Reputation Effects under Short Memories

Harry PEI*

February 10, 2024

Abstract: I analyze a reputation game between a patient player and a sequence of short-run players, who have limited memories and cannot observe the exact sequence of the patient player's actions. I focus on the case in which every short-run player only observes the number of times that the patient player took each of his actions in the last K periods. When players have monotone-supermodular payoffs, I show that the patient player can approximately secure his commitment payoff in all equilibria as long as K is at least one. I also show that the short-run players can approximately attain their highest feasible payoff in all equilibria *if and only if* K is lower than some cutoff. Although a larger K enables more short-run players to punish the patient player once he deviates from his commitment action, it weakens their incentives to execute the punishment.

Keywords: limited memory, coarse information, commitment payoff, equilibrium behavior.

1 Introduction

Economic agents benefit from good reputations. This idea is formalized in the reputation literature pioneered by Fudenberg and Levine (1989), who show that a patient player can secure a high payoff if he builds a reputation in front of a sequence of short-run players. These reputation results assume that the short-run players can observe either the *full history* of actions or *a sufficiently long history* of actions including the *exact sequence* of actions. The intuition is that when the short-run players observe that the patient player has taken an action for a long time, they will be convinced that he is likely to take that same action in the future.

In practice, people may not have long memories and may not know the exact sequence of other players' actions. For example, many online marketplaces disclose in a salient place the number of positive and negative reviews each merchant received in the last two weeks (e.g., the store-level ratings on TMall) or the last few months (e.g., eBay and EachNet). The *exact sequence* with which these reviews were received is either not disclosed, such as on eBay and EachNet (see Dellarocas 2006 and Tadelis 2016 for eBay and Cai, Jin, Liu and Zhou 2014 for EachNet), or is very time-consuming for people to learn about such as on TMall.

*Department of Economics, Northwestern University. Email: harrydp@northwestern.edu. I thank Dilip Abreu, Jie Bai, Dan Barron, James Best, Joyee Deb, Laura Doval, Jeff Ely, Drew Fudenberg, George Georgiadis, Emir Kamenica, Yingkai Li, Qingmin Liu, Daniel Luo, Lucas Maestri, Meg Meyer, Wojciech Olszewski, Alessandro Pavan, Larry Samuelson, Ali Shourideh, Vasiliki Skreta, Andrzej Skrzypacz, Alex Smolin, Takuo Sugaya, Caroline Thomas, Juuso Välimäki, Allen Vong, Alex Wolitzky, and three anonymous referees for helpful comments. I thank the NSF Grant SES-1947021 and the Cowles Foundation for financial support.

This paper takes a first step to analyze reputation effects when the short-run players do not have detailed information about the patient player's history. I study reputation models where each short-run player only observes some summary statistics of the patient player's last K actions but *cannot* observe the exact sequence of these actions. This stands in contrast to the canonical reputation model of Fudenberg and Levine (1989) as well as existing reputation models with limited memories such as Liu and Skrzypacz (2014) and Pei (2023), all of which assume that the short-run players can observe the *exact sequence* of the patient player's actions.

When the short-run players have short memories, it is unclear whether the patient player can still benefit from building reputations. This is because the short-run players may not be convinced that the patient player will sustain his reputation in the future if they can only observe his behavior in the last few periods.

Nevertheless, I show that in a natural class of games, the patient player can secure high returns from building reputations *regardless* of his opponents' memory length K. This stands in contrast to existing reputation results that *require* the short-run players to have long enough memories. I also show that the short-run players can obtain their first-best payoff in all equilibria *if and only if* their memories are short enough. This stands in contrast to the canonical reputation model of Fudenberg and Levine (1989) in which the short-run players can observe the exact sequence of actions and there exist equilibria where they receive low payoffs.

I study an infinitely repeated game between a patient player and a sequence of short-run players. I assume that players' stage-game payoffs are *monotone-supermodular* in the sense that there exists a complete order on each player's action set such that (i) the patient player's payoff decreases in his own action and increases in his opponent's action, and (ii) both players' payoff functions have strictly increasing differences.

My assumption fits, for example, when the patient player's action is a *relationship-specific investment*. To fix ideas, consider the market for *customized goods*, which takes place on platforms such as eBay, EachNet, and TMall. Each period, a buyer places an order (e.g., 100 units of some customized products). The seller decides whether to *customize* his products according to the buyer's detailed demands (action H) or to supply *standardized* products (action L),[1] and the buyer decides whether to renege on her earlier promise by *returning part of the order* (action N) or to *keep the entire order* (action T). I model this as the product choice game:[2]

[1]Whether the seller supplies products that are customized to buyers' demands has significant effects on buyers' reviews in online markets for customized goods. For example, one buyer wrote in their product review for customized shirts that "*The colors printed were too dull and did not match our vision nor did it match what was ordered.*" and another buyer wrote "*The printing was inconsistent, colors were not as they were online, printing was not fitted correctly on each shirt. Would not purchase again.*" Similarly, on platforms that sell customized furniture, one buyer wrote after posting a one-star review that "*I ordered the color beige, and when the box arrived, I double check to make sure that the color was beige, but when I open the box, the couch is more of a brown color.*"

[2]My baseline model focuses on the case in which players move simultaneously, or equivalently, the buyer receives *no information* about the seller's *current-period action* before choosing her action. In Online Appendix I.1, I extend my results to the *sequential-move* case where each buyer observes a *noisy private signal* about the seller's *current-period action* before choosing her own action. The motivation is that major Ecommerce platforms such as EachNet, TMall, and eBay only allow buyers to return the products and obtain a partial refund within a short timeframe (e.g., in a few days). Although a buyer may eventually learn about the seller's action against her, she may only receive a *noisy signal* about the seller's action against her before the return and refund deadline.

-	T	N
H	$1-c, 1$	$-c, x$
L	$1, -x$	$v, 0$

with $v \in (0, 1-c)$, $c > 0$, and $x \in (0, 1)$.

Intuitively, producing customized products costs the seller more since he needs to finetune the production process, where c is the cost difference. The buyer has less incentive to return products that are customized to her needs. If the buyer returns part of the order, then the seller needs to give her a partial refund and then sell the returned products on the spot market. Products that are customized to a particular buyer are valued less by the market relative to the standardized ones. Hence, the seller recoups less value from selling customized products on the spot market, with the difference denoted by $v > 0$. Hence, the seller's *net cost* of choosing H is c when the buyer chooses T, which is strictly less than his net cost $c + v$ when the buyer chooses N.

The patient player privately observes his *type*: He is either a *commitment type* who chooses his highest action (in the example, action H) in every period, or a *strategic type* who maximizes his discounted average payoff. The patient player's *reputation* is the probability his opponents assign to the commitment type.

My baseline model assumes that each short-run player can only observe the number of times that the patient player took each of his actions in the last $K \in \mathbb{N}$ periods but not the exact sequence of these K actions. In order to be consistent with the existing literature on reputation games with limited memories, such as Liu (2011), Liu and Skrzypacz (2014), and Levine (2021), I make a standard assumption that the short-run players *cannot* directly observe calendar time, or equivalently, how long the game has lasted. They have a prior belief about calendar time and update their beliefs via Bayes rule after observing their histories.

Theorem 1 shows that as long as K is at least 1, the patient player receives at least his commitment payoff (in the example, $1 - c$) in every Nash equilibrium, and that he can secure this payoff by taking his highest action in every period. To the best of my knowledge, Theorem 1 is the first reputation result that allows for *arbitrary memory length*, and in particular, it allows the short-run players to have *arbitrarily short memories*. This aspect of my result stands in contrast to the existing reputation results which assume that the short-run players have infinite memories (e.g., Fudenberg and Levine 1989) or long enough memories (e.g., Theorem 2 in Liu and Skrzypacz 2014). My result contributes to the reputation literature by showing that in a natural class of games, the patient player can secure high returns from building reputations even when his opponents do not have long memories and can only observe some summary statistics of his recent actions.

The challenge to prove Theorem 1 comes from the observation that the short-run players may have arbitrarily short memories and cannot observe everything their predecessors observe. As a result, the standard arguments in Fudenberg and Levine (1989,1992), Sorin (1999), and Gossner (2011) do not apply.

