rightarrow S_{j,k})\} \leq z(1 - \delta)$ ,

$$\mathcal{Q}(S_{j,k}^* \rightarrow S_{k+1}) = \mathcal{O}(S_{j,k}) - \mathcal{Q}(S'_{j,k} \rightarrow S_{k-1}) - \mathcal{Q}(S_{j,k} \rightarrow S_k \setminus S_{j,k}) \geq \mathcal{O}(S_{j,k}) - \mathcal{Q}(S_{j,k} \rightarrow S_k \setminus S_{j,k}) - z(1 - \delta).$$

According to Lemma B.2, we have  $\mathcal{Q}(S_{j,k}^* \rightarrow S_{k+1}) \geq \mathcal{I}(S_{j,k}) - \mathcal{Q}(S_{j,k} \rightarrow S_k \setminus S_{j,k}) - \frac{(1-\delta)(1+z\delta)}{\delta}$ . Since at every  $s \in S_{j,k}^*$ , the state in the next period belongs to  $S_{k+1}$  if player 1 does not play  $a^*$  at  $s$ , and belongs to  $S_{j,k}$  if player 1 plays  $a^*$  at  $s$ , we have  $\mathcal{Q}(S_{j,k}^* \rightarrow S_k \setminus S_{j,k}) = 0$ . This implies that  $\mathcal{Q}(S_{j,k} \rightarrow S_k \setminus S_{j,k}) = \mathcal{Q}(S'_{j,k} \rightarrow S_k \setminus S_{j,k})$ . Since  $\mathcal{I}(S_{j,k}) = \mathcal{I}(S_{j,k}^*) + \mathcal{I}(S'_{j,k}) - \mathcal{Q}(S_{j,k}^* \rightarrow S'_{j,k}) - \mathcal{Q}(S'_{j,k} \rightarrow S_{j,k}^*)$ ,

$$\begin{aligned} \mathcal{Q}(S_{j,k}^* \rightarrow S_{k+1}) &\geq \mathcal{I}(S_{j,k}) - \mathcal{Q}(S_{j,k} \rightarrow S_k \setminus S_{j,k}) - \frac{(1-\delta)(1+z\delta)}{\delta} \\ &= \mathcal{I}(S_{j,k}^*) + \mathcal{I}(S'_{j,k}) - \mathcal{Q}(S_{j,k}^* \rightarrow S'_{j,k}) - \mathcal{Q}(S'_{j,k} \rightarrow S_{j,k}^*) - \mathcal{Q}(S'_{j,k} \rightarrow S_k \setminus S_{j,k}) - \frac{(1-\delta)(1+z\delta)}{\delta} \\ &\geq \mathcal{I}(S_{j,k}^*) - \mathcal{Q}(S_{j,k}^* \rightarrow S'_{j,k}) + \underbrace{\mathcal{O}(S'_{j,k}) - \mathcal{Q}(S'_{j,k} \rightarrow S_k \setminus S_{j,k}) - \mathcal{Q}(S'_{j,k} \rightarrow S_{j,k}^*)}_{\geq 0} - \frac{(1-\delta)(2+z\delta)}{\delta} \\ &\geq \mathcal{I}(S_{j,k}^*) - \mathcal{Q}(S_{j,k}^* \rightarrow S'_{j,k}) - \frac{(1-\delta)(2+z\delta)}{\delta} \end{aligned}$$

Since  $\mathcal{Q}(S_{j,k}^* \rightarrow S'_{j,k}) \leq \mathcal{O}(S_{j,k}^*)$  and  $\mathcal{Q}(S_{j,k}^* \rightarrow S'_{j,k}) \leq \mathcal{I}(S'_{j,k})$ ,

$$\mathcal{Q}(S_{j,k}^* \rightarrow S'_{j,k}) \leq \frac{1}{K}\mathcal{O}(S_{j,k}^*) + \frac{K-1}{K}\mathcal{I}(S'_{j,k}) \leq \frac{1}{K}\mathcal{I}(S_{j,k}^*) + \frac{K-1}{K}\mathcal{O}(S'_{j,k}) + \frac{1-\delta}{\delta}.$$

This together with the lower bound on  $\mathcal{Q}(S_{j,k}^* \rightarrow S_{k+1})$  that we derived earlier implies that:

$$\mathcal{Q}(S_{j,k}^* \rightarrow S_{k+1}) \geq \frac{K-1}{K} \left( \mathcal{I}(S_{j,k}^*) - \mathcal{O}(S'_{j,k}) \right) - \frac{(1-\delta)(3+z\delta)}{\delta}.$$