My proof establishes a *no-back-loop property*, that it is never optimal for the patient player to *milk his*

reputation when it is strictly positive and later restore his reputation. Intuitively, since the patient player's payoff increases in the short-run players' action but decreases in his own action, he has an incentive to restore his reputation only if the short-run players take higher actions after he does that. Since the patient player's payoff is supermodular, he has a stronger incentive to take higher actions when the short-run players' actions are higher. Hence, if it is optimal for the patient player to restore his reputation when the short-run players take a lower action, then it is not optimal for him to milk his reputation when the short-run players take a higher action.

The no-back-loop property implies that in every equilibrium, there is *at most one period over the infinite horizon* where the patient player took his highest action in all of the last K periods but will take a lower action in the current period. Since the commitment type takes the highest action in every period, the short-run players believe that the patient player will take the highest action with probability close to one after they observe him taking the highest action in all of the last K periods. This provides the short-run players an incentive to play a best reply against the highest action. Therefore, the patient player obtains approximately his commitment payoff once he deviates to *playing the highest action in every period* and his equilibrium payoff must be greater than his payoff under any deviation.

Next, I examine the short-run players' welfare. This is not covered by Theorem 1, which only shows that the patient player will obtain his commitment payoff if he builds a reputation. However, there might exist other strategies that give the patient player weakly higher payoffs. As a result, it is unclear whether he will build a reputation *in equilibrium* and whether the short-run players can attain a high welfare.

My next set of results establish an equivalence between (i) the short-run players' memory length K being lower than some cutoff, (ii) the patient player taking his highest action in almost all periods in *all* equilibria, and (iii) the short-run players approximately obtaining their first-best welfare in *all* equilibria. I also show that when K is below the cutoff, the patient player will take his highest action with probability close to one except for the initial few periods and periods in the distant future that have negligible payoff consequences.

Take the product choice game example. The intuition is that an increase in K has two effects on the patient player's incentives. First, once he chooses L, more short-run players can observe it when K is larger, in which case more of them *have the ability to punish* him. However, a larger K also makes it *more difficult to motivate* the short-run players to execute the punishment. To see this, suppose that the short-run players believe that the patient player will play L in periods $K-1, 2K-1, ...$ and will play H in other periods. If a short-run player observes that L was played once in the last K periods, then she knows that the patient player is the strategic type. According to Bayes rule, she believes that the patient player will play L in the current period with probability close to $\frac{1}{K}$. When K is large, $\frac{1}{K}$ is small, so a short-run player who does not observe

calendar time believes that it is unlikely that the patient player will play L in the current period, and thus has no incentive to play N. If the short-run players play T when L occurred once in the last K periods, then the patient player prefers *playing L once every K periods* to *playing H in every period*, making the short-run players' beliefs self-fulfilling. This leads to an equilibrium where the patient player chooses L periodically.

When K is small, I show that in *every* equilibrium, the patient player's payoff is bounded below his commitment payoff after he loses his reputation. Since Theorem 1 implies that the patient player can secure his commitment payoff by taking the highest action in every period, he will do so in *all* equilibria.

My proof introduces new techniques that can characterize the common properties of the patient player's behavior in *all* equilibria. Some of my arguments require no assumption on players' payoffs (e.g., Lemma B.1, B.2 and B.3), which are portable to repeated games where players observe random samples of their opponents' past actions and repeated games where players use finite automaton strategies.

This paper contributes to the literature on reputations with limited memories, cooperation under limited information, and the sustainability of reputations. In contrast to the existing reputation models with limited memories such as Liu (2011), Liu and Skrzypacz (2014), and Pei (2023), I study reputation models in which the short-run players *cannot* observe the exact sequence of actions.[3] In contrast to the existing reputation results that require long enough memories, I show that the patient player can secure high payoffs regardless of his opponents' memory. My results also shed light on the effects of memory length on the short-run players' welfare which, to the best of my knowledge, has not been examined in the existing reputation literature.[4]

My paper is also related to the literature on sustaining cooperation when players have limited information about the game's history. This has been studied in repeated games with random matching by Kandori (1992), Ellison (1994), Takahashi (2010), Heller and Mohlin (2018), and Clark, Fudenberg and Wolitzky (2021). Most of these papers focus on the repeated prisoner's dilemma when all players are patient and provide conditions on the monitoring technology under which *either* a folk theorem holds *or* players obtain their minmax payoffs in all equilibria.[5] In contrast, I study repeated games between a patient player and a sequence of short-run players with one-sided lack of commitment (e.g., product choice games) rather than the prisoner's dilemma. My results provide conditions under which players receive high payoffs in *all* equilibria.

Bhaskar and Thomas (2019) study a repeated *complete information* game between a patient player and

[3]Levine (2021) assumes that the short-run players have 1-period memory, that is $K = 1$, in which case whether they can observe the exact sequence of the last K actions is irrelevant. In Jehiel and Samuelson (2012), the short-run players mistakenly believe that the strategic-type of the patient player uses a stationary strategy. In contrast, the short-run players in my model understand that the patient player's behavior when he has a positive reputation can be different from his behavior after he has lost his reputation.

[4]Kaya and Roy (2022) study a repeated signaling game where the receivers' payoffs depend on the sender's type. They show that longer memories *encourage* the low-quality sender to imitate the high-quality one. In contrast, the short-run players' payoff depends only on players' actions in my model and longer memories *undermine* the patient player's incentive to imitate the commitment type.

[5]For games with general payoffs, see Deb (2020), Deb, Sugaya and Wolitzky (2020), and Sugaya and Wolitzky (2020).

a sequence of short-run players. They assume that the short-run players do not have any information about actions that were taken more than K periods ago. They construct information structures under which players can cooperate in some equilibria. By contrast, I study a repeated *incomplete information* game and provide conditions under which the short-run players approximately attain their first-best payoff in *all* equilibria.

Ekmekci (2011) and Vong (2022) focus on games in which there is no complementarity in players' actions and construct rating systems under which there *exists* an equilibrium where the patient player exerts effort in almost all periods. Although I do not explicitly study an information design problem, my results imply that when players' actions are complements, the short-run players can attain their highest feasible payoff in *all* equilibria when they can only observe the summary statistics of the patient player's recent actions.

My results also contribute to the literature on reputation sustainability. Theorem 2 focuses on a novel notion of reputation sustainability, that whether the patient player will take his commitment action with *discounted frequency* close to 1 in all equilibria. Compared to the criteria in Cripps, Mailath and Samuelson (2004) which focuses on the patient player's behavior and reputation as $t \to +\infty$, my notion of sustainability is better suited for evaluating the short-run players' welfare and social welfare.[6] Pei (2020) and Ekmekci and Maestri (2022) study *interdependent value* models and provide conditions under which the patient player takes his commitment action in almost all periods in all equilibria. In contrast, I study a *private value* model in which the short-run players cannot observe the exact sequence of actions, and therefore, cannot fine-tune their punishments based on the details of the game's history. The patient player has stronger incentives to sustain his reputation since he will be punished at a larger set of histories after he loses his reputation.

2 Baseline Model

Time is indexed by $t = 0, 1....$ A long-lived player 1 interacts with a different player 2 in each period. After each period, the game ends with probability $1 - \delta$ with $\delta \in (0, 1)$, after which players' stage-game payoffs are normalized to 0. Player 1 is indifferent between receiving one unit of utility in the current period and in the next period, so he discounts his future payoffs by δ. In period t, player 1 chooses $a_t \in A$ and player 2_t chooses $b_t \in B$ simultaneously from finite sets A and B. My baseline model focuses on games where player 2's action choice is binary, that is, $|B| = 2$.[7] This is a primary focus of the reputation literature, including Mailath and Samuelson (2001), Ekmekci (2011), Liu (2011), and Levine (2021). Players' stage-game payoffs

[6] Ekmekci, Gossner and Wilson (2012) and Liu and Skrzypacz (2014) propose an alternative notion of reputation sustainability, that the patient player can secure his commitment payoff *at every history* in every equilibrium, rather than just securing his commitment payoff in period 0. Theorem 1 implies that in my model, reputation is sustainable under their notion as long as $K \geq 1$.

[7] I discuss extensions of my theorems to games where $|B| \geq 3$ in Section 4, which include but are not limited to the games studied by Liu and Skrzypacz (2014) where player 2 has a unique best reply to every $\alpha \in \Delta(A)$.

are $u_1(a_t, b_t)$ and $u_2(a_t, b_t)$, which satisfy the following *monotone-supermodularity* assumption:

Assumption 1. *There exist a complete order $\succ_A$ on A and a complete order $\succ_B$ on B such that first, $u_1(a, b)$ is strictly increasing in b and is strictly decreasing in a, and second, both $u_1(a, b)$ and $u_2(a, b)$ have strictly increasing differences in a and b.*

The product choice game satisfies Assumption 1 once players' actions are ranked according to $H \succ_A L$ and $T \succ_B N$. The requirements that (i) $u_1(a, b)$ strictly increases in b and strictly decreases in a and (ii) $u_2(a, b)$ has strictly increasing differences, are standard in the literature and are also assumed in Ekmekci (2011), Liu (2011), and Liu and Skrzypacz (2014). The assumption that $u_1(a, b)$ has strictly increasing differences stands in contrast to those papers, which assume that $u_1(a, b)$ has weakly decreasing differences.

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 ω_c stand for a *commitment type* who plays his highest action $a^* \equiv \max A$, or his *commitment action*, in every period. Let ω_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. Let π_t be the probability that player 2_t's belief assigns to the commitment type, which I call player 1's *reputation*.

Before choosing b_t, player 2_t only observes the number of times that player 1 took each of his actions in the last $\min\{t, K\}$ periods, where $K \in \{1, 2, ...\}$ is a parameter that measures the society's memory length. An implication is that player 2 does not know the order with which player 1 took his last K actions. For example, if $K = 2$, then player 2 cannot distinguish between $(a_1, a_2) = (a^*, a')$ and $(a_1, a_2) = (a', a^*)$. This is the modeling innovation relative to Liu (2011), Liu and Skrzypacz (2014), and Pei (2023), which assume that the short-run players can *perfectly* observe the *exact sequence* of the patient player's last K actions.

I also assume that player 2 *cannot* directly observe calendar time.[8] As in Liu and Skrzypacz (2014), player 2 has a full support prior belief about calendar time and update their beliefs using Bayes rule after observing their histories. The standard interpretation is that due to limited memories, it is hard for the short-run players to know exactly *how long the game has lasted*, especially when the game has lasted for much longer relative to their memories. Nevertheless, the short-run players who arrive before period K can perfectly *infer* calendar time based on the history they observe. This is consistent with my formulation, since the first K players observe fewer than K actions, so their *posterior beliefs* will assign probability 1 to the true calendar time.

What is a reasonable prior belief about calendar time? Since the game ends with probability $1 - \delta$ after each period, for every $t \in \{0, 1, ...\}$, the probability that player 2's prior assigns to calendar time being $t + 1$

[8]This is also assumed in Liu and Skrzypacz (2014), Section 2 in Acemoglu and Wolitzky (2014), Cripps and Thomas (2019), Levine (2021), among many others. Liu (2011) and Heller and Mohlin (2018) focus on *stationary equilibria* where strategies are *required* to be time-independent, which is equivalent to a model where the short-run players have an improper uniform prior.

should equal δ times the probability her prior assigns to calendar time being t. The unique prior belief that satisfies this restriction for every t is the one that assigns probability $(1-\delta)\delta^t$ to calendar time being t. Hu (2020) provides a foundation for this exponential prior by constructing a distribution over entry processes.

The set of player 1's histories is $\mathcal{H}_1 \equiv \{(a_s, b_s)_{s=0}^{t-1} \text{ s.t. } t \in \mathbb{N} \text{ and } (a_s, b_s) \in A \times B\}$ with a typical element denoted by h^t. The set of player 2's histories is $\mathcal{H}_2 \equiv \{(n_1, ..., n_{|A|}) \in \mathbb{N}^{|A|} \text{ s.t. } n_1 \geq 0, ..., n_{|A|} \geq 0 \text{ and } n_1 + ... + n_{|A|} \leq K\}$, where $n_1, ..., n_{|A|}$ denote the number of times that player 1 played each of his actions in the last K periods. The strategic type of player 1's strategy is $\sigma_1 : \mathcal{H}_1 \to \Delta(A)$. Player 2's strategy is $\sigma_2 : \mathcal{H}_2 \to \Delta(B)$. Let Σ_i be the set of player i's strategies, where $i \in \{1, 2\}$. Under my formulation, player 2's action depends only on the history she observes. For example, player 2_t and player 2_{t+1} will take the same (possibly mixed) action if they observe the same history. This is a standard requirement in reputation models with limited memories, see for example Liu (2011), Liu and Skrzypacz (2014), and Levine (2021).

3 Main Results

Section 3.1 shows that the patient player will receive at least his commitment payoff in all equilibria regardless of his opponent's memory length. My proof uses a *no-back-loop property* that applies to all of the patient player's best replies in the repeated game. Section 3.2 shows that the short-run players can approximately attain their highest feasible payoff in all equilibria and that the patient player will play a^* in almost all periods in all equilibria *if and only if* the short-run players' memory length is lower than some cutoff.

3.1 Reputation Result for Arbitrary Memory Length

Let b^* denote player 2's lowest best reply to a^*. Let $u_1(a^*, b^*)$ be player 1's *commitment payoff*. Let $\underline{a} \equiv \min A$ denote player 1's lowest action. Let $\underline{b}$ denote player 2's lowest best reply to $\underline{a}$. For every $\pi_0 \in (0, 1)$, there exists $\underline{\delta}(\pi_0) \in (0, 1)$ such that for every $\delta > \underline{\delta}(\pi_0)$, each of player 2's best reply to mixed action

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

is no less than b^*. Such $\underline{\delta}(\pi_0) \in (0, 1)$ exists for every $\pi_0 \in (0, 1)$ since b^* is the lowest best reply to a^*, the value of $\frac{(1-\delta)(1-\pi_0)}{\pi_0}$ converges to 0 as $\delta \to 1$, and best reply correspondences are upper-hemi-continuous.

Theorem 1. *For any $\pi_0 \in (0, 1)$, $\delta > \underline{\delta}(\pi_0)$, and $K \geq 1$, player 1's payoff in any Nash equilibrium is at least*

$$(1 - \delta^K)u_1(a^*, \underline{b}) + \delta^K u_1(a^*, b^*). \tag{3.1}$$

The payoff lower bound converges to $u_1(a^*, b^*)$ as $\delta \to 1$. Therefore, Theorem 1 identifies a class of games such that regardless of the short-run players' memory length K,[9] a patient player can secure at least his commitment payoff $u_1(a^*, b^*)$ by building a reputation for playing a^*. As will become clear later in the proof, at every history of every equilibrium, if the patient player deviates by playing a^* in every subsequent period, then his continuation value after K periods is at least $u_1(a^*, b^*)$. My result stands in contrast to the repeated complete information game *without* any commitment type, in which there are equilibria where players play $(\underline{a}, \underline{b})$ in every period and the patient player receives his minmax payoff $u_1(\underline{a}, \underline{b})$.

To the best of my knowledge, Theorem 1 is the first reputation result in the literature that allows the short-run players to have arbitrary memory length, and in particular, they may have *arbitrarily short memories*. This aspect of my result stands in contrast to the existing reputation results which require the short-run players to have *infinite memories* (e.g., Fudenberg and Levine 1989), or *long enough memories* (e.g., Theorem 2 in Liu and Skrzypacz 2014),[10] or *infinite memories* about some noisy signal that can statistically identify the patient player's action (e.g., Fudenberg and Levine 1992, Gossner 2011, Theorem 2 in Pei 2023).

Since the short-run players have limited memories and cannot observe everything their predecessors observe, the standard techniques to show reputation results such as the ones in Fudenberg and Levine (1989, 1992), Sorin (1999), and Gossner (2011) do not apply. To overcome these challenges, my proof uses an observation called the *no-back-loop property*, that it is *never* optimal for the patient player to milk his reputation when it is strictly positive and later restore his reputation.[11] I state this observation as a lemma, provide a heuristic explanation, and use this result to show Theorem 1 by the end of this section.

Let $\mathcal{H}_1^* \equiv \left\{(a_s, b_s)_{s=0}^{t-1} \text{ such that } t \geq K \text{ and } (a_{t-K}, ..., a_{t-1}) = (a^*, ..., a^*)\right\}$ be the set of player 1's histories where calendar time is at least K and his last K actions were a^*. Let $\mathcal{H}_1(\sigma_1, \sigma_2)$ be the set of histories that occur with positive probability under (σ_1, σ_2). Let $U_1(\sigma_1, \sigma_2)$ be player 1's discounted average payoff from (σ_1, σ_2). I say that strategy $\widehat{\sigma}_1$ best replies to σ_2 if $\widehat{\sigma}_1 \in \arg\max_{\sigma_1 \in \Sigma_1} U_1(\sigma_1, \sigma_2)$.