Since  $\mathcal{O}(S'_{j,k}) = \mathcal{Q}(S'_{j,k} \rightarrow S_k \setminus S_{j,k}) + \mathcal{Q}(S'_{j,k} \rightarrow S_{j,k}^*) + \mathcal{Q}(S'_{j,k} \rightarrow S_{k-1})$ , and  $\mathcal{Q}(S'_{j,k} \rightarrow S_{k-1})$  is assumed to be less than  $z(1-\delta)$ , we know that

$$\mathcal{Q}(S_{j,k}^* \rightarrow S_{k+1}) \geq \frac{K-1}{K} \left( \mathcal{I}(S_{j,k}^*) - \mathcal{Q}(S'_{j,k} \rightarrow S_k \setminus S_{j,k}) - \mathcal{Q}(S'_{j,k} \rightarrow S_{j,k}^*) \right) - \frac{(1-\delta)(3+2z\delta)}{\delta}.$$

Hence, when  $\delta$  is close to 1, inequality (B.12) is implied by

$$\begin{aligned} \sum_{s \in S_{j,k}^*} \mu(s) &< (K-1) \left\{ \sum_{s \in S'_{j,k}} \mu(s) - \mathcal{Q}(S'_{j,k} \rightarrow S_k \setminus S_{j,k}) \right\} \\ &\quad + (K-1) \left\{ \mathcal{I}(S_{j,k}^*) - \mathcal{Q}(S'_{j,k} \rightarrow S_{j,k}^*) \right\}. \end{aligned} \tag{B.13}$$

I say that a sequence of states  $\{s_1, \dots, s_j\} \subset S_{j,k}$  form a *connected sequence* if for every  $i \in \{1, 2, \dots, i-1\}$ , there exists  $a_i \in A$  such that the state in the next period is  $s_{i+1}$  when the state in the current period is  $s_i$  and player 1 takes action  $a_i$ . A useful observation is that for every  $s \in S_{j,k}$ , there exists a unique state  $s'$  in



$S_{j,k}$  such that playing some  $a \in A$  in state  $s'$  leads to state  $s$ . Using this observation, we construct for any  $s_1 \in S'_{j,k}$ , a finite sequence of states  $\{s_1, \dots, s_m\} \subset S_{j,k}$  with length  $m$  at least one such that (i) for every  $i \in \{1, 2, \dots, m-1\}$ , there exists an action  $a_i \in A$  such that playing  $a_i$  in state  $s_i$  leads to state  $s_{i+1}$  in the next period, (ii) if  $m \geq 2$ , then  $\{s_2, \dots, s_m\} \subset S_{j,k}^*$ , and (iii) no matter which action player 1 takes in state  $s_m$ , the state in the next period does not belong to  $S_{j,k}^*$ , or equivalently, there exists an action  $a \in A$  such that taking action  $a$  at state  $s_m$  leads to a state that belongs to  $S'_{j,k}$ . Lemma B.1 implies that

$$\begin{aligned} \mu(s_2) &\leq \mu(s_1)Q(s_1 \rightarrow s_2) + \mathcal{Q}(S \setminus S_{j,k} \rightarrow \{s_2\}) + \frac{1-\delta}{\delta} \\ &= \mu(s_1) - \mathcal{Q}(\{s_1\} \rightarrow S_k \setminus S_{j,k}) + \mathcal{Q}(S \setminus S_{j,k} \rightarrow \{s_2\}) + \frac{1-\delta}{\delta}, \end{aligned} \quad (\text{B.14})$$

and for every  $i \geq 2$ , we have:

$$\begin{aligned} \mu(s_{i+1}) &\leq \mu(s_i)Q(s_i \rightarrow s_{i+1}) + \mathcal{Q}(S \setminus S_{j,k} \rightarrow \{s_{i+1}\}) + \frac{1-\delta}{\delta} \\ &= \mu(s_i) + \mathcal{Q}(S \setminus S_{j,k} \rightarrow \{s_{i+1}\}) + \frac{1-\delta}{\delta}. \end{aligned} \quad (\text{B.15})$$