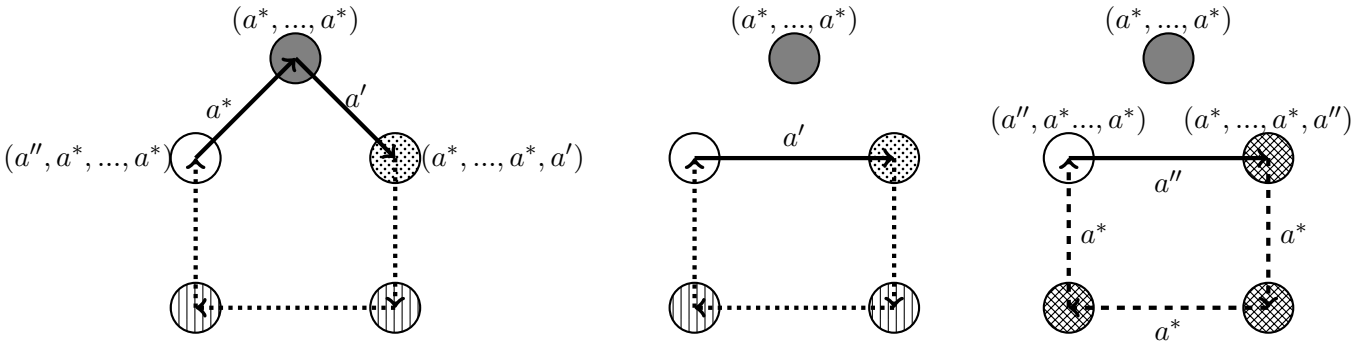
No-Back-Loop Lemma. *For every $\sigma_2 : \mathcal{H}_2 \to \Delta(B)$ and pure strategy $\widehat{\sigma}_1 : \mathcal{H}_1 \to A$ that best replies to σ_2, there does not exist any $h^t \in \mathcal{H}_1(\widehat{\sigma}_1, \sigma_2) \bigcap \mathcal{H}_1^*$ such that when player 1 uses strategy $\widehat{\sigma}_1$, he plays an action that is not a^* at h^t and reaches another history that belongs to $\mathcal{H}_1(\widehat{\sigma}_1, \sigma_2) \bigcap \mathcal{H}_1^*$ in the future.*

My *no-back-loop lemma* rules out situations depicted in the left panel of Figure 1, that is, player 1's best

[9] When $K = +\infty$, one can use Fudenberg and Levine (1989)'s argument to show that player 1 can secure payoff approximately $u_1(a^*, b^*)$ in every equilibrium as $\delta \to 1$. My theorem focuses on the novel case in which K is finite.

[10] Theorem 2 in Liu and Skrzypacz (2014) shows that for every $\pi_0 > 0$, there exists $\widehat{K} \in \mathbb{N}$ such that the patient player can approximately secure his commitment payoff in every equilibrium when $K > \widehat{K}$. That is, K needs to be large enough.

[11] My no-back-loop property is reminiscent of the bad news model in Board and Meyer-ter-Vehn (2013), in which the long-run player has a stronger incentive to exert effort when he has a higher reputation. This implies that the high-reputation state is absorbing in their model. Although there are similarities, my no-back-loop property does *not* imply that the high-reputation state is absorbing. In fact, there exist equilibria in which player 1 has a strict incentive to play actions other than a^* when his last K actions were a^*.



Player 1 uses strategy $\widehat{\sigma}_1$ that violates the no-back-loop property

Player 1 uses Deviation A

Player 1 uses Deviation B

Figure 1: The gray circle represents a history that belongs to $\mathcal{H}_1^*$. The white circle represents a history where player 1 is one-period-away from $\mathcal{H}_1^*$. The dotted circle represents a history that is reached after player 1 plays a' at the gray circle. The lined circles represent histories that are reached when player 1 plays $\widehat{\sigma}_1$. The cross-hatched circles represent histories where a'' occurred once and a^* occurred $K-1$ times.

reply $\widehat{\sigma}_1$ asks him to play $a'(\neq a^*)$ when his last K actions were a^* (the gray circle) and after a finite number of periods, returns to a history where his last K actions were a^*. This property applies to *all* of player 1's best replies in the repeated game under any discount factor, not just to player 1's equilibrium strategies.

Due to two modeling differences, the *reputation cycles* that my lemma ruled out occur in all equilibria in the reputation model of Liu and Skrzypacz (2014). First, they assume that u_1 has *strictly decreasing differences* while I assume that u_1 has strictly increasing differences. Second, they assume that player 2 can *perfectly* observe the exact sequence of player 1's last K actions while I assume that player 2 cannot observe the exact sequence of actions. In Online Appendices G and H, I discuss two alternative models which are only one-step-away from both my baseline model and the model of Liu and Skrzypacz (2014).

My lemma does *not* follow from existing results on supermodular games, most of which focus on static games. The intuition is that even when player 1's stage-game payoff $u_1(a, b)$ has strictly increasing differences, it is not necessarily the case that when the game is played *repeatedly*, player 1 has a stronger incentive to play higher actions at histories where player 2's actions are higher. This is because player 1's current-period action affects future player 2's observations, which in turn affects future player 2's actions as well as player 1's continuation value. I explain the ideas behind my proof. The detailed calculations are in Appendix A.

Proof Sketch: Since player 2 has no information about player 1's action more than K periods ago, it is without loss to focus on player 1's pure-strategy best replies that depend only on his last K actions, including the order of these K actions. Suppose by way of contradiction that there exists a pure strategy $\widehat{\sigma}_1$ that best replies to σ_2

such that $\widehat{\sigma}_1$ plays $a' \ (\neq a^*)$ at a history h^t where $(a_{t-K}, ..., a_{t-1}) = (a^*, ..., a^*)$, and after a finite number of periods, reaches a history h^s that satisfies $(a_{s-K}, ..., a_{s-1}) = (a'', a^*, ..., a^*)$ where $a'' \neq a^*$, and then plays a^* at h^s after which all of the last K actions are a^* again. Note that a' and a'' can be the same action.

I depict strategy $\widehat{\sigma}_1$ in the left panel of Figure 1, where h^t is represented by the gray circle and h^s is represented by the white circle. I propose *two deviations* for player 1 starting from the white circle:

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

I depict these deviations in Figure 1. Next, I compare player 1's continuation value at the white circle when he uses $\widehat{\sigma}_1$ to those under the two deviations. I argue that at least one of these deviations is strictly profitable.

1. Compared to $\widehat{\sigma}_1$, Deviation A takes a lower-cost action a' at the white circle, skips the gray circle, and frontloads the payoffs along the dotted lines (i.e., the dotted, lined, and white circles). If player 1 prefers $\widehat{\sigma}_1$ to Deviation A, then his average payoff from the circles along the dotted lines (i.e., the payoff that Deviation A frontloads) must be strictly lower than his stage-game payoff at the gray circle.

2. Compared to $\widehat{\sigma}_1$, Deviation B takes a lower-cost action a'' at the white circle, skips the gray circle, and induces payoffs along the dashed lines in the next $K-1$ periods (i.e., the cross-hatched circles). If player 1 prefers $\widehat{\sigma}_1$ to Deviation B, then his average payoff along the dashed lines must be strictly less than a convex combination of his payoff at the gray circle and his average payoff along the dashed lines.

If $\widehat{\sigma}_1$ is player 1's best reply, then both Deviation A and Deviation B are unprofitable. Therefore, player 1's stage-game payoff at the gray circle must be strictly greater than his average payoff along the dashed lines.

If $\widehat{\sigma}_1$ best replies to σ_2, then player 1 prefers a' to a^* at the gray circle and prefers a^* to a' at the white circle. No matter whether player 1 is currently at the gray circle or at the white circle, he will reach the gray circle after playing a^* and will reach the dotted circle after playing a'. Hence, the difference in player 1's incentives at the gray circle and at the white circle *cannot* be driven by his continuation value. This implies that such a difference in incentives can only be driven by player 1's stage-game payoff, which is affected by player 2's actions at the gray circle and at the white circle. Since $u_1(a, b)$ has strictly increasing differences and $a^* \succ_A a'$, the Topkis Theorem implies that it *cannot* be the case that player 2's mixed action at the gray circle strictly FOSDs her mixed action at the white circle. Since $|B| = 2$, player 2's action at the white circle weakly FOSDs her action at the gray circle. Since player 2 cannot observe the order of player 1's last K actions, player 2's action at every circle along the dashed line coincides with her action at the white circle.

This leads to a contradiction since on the one hand, player 1's stage-game payoff at the gray circle is strictly greater than his average payoff along the dashed lines, and on the other hand, player 2's action at the white circle weakly FOSDs her action at the gray circle and player 1's stage-game payoff is strictly increasing in player 2's action. This contradiction implies that at the white circle, either Deviation A or Deviation B yields a strictly higher payoff for player 1 compared to strategy $\widehat{\sigma}_1$. Therefore, $\widehat{\sigma}_1$ cannot be a best reply. □

In the remainder of this section, I use the no-back-loop lemma to show Theorem 1. Let Σ_1^* denote the set of player 1's pure strategies that satisfy the no-back-loop property. For any Nash equilibrium $(\widetilde{\sigma}_1, \sigma_2)$, the no-back-loop lemma implies that every pure strategy in the support of $\widetilde{\sigma}_1$ belongs to Σ_1^*.

Proof of Theorem 1: For every $t \in \mathbb{N}$, let E_t denote the event that *player 1 is the strategic type* and *no action other than* a^* *was played from period* $\max\{0, t-K\}$ *to period* $t-1$. Fix any $\sigma_1 \in \Sigma_1^*$ and σ_2, let $p_t(\sigma_1, \sigma_2)$ denote the *ex ante* probability of event E_t when the strategic type of player 1 plays σ_1 and player 2 plays σ_2. Since player 1 is the commitment type with probability π_0, we know that $p_t(\sigma_1, \sigma_2) \leq 1 - \pi_0$ for every $t \in \mathbb{N}$. Let $\mathbb{N}^*(\sigma_1, \sigma_2) \subset \mathbb{N}$ denote the set of calendar times 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 $q_t(\sigma_1, \sigma_2)$ denote the probability that player 1 *does not* play a^* in period t conditional on event E_t. The definition of Σ_1^* 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 any arbitrary Nash equilibrium $(\widetilde{\sigma}_1, \sigma_2)$. For every $\sigma_1 \in \Sigma_1^*$, let $\widetilde{\sigma}_1(\sigma_1)$ be the probability that $\widetilde{\sigma}_1$ assigns to σ_1, which is well-defined since Σ_1^* is a countable set. Recall that player 2's prior belief assigns probability $(1-\delta)\delta^t$ to the calendar time being t. According to Bayes rule, at any history after period K where all of player 1's last K actions were a^*, player 2 believes that player 1's action is *not* a^* with probability

$$\frac{\sum_{\sigma_1 \in \Sigma_1^*} \widetilde{\sigma}_1(\sigma_1) \sum_{t \in \mathbb{N}^*(\sigma_1, \sigma_2)} (1-\delta)\delta^t p_t(\sigma_1, \sigma_2) q_t(\sigma_1, \sigma_2)}{\pi_0 \sum_{t=K}^{+\infty} (1-\delta)\delta^t + \sum_{\sigma_1 \in \Sigma_1^*} \widetilde{\sigma}_1(\sigma_1) \sum_{t \in \mathbb{N}^*(\sigma_1, \sigma_2)} (1-\delta)\delta^t p_t(\sigma_1, \sigma_2)}. \tag{3.2}$$

The denominator of (3.2) 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 $\sigma_1 \in \Sigma_1^*$, Σ_1^* is a countable set, and $t \geq K$ for every $t \in \mathbb{N}^*(\sigma_1, \sigma_2)$, the numerator of (3.2) is no more than $(1-\delta)(1-\pi_0)\delta^K$. This suggests that (3.2) 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 when player 1's last K actions were a^*. Moreover, since $\underline{b}$ is player 2's lowest best reply to player 1's lowest action $\underline{a}$, actions strictly lower than $\underline{b}$ are never optimal for player 2. Hence, in any Nash equilibrium, if player 1 plays a^* in every period, his discounted average payoff is at least $(1-\delta^K) u_1(a^*, \underline{b}) + \delta^K u_1(a^*, b^*)$. This is a lower bound for player 1's equilibrium payoff. □

3.2 Equilibrium Behavior & The Short-Run Players' Welfare

Although Theorem 1 shows that player 1 can secure his commitment payoff $u_1(a^*, b^*)$ *if* he plays a^* in every period, it does not imply that he will actually play a^* in equilibrium since he may have other strategies that lead to weakly higher payoffs. In addition, Theorem 1 also remains silent about the short-run players' welfare.

The results in this section examine the patient player's behavior and the short-run players' welfare. Since the game ends with probability $1-\delta$ after each period, the sum of the short-run players' payoffs is $\mathbb{E}^\sigma\Big[\sum_{t=0}^{+\infty}\delta^t u_2(a_t,b_t)\Big]$, where $\mathbb{E}^\sigma[\cdot]$ denotes the expectation induced by $\sigma \equiv (\sigma_1,\sigma_2)$. I use the following *normalized sum* of the short-run players' payoffs, denoted by U_2^σ, to measure their welfare:

$$U_2^\sigma \equiv \mathbb{E}^\sigma\Big[\sum_{t=0}^{+\infty}(1-\delta)\delta^t u_2(a_t,b_t)\Big]. \tag{3.3}$$

By definition, $U_2^\sigma = \sum_{(a,b)\in A\times B} F^\sigma(a,b)u_2(a,b)$ where

$$F^\sigma(a,b) \equiv \mathbb{E}^\sigma\Big[\sum_{t=0}^{+\infty}(1-\delta)\delta^t \mathbf{1}\{a_t=a, b_t=b\}\Big] \tag{3.4}$$

is called the *discounted frequency* (or the *occupation measure*) of action profile (a,b). Hence, the short-run players' welfare depends on σ only through the discounted action frequencies $\{F^\sigma(a,b)\}_{(a,b)\in A\times B}$.

Theorem 2. *There exists a cutoff $\overline{K}\in\mathbb{N}$ that depends only on (u_1,u_2) such that:*

1. *There exists a constant $C\in\mathbb{R}_+$ that is independent of δ such that for every $1\le K<\overline{K}$, we have $\sum_{b\succeq b^*}F^\sigma(a^*,b)\ge 1-(1-\delta)C$ in every Nash equilibrium σ under K and δ.*
2. *There exists $\eta>0$ such that for every $K\ge\overline{K}$, there exists $\underline{\delta}\in(0,1)$ such that for every $\delta>\underline{\delta}$, there exists a PBE with strategy profile σ such that $\sum_{b\in B}F^\sigma(a^*,b)\le 1-\eta$.*

The proof is in Appendix B, with an intuitive explanation provided by the end of this section. Theorem 2 implies that (i) player 1 plays a^* and player 2's action is at least b^* in almost all periods in all Nash equilibria when player 2's memory K is lower than some cutoff $\overline{K}$, and (ii) there exists at least one PBE in which player 1 plays a^* with frequency bounded below 1 when K is above the cutoff. The presence of the commitment type is necessary for the first part of this result, since in a repeated complete information game *without* any commitment type, there exist equilibria where players play $(\underline{a},\underline{b})$ in every period regardless of δ and K.

Although Theorem 2 focuses on player 1's discounted action frequencies, it leads to a sharp prediction on the *dynamics* of player 1's behavior and reputation when K is small. My next result shows that when K

is below the cutoff $\overline{K}$, the strategic type of player 1 has a strictly positive reputation with probability close to 1 after the initial few periods, after which he will play a^* with probability close to 1 in every period until calendar time t is so large that δ^t is close to 0. I state this as Corollary 1 with proof in Online Appendix C.