Iteratively apply (B.15) and (B.14) for every  $i \geq 2$ , we obtain:

$$\mu(s_i) \leq \mu(s_1) + \mathcal{Q}(S \setminus S_{j,k} \rightarrow \{s_2, \dots, s_i\}) - \mathcal{Q}(\{s_1\} \rightarrow S_k \setminus S_{j,k}) + \frac{(1-\delta)(i-1)}{\delta}. \quad (\text{B.16})$$

Summing up inequality (B.16) for  $i \in \{2, \dots, m\}$ , we obtain:

$$\begin{aligned} \sum_{i=2}^m \mu(s_i) &\leq (m-1) \left\{ \mu(s_1) + \mathcal{Q}(S_{j,k}^c \rightarrow \{s_2, \dots, s_m\}) - \mathcal{Q}(\{s_1\} \rightarrow S_k \setminus S_{j,k}) \right\} + \frac{m(m-1)(1-\delta)}{2\delta} \\ &= (m-1) \left\{ \underbrace{\mu(s_1) - \mathcal{Q}(\{s_1\} \rightarrow S_k \setminus S_{j,k})}_{\geq -\frac{1-\delta}{\delta}} \right\} \\ &\quad + (m-1) \left\{ \underbrace{\mathcal{Q}(S \setminus S_{j,k}^* \rightarrow \{s_2, \dots, s_m\}) - \mathcal{Q}(S'_{j,k} \rightarrow \{s_2, \dots, s_m\})}_{\geq 0} \right\} + \frac{m(m-1)(1-\delta)}{2\delta} \\ &\leq (K-1) \left\{ \mu(s_1) - \mathcal{Q}(\{s_1\} \rightarrow S_k \setminus S_{j,k}) \right\} \\ &\quad + (K-1) \left\{ \mathcal{Q}(S \setminus S_{j,k}^* \rightarrow \{s_2, \dots, s_m\}) - \mathcal{Q}(S'_{j,k} \rightarrow \{s_2, \dots, s_m\}) \right\} + \left\{ \frac{m(m-1)(1-\delta)}{2\delta} + \frac{K(1-\delta)}{\delta} \right\}. \end{aligned}$$

One can obtain (B.13) by summing up the above equation for every  $s_1 \in S'_{j,k}$  and taking  $\delta \rightarrow 1$ . This is because the left-hand-side of this sum equals  $\sum_{s \in S_{j,k}^*} \mu(s)$ , and after ignoring the last term that vanishes to

0 as  $\delta \rightarrow 1$ , and uses the additive property of the operator  $\mathcal{Q}$  in (B.6) and (B.7), the right-hand-side equals

$$(K-1) \left\{ \sum_{s \in S'_{j,k}} \mu(s) - \mathcal{Q}(S'_{j,k} \rightarrow S_k \setminus S_{j,k}) \right\} + (K-1) \left\{ \mathcal{I}(S_{j,k}^*) - \mathcal{Q}(S'_{j,k} \rightarrow S_{j,k}^*) \right\}.$$

□

Since  $b^*$  does not best reply to  $\frac{K-1}{K}a^* + \frac{1}{K}a'$  for every  $a' \neq a^*$ , Lemma B.4 implies that when  $\mathcal{Q}(S_{k-1} \rightarrow S_{j,k})$  and  $\mathcal{Q}(S_{j,k} \rightarrow S_{k-1})$  are both bounded above by a linear function of  $1 - \delta$ , then either  $\sum_{s \in S_{j,k}} \mu(s)$  is also bounded above by a linear function of  $1 - \delta$  or  $b^*$  is strictly suboptimal for player 2 when she observes that the state belongs to  $S_{j,k}$ .

**Lemma B.5.** *Suppose  $b^*$  does not best reply to  $\frac{K-1}{K}a^* + \frac{1}{K}a'$  for every  $a' \neq a^*$ . For every  $y > 0$ , there exists  $z > 0$  and  $\underline{\delta} \in (0, 1)$  such that for every  $\delta > \underline{\delta}$ , every equilibrium under  $\delta$ , and every  $k \in \{1, 2, \dots, K\}$ . If  $\max\{\mathcal{Q}(S_{k-1} \rightarrow S_k), \mathcal{Q}(S_k \rightarrow S_{k-1})\} < y(1 - \delta)$ , then  $\sum_{s \in S_k} \mu(s) \leq z(1 - \delta)$ .*