Corollary 1. *Suppose $K < \overline{K}$. For every $\varepsilon > 0$, there exist a constant $C_\varepsilon \in \mathbb{R}_+$ and $\underline{\delta} \in (0,1)$ such that for every $\delta > \underline{\delta}$, every equilibrium under δ, and every $t \in \mathbb{N}$ that satisfies $\delta^t \in (\varepsilon, 1-\varepsilon)$, the probability that $h^t \in \mathcal{H}_1^*$ is at least $1-(1-\delta)C_\varepsilon$ and player 1 plays a^* with probability at least $1-(1-\delta)C_\varepsilon$ in period t.*

Theorem 2 also implies that when K is below the cutoff $\overline{K}$, the short-run players' welfare is arbitrarily close to $u_2(a^*, b^*)$ in all Nash equilibria. When $u_2(a,b)$ is strictly increasing in a, $u_2(a^*, b^*)$ is the short-run players' *highest feasible payoff*. This is because $u_2(a^*, b) > u_2(a, b)$ for every $a \neq a^*$ and $b \in B$, and $u_2(a^*, b^*) \geq u_2(a^*, b)$ for every $b \in B$ given that b^* best replies to a^*. Theorem 3 provides a necessary and sufficient condition under which the short-run players attain their highest feasible payoff in all equilibria.

Theorem 3. *Suppose (u_1, u_2) satisfies Assumption 1 and $u_2(a,b)$ is strictly increasing in a.*

1. *There exists a constant $C_0 \in \mathbb{R}_+$ that is independent of δ such that for every $1 \leq K < \overline{K}$, we have $U_2^\sigma \geq u_2(a^*, b^*) - C_0(1-\delta)$ in every Nash equilibrium σ under K and δ.*

2. *There exists $\xi > 0$ such that for every $K \geq \overline{K}$, there exists $\underline{\delta} \in (0,1)$ such that for every $\delta > \underline{\delta}$, there exists a PBE with strategy profile σ such that $U_2^\sigma < u_2(a^*, b^*) - \xi$.*

The proof of Theorem 3 follows from that of Theorem 2, which I omit in order to avoid repetition. Theorem 3 implies that the short-run players obtain their highest feasible payoff in all equilibria when K is below the cutoff $\overline{K}$ but their payoffs are bounded below first best in some equilibria when K is above the cutoff. Therefore, longer memories, modeled as a larger K, may *lower* the short-run players' welfare.

Two natural questions follow from Theorems 2 and 3. First, how to compute the cutoff $\overline{K}$ from the primitives u_1 and u_2? Second, how large can η be? My proof of Theorems 2 and 3 sheds light on these questions as well. To preview the answers, I say that a^* is player 1's *optimal pure commitment action* if

$$u_1(a^*, b^*) > \max_{a \neq a^*} \max_{b \in \mathrm{BR}_2(a)} u_1(a, b). \tag{3.5}$$

If (u_1, u_2) violates (3.5), then $\overline{K} = 1$ and η can be as large as 1 for every $K \geq 1$. That is, the frequency with which player 1 plays a^* is 0 in some PBE no matter how small K is.

The interesting case is the one in which (u_1, u_2) satisfies (3.5), that is, a^* is player 1's optimal pure commitment action. The cutoff $\overline{K}$ is the smallest integer $\widehat{K} \in \mathbb{N}$ such that b^* best replies to the mixed action

$\frac{\widehat{K}-1}{\widehat{K}}a^* + \frac{1}{\widehat{K}}a'$ for some $a' \neq a^*$. Moreover, η can be as large as $\frac{m}{K}$, where $m \in \{1, 2, ..., K\}$ is such that b^* best replies to the mixed action $\frac{K-m}{K}a^* + \frac{m}{K}a'$ for some $a' \neq a^*$. As $K \to +\infty$, η converges to η^* where η^* is the largest $\widetilde{\eta}$ such that b^* best replies to $(1-\widetilde{\eta})a^* + \widetilde{\eta}a'$ for some $a' \neq a^*$. In the product choice game, players' payoffs satisfy (3.5) since H is player 1's highest action and committing to play H results in a strictly higher payoff for player 1 relative to committing to play L. The above algorithm implies that $\overline{K} = \left\lceil \frac{1}{1-x} \right\rceil$.

However, due to integer constraints, it is not necessarily the case that the short-run players' worst equilibrium payoff decreases in K. Nevertheless, as I explained earlier that when $K \to +\infty$, the lowest frequency with which player 1 plays a^* converges to η^* where η^* is the largest $\widetilde{\eta}$ such that b^* best replies to $(1-\widetilde{\eta})a^*+\widetilde{\eta}a'$ for some $a' \neq a^*$. Hence, there exists a uniform bound $u' < u_2(a^*, b^*)$ such that for every $K \geq \overline{K}$, there exists an equilibrium where the short-run players' welfare is no more than u'. In contrast, when K is below $\overline{K}$, the short-run players' welfare is arbitrarily close to $u_2(a^*, b^*)$ in *all* equilibria.

Mechanism Behind Theorems 2 and 3: I use the product choice game to explain the mechanism behind Theorems 2 and 3. Intuitively, a longer memory has two effects on the patient player's reputational incentives. An obvious effect is that when K is larger, each of the patient player's actions is observed by more short-run players, so that he can be punished by more people after he plays L. This encourages him to play H.

However, there is another countervailing effect, which is that a larger K makes it more difficult to motivate the short-run players to execute the punishment. To the best of my knowledge, this effect has not been discussed in the existing reputation literature. I explain this novel effect using a thought experiment.

Suppose that the short-run players believe that the strategic type of player 1 will play L in periods $K-1, 2K-1, 3K-1, ...$ and will play H in other periods. After they observe that L was played once in the last K periods, their posterior belief assigns probability close to $\frac{1}{K}$ to L being played in the current period. Hence, the short-run players have an incentive to play T at such histories only when $x \leq \frac{K-1}{K}$, or equivalently when $K \geq \frac{1}{1-x}$. If this is the case, then player 1 prefers *playing L once every K periods* to *playing H in every period*, making the short-run players' beliefs self-fulfilling. If this is not the case (i.e., the short-run players prefer to play N after observing one L in the last K periods), then playing L once every K periods gives player 1 a strictly lower payoff compared to playing H in every period, so that in equilibrium, the short-run players cannot entertain the belief that the strategic type of player 1 will play L once every K periods.

Using similar ideas, one can also show that for every $m \in \{1, 2, ..., K-1\}$ and $a' \neq a^*$ such that b^* best replies to $\frac{K-m}{K}a^* + \frac{m}{K}a'$, there exists an equilibrium 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 at most m times and a^* being played at least $K-m$ times in the last K periods.

Comparison with Fudenberg and Levine (1989): When $K < \overline{K}$, my results lead to sharp predictions not only on the patient player's equilibrium payoff, but also on his equilibrium behaviors as well as on the short-run players' welfare. This aspect of my results stands in contrast to most of the existing results in the reputation literature, including those in Fudenberg and Levine (1989), that focus exclusively on the patient player's payoff but do *not* lead to sharp predictions on the short-run players' welfare and players' behaviors.

For example, in Fudenberg and Levine (1989)'s reputation model where the full history is publicly observed, Li and Pei (2021) characterize the action frequencies that can arise in equilibrium and find that a wide range of behaviors are plausible. In the product choice game of the introduction, the discounted frequency with which the patient player plays H can be anything between $\frac{x(1-c-v)}{1-xc-v}$ and 1 and the short-run players' equilibrium payoff can be anything between their highest feasible payoff 1 and their minmax value 0.

According to Theorem 2, when the short-run players have short memories and cannot observe the exact sequence of player 1's actions, the only equilibria that survive are those where player 1 plays a^* in almost all periods. It implies that in Fudenberg and Levine (1989)'s model, equilibria where the short-run players obtain a low payoff rely on strategies that depend either on events that happened in the distant past or on the fine details of the game's history. In Online Appendix D, I compare the predictions on player 1's *discounted action frequency* in Fudenberg and Levine (1989)'s model to those in my model under any arbitrary $K \in \mathbb{N}$.

Proof Sketch: The idea behind the proof for the second part is contained in the thought experiment, with details in Appendix B.2. The main challenge is in the proof for the first part. This is because characterizing *all* equilibria in a repeated game is not tractable and given that player 2s are Bayesian, their expectation of player 1's current-period action *may not* be close to player 1's action frequency in the last K periods. This suggests that ruling out the type of equilibria in the thought experiment is *insufficient* to show that a^* will be played with frequency close to 1 in *all* equilibria. I explain the ideas behind my proof with details in Appendix B.1.

Starting from period K, players' incentives depend only on player 1's last K actions. Therefore, I treat every sequence of player 1's actions with length K as a *state* with $S \equiv A^K$ the set of states. In order to show that player 1 plays a^* in almost all periods, I only need to show that there exists a constant $C > 0$ such that the discounted frequency of state $s^* \equiv (a^*, ..., a^*)$ is at least $1 - (1-\delta)C$ in every Nash equilibrium.

For any subset of states $S' \subset S$, let $\mathcal{I}(S')$ be the probability that the state moves from $S \backslash S'$ to S', which I call the *inflow* to S', and let $\mathcal{O}(S')$ be the probability that the state moves from S' to $S \backslash S'$, which I call the *outflow* from S'. I show that the difference between the inflow to S' and the outflow from S' must be small:

$$|\mathcal{I}(S') - \mathcal{O}(S')| \leq \frac{1-\delta}{\delta}. \tag{3.6}$$

Fix any equilibrium (σ_1, σ_2), I argue that the inflow to $\{s^*\}$ and the outflow from $\{s^*\}$ must be no more than $\frac{2(1-\delta)}{\delta}$. To see why, let $S' \subset S$ be the set of states such that $s' \in S'$ if and only if it is optimal for player 1 to reach state s^* from s' in finite time. The *no-back-loop lemma* implies that *either* player 1 has no incentive to play actions other than a^* at s^* *or* player 1 never returns to s^* after playing actions other than a^* at s^*, which is the case only if he never reaches states in S' after leaving s^*. In the first case, the outflow from $\{s^*\}$ is 0, so the inflow from $\{s^*\}$ is no more than $\frac{1-\delta}{\delta}$. In the second case, the inflow to S' is 0, so the outflow from S' is no more than $\frac{1-\delta}{\delta}$. The definition of S' implies that the inflow to $\{s^*\}$ is no more than the outflow from S', which in turn implies that both the inflow to $\{s^*\}$ and the outflow from $\{s^*\}$ are no more than $\frac{2(1-\delta)}{\delta}$.

Let S_k be the set of states such that k of the last K actions were *not* a^*. By definition, $S_0 = \{s^*\}$. I show that for every $k \in \{1, 2, ..., K\}$, if the probability that the state moves from S_{k-1} to S_k and from S_k to S_{k-1} are both bounded above by some linear function of $1-\delta$, then the probability that the state belongs to S_k must also be bounded above by some linear function of $1-\delta$. The previous step implies that the probability that the state moves from S_0 to S_1 and from S_1 to S_0 are low. This implies that states in S_1 occur with low probability, which further implies that the probability that the state moves from S_1 to S_2 is low. Inequality (3.6) then implies that the probability that the state moves from S_2 to S_1 is low. Iterate this argument, one can obtain that the probability of states other than s^* is bounded above by a linear function of $1-\delta$.

4 Concluding Remarks

I analyze reputation games where the short-run players have limited memories and cannot observe the exact sequence of actions. I show that the patient player receives at least his commitment payoff in all Nash equilibria regardless of the short-run players' memory length. I also show that the patient player will play his commitment action in almost all periods in all equilibria *if and only if* the short-run players' memory length is lower than some cutoff. I conclude by discussing my modeling assumptions and several extensions.

Games with $|B| \geq 3$: In contrast to games where $|B| = 2$, one can no longer rank any pair of player 2's mixed actions via FOSD when $|B| \geq 3$. For example, when $B = \{b_1, b_2, b_3\}$ with $b_1 \succ_B b_2 \succ_B b_3$, mixed actions $\frac{1}{2}b_1 + \frac{1}{2}b_3$ and b_2 cannot be ranked via FOSD. The proof of the *no-back-loop lemma* sketched in Section 3 breaks down since player 2's action at the gray circle does not strictly FOSD her action at the white circle *cannot imply* that her action at the white circle weakly FOSDs her action at the gray circle.

Let

$$\mathcal{B}^* \equiv \left\{\beta \in \Delta(B) \middle| \text{ there exists } \alpha \in \Delta(A) \text{ s.t. } \beta \text{ best replies to } \alpha\right\} \tag{4.1}$$

be the set of player 2's *mixed-strategy best replies*. The issue I identified does not arise under Assumption 2.

Assumption 2. *For every $\beta, \beta' \in \mathcal{B}^*$, either $\beta \succeq_{FOSD} \beta'$ or $\beta' \succeq_{FOSD} \beta$ or both.*

Assumption 2 is trivially satisfied in games where $|B| = 2$. When $|B| \geq 3$, Assumption 2 is satisfied in the games studied by Liu and Skrzypacz (2014), where player 2 has a unique best reply to every $\alpha \in \Delta(A)$. Under their assumption, all actions in $\mathcal{B}^*$ are pure, so any pair of them can be ranked according to FOSD. A more general sufficient condition for Assumption 2 is characterized by Quah and Strulovici (2012).

I restate the no-back-loop lemma for games that satisfy Assumptions 1 and 2: For every $\sigma_2 : \mathcal{H}_2 \to \mathcal{B}^*$ and pure strategy $\widehat{\sigma}_1 : \mathcal{H}_1 \to A$ that best replies to σ_2, there is *no* $h^t \in \mathcal{H}_1(\widehat{\sigma}_1, \sigma_2) \bigcap \mathcal{H}_1^*$ such that strategy $\widehat{\sigma}_1$ plays an action that is not a^* at h^t and reaches a history in $\mathcal{H}_1(\widehat{\sigma}_1, \sigma_2) \bigcap \mathcal{H}_1^*$ in the future.

My proof in Appendix A covers this case. The intuition is that as long as player 2's action at every history belongs to $\mathcal{B}^*$, her action at any pair of histories can be ranked via FOSD, and the proof of the no-back-loop lemma in the two-action case can be directly applied to this more general case.

Prior Belief about Calendar Time: My no-back-loop lemma requires no assumption on the prior belief about calendar time. My theorems extend to settings where player 2 has an improper uniform prior belief.

Next, I explain how to extend these theorems to settings where player 1's discount factor differs from the game's continuation probability. Suppose player 1 discounts future payoffs for two reasons (i) he is indifferent between receiving one unit of utility in period t and receiving $\widehat{\delta}$ unit of utility in period $t-1$, and (ii) the game continues with probability $\overline{\delta}$ after each period. In this model, player 1 discounts future payoffs by $\delta \equiv \overline{\delta} \cdot \widehat{\delta}$ and player 2's prior belief assigns probability $(1-\overline{\delta})\overline{\delta}^t$ to the calendar time being $t \in \mathbb{N}$.

The statement of Theorem 1 remains the same. The statements of Theorems 2 and 3 need to be modified in order to take into account the difference between the game's continuation probability $\overline{\delta}$ and player 1's discount factor δ, which can be done by replacing δ, $F^\sigma(a,b)$, and U_2^σ with $\overline{\delta}$,

$$\overline{F}^\sigma(a,b) \equiv \mathbb{E}^\sigma\Big[\sum_{t=0}^{+\infty}(1-\overline{\delta})\overline{\delta}^t \mathbf{1}\{a_t = a, b_t = b\}\Big], \tag{4.2}$$

and

$$\overline{U}_2^\sigma \equiv \mathbb{E}^\sigma\Big[\sum_{t=0}^{+\infty}(1-\overline{\delta})\overline{\delta}^t u_2(a_t,b_t)\Big] = \sum_{(a,b)\in A\times B} \overline{F}^\sigma(a,b)u_2(a,b). \tag{4.3}$$

Imperfect Monitoring: I study two extensions that allow for *imperfect monitoring*. First, I assume that there is some noisy signal $\widetilde{a}_t \in A$ about a_t such that $\widetilde{a}_t = a_t$ with probability $1-\varepsilon$, and $\widetilde{a}_t$ is drawn according to some arbitrary distribution $\alpha \in \Delta(A)$ with probability ε. Player 1 observes the entire history

$h^t = \{a_s, b_s, \widetilde{a}_s\}_{s=0}^{t-1}$. Player 2_t only observes the number of times that each *signal realization* occurred in the last $\min\{t, K\}$ periods. My baseline model considers a special case in which $\varepsilon = 0$. In Online Appendix E, I establish a version of my no-back-loop lemma which holds for all α and all small enough ε. This new no-back-loop lemma can be applied to extend my theorems to settings where ε is small but positive.

Second, I study a model where the short-run players learn from *coarse summary statistics*, Following Acemoglu, Makhdoum, Malekian and Ozdaglar (2022), I assume that there exists a partition $A \equiv A_1 \cup ... \cup A_n$ such that player 2_t only observes the number of times that player 1's actions in the last $\min\{t, K\}$ periods belong to each partition element. My baseline model corresponds to the finest partition of A.