*Proof.* Suppose by way of contradiction that for every  $y > 0$  and  $\underline{\delta} \in (0, 1)$ , there exist  $\delta > \underline{\delta}$ , an equilibrium under  $\delta$ , and  $k \geq 1$ , such that in this equilibrium,  $\max\{\mathcal{Q}(S_{k-1} \rightarrow S_k), \mathcal{Q}(S_k \rightarrow S_{k-1})\} < y(1 - \delta)$  but  $\sum_{s \in S_k} \mu(s) > z(1 - \delta)$ . Pick a large enough  $z$ , Lemma B.4 implies that for every partition element  $S_{j,k} \subset S_k$ , either  $\sum_{s \in S_{j,k}} \mu(s) < \frac{z}{2K}(1 - \delta)$ , or player 2 has a strict incentive not to play  $a^*$  at  $S_{j,k}$ . The hypothesis that  $\sum_{s \in S_k} \mu(s) > z(1 - \delta)$  implies that there exists at least one partition element  $S_{j,k}$  such that player 2 has a strict incentive not to play  $a^*$  at  $S_{j,k}$ . Let  $S'_k$  be the union of such partition elements.

I start from deriving an upper bound on the ratio between  $\sum_{s \in S'_k} \mu(s)$  and  $\mathcal{Q}(S'_k \rightarrow S_{k-1})$ . Let  $V(s)$  be player 1's continuation value in state  $s$  and let  $\bar{V} \equiv \max_{s \in S} V(s)$ . Let  $\underline{v}$  be player 1's lowest stage-game payoff. Let  $v' \equiv \max_{a \in A, b < b^*} u_1(a, b)$  and  $v^* \equiv u_1(a^*, b^*)$ . Assumptions 1 and 3 imply that  $v^* > v' > \underline{v}$ . Since player 1 can reach any state within  $K$  periods, we have  $V(s) \geq (1 - \delta^K)\underline{v} + \delta^K\bar{V}$  for every  $s \in S$ . Theorem 1 suggests that player 1's continuation value at  $s^*$  is at least  $u_1(a^*, b^*)$ . Therefore,  $\bar{V} \geq v^*$ . Let  $M$  be the largest integer  $m$  such that

$$(1 - \delta^m)v' + \delta^m\bar{V} \geq (1 - \delta^K)\underline{v} + \delta^K\bar{V}. \quad (\text{B.17})$$

Applying the L'Hospital Rule, (B.17) implies that when  $\delta$  is close to 1, we have  $M \leq K \frac{\bar{V} - \underline{v}}{\bar{V} - v'}$ . Therefore, for any  $t \in \mathbb{N}$  and  $s \in S'_k$ , and under any pure-strategy best reply of player 1, if the state is  $s$  in period  $t$ , then there exists  $\tau \in \{t + 1, \dots, t + M\}$  such that when player 1 uses this pure-strategy best reply, the state in period  $\tau$  does not belong to  $S'_k$ . Therefore,  $\frac{\sum_{s \in S'_k} \mu(s)}{\mathcal{Q}(S'_k)} \leq \frac{1 - \delta^M}{\delta^M(1 - \delta)}$ . When  $\delta \rightarrow 1$ , the RHS of the above

inequality converges to  $M$ , which implies that

$$\sum_{s \in S'_k} \mu(s) \leq K \cdot \frac{\bar{V} - v}{\bar{V} - v'} \cdot \mathcal{O}(S'_k). \quad (\text{B.18})$$

Since  $\sum_{s \in S_k \setminus S'_k} \mu(s)$  is bounded above by some linear function of  $1 - \delta$ , it must be the case that  $\mathcal{Q}(S_{k-1} \rightarrow S'_k) \geq \frac{\sum_{s \in S'_k} \mu(s)}{2M}$ . This implies that there exists  $s \in S_{k+1}$  and a canonical pure best reply  $\hat{\sigma}_1$  such that:

1. the state in the next period, denoted by  $s'$ , belongs to  $S'_k$ , and the state belongs to  $S'_k$  for  $m$  periods,
2. the state returns to  $S_{k+1}$  after these  $m$  periods, returns to  $s$  after a finite number of periods, and the state never reaches  $\cup_{n=0}^{k-1} S_n$  when play starts from  $s$ .

By definition, player 1 plays  $a^*$  in state  $s$  under  $\hat{\sigma}_1$  and  $\hat{\sigma}_1$  induces a cycle of states. Moreover, it is without loss of generality to focus on best replies that induce a cycle where each state occurs at most once.

I show that  $m \leq K - 1$ . Suppose by way of contradiction that  $m \geq K$ , namely, after reaching state  $s'$ , the state belongs to  $S'_k$  for at least  $K$  periods under player 1's pure-strategy best reply  $\hat{\sigma}_1$ . Recall the definition of a *minimal connected sequence*. Every minimal connected sequence contains either one state (if  $k = K$ ) or  $K$  states in category  $k$ . Therefore, the category  $k$  state after  $K$  periods is also  $s'$ . As a result, there exists a best-reply of player 1 such that under this best reply and starting from state  $s'$ , the state remains in category  $k$  forever. Due to the hypothesis that player 2 has no incentive to play  $b^*$  when the state belongs to  $S'_k$ , player 1's continuation value under such a best reply is at most  $v'$ , which is strictly less than his guaranteed continuation value  $(1 - \delta^K)v + \delta^K v^*$ . This contradicts the conclusion of Theorem 1.

Given that  $m \leq K - 1$ , let us consider an alternative strategy of player 1 under which he plays an action other than  $a^*$  in state  $s$ , then follows strategy  $\hat{\sigma}_1$ . Starting from state  $s$ , this strategy and  $\hat{\sigma}_1$  lead to the same state after  $m + 1$  periods. This strategy leads to a strictly higher payoff since the stage-game payoff at state  $s$  is strictly greater, and the payoffs after the first period are weakly greater. This contradicts the hypothesis that  $\hat{\sigma}_1$  is player 1's best reply to player 2's equilibrium strategy.  $\square$

In summary, Lemma B.3 implies that  $\max\{\mathcal{Q}(S_0 \rightarrow S_1), \mathcal{Q}(S_1 \rightarrow S_0)\} \leq \frac{2(1-\delta)}{\delta}$ . Lemma B.4 and Lemma B.5 together imply that  $\sum_{s \in S_1} \mu(s)$  is bounded from above by a linear function of  $1 - \delta$  given that  $\max\{\mathcal{I}(S_0), \mathcal{O}(S_0)\} \leq \frac{2(1-\delta)}{\delta}$ , which then implies that  $\mathcal{Q}(S_1 \rightarrow S_2)$  and  $\mathcal{Q}(S_2 \rightarrow S_1)$  are also bounded from above by a linear function of  $1 - \delta$ . Iteratively apply this argument, we obtain that for every  $k \in \{1, 2, \dots, K\}$ ,  $\sum_{s \in S_k} \mu(s)$  is bounded from above by a linear function of  $1 - \delta$ .

## References

- [1] Acemoglu, Daron, Ali Makhdoum, Azarakhsh Malekian and Asu Ozdaglar (2020) “Learning from Reviews: The Selection Effect and the Speed of Learning,” Working Paper.
- [2] Banerjee, Abhijit and Drew Fudenberg (2004) “Word-of-mouth Learning,” *Games and Economic Behavior*, 46, 1-22.
- [3] Bhaskar, V. and Caroline Thomas (2019) “Community Enforcement of Trust with Bounded Memory,” *Review of Economic Studies*, 86, 1010-1032.
- [4] Clark, Daniel, Drew Fudenberg and Alexander Wolitzky (2021) “Record-Keeping and Cooperation in Large Societies,” *Review of Economic Studies*, 88, 2179-2209.
- [5] Cripps, Martin, George Mailath and Larry Samuelson (2004) “Imperfect Monitoring and Impermanent Reputations,” *Econometrica*, 72, 407-432.
- [6] Cripps, Martin and Caroline Thomas (2019) “Strategic Experimentation in Queues,” *Theoretical Economics*, 14, 647-708.
- [7] Deb, Joyee (2020) “Cooperation and Community Responsibility,” *Journal of Political Economy*, 128, 1976-2009.
- [8] Echenique, Federico (2004) “Extensive-Form Games and Strategic Complementarities,” *Games and Economic Behavior*, 46, 348-364.
- [9] Ekmekci, Mehmet (2011) “Sustainable Reputations with Rating Systems,” *Journal of Economic Theory*, 146, 479-503.
- [10] Ekmekci, Mehmet, Olivier Gossner and Andrea Wilson (2012) “Impermanent Types and Permanent Reputations,” *Journal of Economic Theory*, 147, 162-178.
- [11] Ekmekci, Mehmet and Lucas Maestri (2022) “Wait or Act Now? Learning Dynamics in Stopping Games,” Working Paper.
- [12] Ellison, Glenn (1994) “Cooperation in the Prisoner’s Dilemma with Anonymous Random Matching,” *Review of Economic Studies*, 61, 567-588.
- [13] Ely, Jeffrey and Juuso Välimäki (2003) “Bad Reputation,” *Quarterly Journal of Economics*, 118, 785-814.
- [14] Fudenberg, Drew, David Kreps and Eric Maskin (1990) “Repeated Games with Long-Run and Short-Run Players,” *Review of Economic Studies*, 57, 555-573.
- [15] Fudenberg, Drew and David Levine (1989) “Reputation and Equilibrium Selection in Games with a Patient Player,” *Econometrica*, 57, 759-778.
- [16] Fudenberg, Drew and David Levine (1992) “Maintaining a Reputation when Strategies are Imperfectly Observed,” *Review of Economic Studies*, 59, 561-579.
- [17] Fudenberg, Drew and Jean Tirole (1991) “Perfect Bayesian Equilibrium and Sequential Equilibrium,” *Journal of Economic Theory*, 53, 236-260.
- [18] Gossner, Olivier (2011) “Simple Bounds on the Value of a Reputation,” *Econometrica*, 79, 1627-1641.

- [19] Heller, Yuval and Erik Mohlin (2018) “Observations on Cooperation,” *Review of Economic Studies*, 88, 1892-1935.
- [20] Hu, Ju (2020) “On the Existence of the Ex Post Symmetric Random Entry Model,” *Journal of Mathematical Economics*, 90, 42-47.
- [21] Jehiel, Philippe and Larry Samuelson (2012) “Reputation with Analogical Reasoning,” *Quarterly Journal of Economics*, 127(4), 1927-1970.
- [22] Kandori, Michihiro (1992) “Social Norms and Community Enforcement,” *Review of Economic Studies*, 59, 63-80.
- [23] Kaya, Ayça and Santanu Roy (2022) “Market Screening with Limited Records,” *Games and Economic Behavior*, 132, 106-132.
- [24] Li, Yingkai and Harry Pei (2021) “Equilibrium Behaviors in Repeated Games,” *Journal of Economic Theory*, 193, 105222.
- [25] Liu, Qingmin (2011) “Information Acquisition and Reputation Dynamics,” *Review of Economic Studies*, 78, 1400-1425.
- [26] Liu, Qingmin and Andrzej Skrzypacz (2014) “Limited Records and Reputation Bubbles,” *Journal of Economic Theory* 151, 2-29.
- [27] Mailath, George and Larry Samuelson (2001) “Who Wants a Good Reputation?” *Review of Economic Studies*, 68, 415-441.
- [28] Mailath, George and Larry Samuelson (2015) “Reputations in Repeated Games,” *Handbook of Game Theory with Economic Applications*, 165-238.
- [29] Niehaus, Paul (2011) “Filtered Social Learning,” *Journal of Political Economy*, 119, 686-720.
- [30] Pei, Harry (2020) “Reputation Effects under Interdependent Values,” *Econometrica*, 88(5), 2175-2202.
- [31] Pei, Harry (2022) “Reputation Building under Observational Learning,” *Review of Economic Studies*, forthcoming.
- [32] Quah, John and Bruno Strulovici (2012) “Aggregating the Single Crossing Property,” *Econometrica*, 80, 2333-2348.
- [33] Renault, Jérôme, Eilon Solan and Nicolas Vieille (2013) “Dynamic Sender-Receiver Games,” *Journal of Economic Theory*, 148, 502-534.
- [34] Sugaya, Takuo and Alexander Wolitzky (2020) “Do a Few Bad Apples Spoil the Barrel?: An Anti-Folk Theorem for Anonymous Repeated Games with Incomplete Information,” *American Economic Review*, 110, 3817-3835.
- [35] Sugaya, Takuo and Alexander Wolitzky (2021) “Communication and Community Enforcement,” *Journal of Political Economy*, 129, 2595-2628.
- [36] Takahashi, Satoru (2010) “Community Enforcement When Players Observe Partners’ Past Play,” *Journal of Economic Theory*, 145, 42-62.