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

In Online Appendix F, I use this simple observation to show that for any $K \in \mathbb{N}$, if there exists a partition of A under which player 1 plays a^* with frequency close to one in all equilibria, then player 1 plays a^* with frequency close to one in all equilibria under partition $A = \{a^*\} \bigcup \Big(A \backslash \{a^*\}\Big)$. Hence, if the objective is to maximize the short-run players' welfare in the *worst* equilibrium, then it is optimal to disclose *only* the number of times that the patient player took the commitment action a^* in the last K periods. I also show that coarsening the summary statistics *cannot* improve the short-run players' welfare when $|A| = 2$. However, it may improve welfare when $|A| \geq 3$ and K is intermediate such that (i) b^* does not best reply to $\frac{K-1}{K}a^* + \frac{1}{K}\underline{a}$, and (ii) b^* is a strict best reply to $\frac{K-1}{K}a^* + \frac{1}{K}a'$ for some $a' \notin \{a^*, \underline{a}\}$. The intuition is that by pooling actions other than a^*, the short-run players believe that the patient player's action is $\underline{a}$ whenever his action is not a^*, which provides them a stronger incentive to punish after the patient player loses his reputation.

Observing the Exact Sequence of Actions: My baseline model focuses on the case where the short-run players *cannot* observe the exact sequence of actions. My theorems extend to the case where every short-run player observes the exact sequence of the last K actions with some small but positive probability. The proof of this robustness result is similar to the one under noisy information, which I have discussed earlier.

In Online Appendix G, I study the other extreme case where player 2 can *perfectly* observe the exact sequence of player 1's last K actions. In the *product choice game* of the introduction, I show that when K is large enough,[12] there exist equilibria in which players' payoffs are strictly bounded below $u_1(a^*, b^*)$ and $u_2(a^*, b^*)$. This suggests that observing the exact sequence of actions can lead to bad equilibrium outcomes.[13]

[12] I need K to be large enough since whether the exact sequence of actions can be observed is irrelevant when $K = 1$.

[13] This stands in contrast to repeated Bayesian games where all players are arbitrarily patient, in which Renault, Solan and Vieille (2013) show that it is sufficient for the patient uninformed player to keep track of the informed player's action frequency.

For some intuition, consider the case in which $K = 2$. Allowing player 2 to observe the exact sequence of actions opens up the possibility that (i) player 1 has a stronger incentive to play L at (H, H) than at (L, H) since player 2 plays T with lower probability at (H, H), and (ii) player 1 has an incentive to play H at (L, H) since he will receive a low payoff at (H, L), which will be reached after he plays L at (L, H). Player 1's incentive to milk his reputation and then rebuild it provides player 2 a rationale to play N at the clean history.

Such a possibility is ruled out when player 2 *cannot* observe the exact sequence since as long as player 2 plays T with higher probability at (L, H), she will also play T with higher probability at (H, L). As a result, whenever player 1 has a stronger incentive to play L at (H, H) relative to (H, L), she will have a strict incentive to play L at (L, H). This is because her payoff at (H, L) is greater than her payoff at (H, H).

Bhaskar and Thomas (2019) show that in a repeated trust game with complete information and finite memory, the patient player will shirk in all *purifiable* equilibria when his opponents can observe the exact sequence of his actions. Their result is driven by the *lack-of incentives* implied by their purifiability requirement. To see why, suppose $K = 2$. Since player 2's action does not depend on player 1's action 3 periods ago, player 1 has the *same* incentive at histories (H, H) and (L, H), which implies that player 2's action does not depend on player 1's action 2 periods ago in any purifiable equilibrium. Iterate this argument, one can obtain that *player* 2*'s action cannot depend on player* 1*'s history in any purifiable equilibrium*. Compared to the lack-of incentives to take costly actions in their model, player 1's incentive to take the highest action when he is one-step away from the clean history is *necessary* for any low-payoff equilibrium in my model.

Monotone-Submodular Payoffs: Online Appendix H examines a model where the short-run players cannot observe the exact sequence of the patient player's actions, $u_1(a, b)$ is strictly increasing in b and is strictly decreasing in a, $u_2(a, b)$ has strictly increasing differences, but $u_1(a, b)$ has *weakly decreasing differences*. This model is only one-step-away from both my baseline model and the model of Liu and Skrzypacz (2014).

Due to the challenges in constructing equilibria, I focus on a *submodular* product choice game. I show that when δ is large enough, (i) the no-back-loop lemma fails, that is, there exists a best reply of the patient player under which he milks his reputation when it is strictly positive and later restores his reputation, and (ii) if in addition, that the prior probability of commitment type π_0 is not too large, there exist equilibria where player 1's payoff is bounded below his commitment payoff when $u_1(a, b)$ is *sufficiently submodular*.

Sequential-Move Stage Games: I consider two instances in which players move sequentially in the stage game. First, suppose player 1 moves *before* player 2. If player 2 can *perfectly* observe player 1's current-period action, then player 1 will face no lack-of-commitment problem and he can secure payoff $u_1(a^*, b^*)$

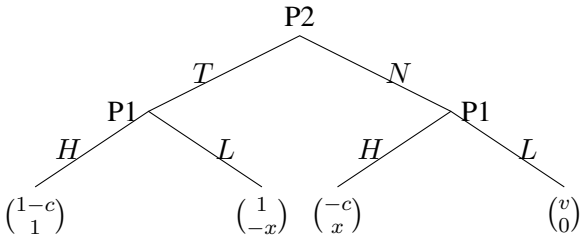


Figure 2: Sequential-Move Product Choice Game with Supermodular Payoffs

by playing a^* in every period. If player 2 observes a *noisy private signal* $y_t \sim F(\cdot|a_t)$ about player 1's current-period action a_t before choosing her own action b_t, then my no-back-loop lemma and my theorems extend as long as the distribution of private signal y_t has full support and is noisy enough. Online Appendix I.1 establishes the no-back-loop lemma in this setting. The proofs of the theorems follow from the same steps.

Online Appendix I.2 studies an extension where player 1 observes player 2's current-period action before choosing his own action. I focus on the sequential-move product choice game in Figure 2 where $v \in (0, 1-c)$, $c > 0$, and $x \in (0, 1)$. With probability $\pi_0 \in (0, 1)$, player 1 is a commitment type who plays H at every history. Focusing on the case in which $K = 1$, I use the techniques developed in the proof of Theorem 2 to fully characterize player 1's equilibrium payoff set as $\delta \to 1$. My result implies a necessary and sufficient condition under which the patient player can secure his commitment payoff $1 - c$ in all equilibria and unveils how the patient player's equilibrium payoff set depends on his reputation. In that appendix, I also discuss the economics behind the comparison between the simultaneous-move case and the sequential-move case.

Observing Previous Short-Run Players' Actions: My baseline model assumes that player 2 *cannot* observe previous short-run players' actions. This is a standard assumption in reputation models with limited memories, which is also assumed in Liu (2011), Liu and Skrzypacz (2014), Levine (2021), and so on.

An exception is Pei (2023), which shows that there exist equilibria in which player 1 receives his minmax value as long as each player 2 (i) can observe at most a bounded number of player 1's past actions and (ii) can perfectly observe *the action profile in the period before*. The intuition is that player 2's ability to observe her predecessors' actions enables her to *imitate* her predecessors and her imitation behavior can be rationalized as long as she can only observe a bounded number of the patient player's actions.

The second requirement for Pei (2023)'s result is violated in the current model not only because player 2 cannot observe her predecessors' actions, but also because when $K \geq 2$, she cannot observe which actions were taken in the period before. As a result, when player 2 observes the *summary statistics* of player 1's last K actions and the *summary statistics* of player 2's last M actions with $K, M \geq 2$, the bad equilibria in Pei

(2023) break down and characterizing the set of equilibrium payoffs, to the best of my knowledge, remains an open question.

A Proof of the No-Back-Loop Lemma

I establish the no-back-loop lemma when B is a finite set and (u_1, u_2) satisfies Assumptions 1 and 2. For every $t \geq K$, player 1's continuation value and incentive in period t depend only on $(a_{t-K}, ..., a_{t-1})$. A pure strategy $\widehat{\sigma}_1$ is *canonical* if it depends only on $(a_{t-K}, ..., a_{t-1})$. For every (σ_1, σ_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 such that (i) $h^s \succ h^t$ and (ii) h^s occurs with positive probability when the game starts from history h^t and players use strategies (σ_1, σ_2). If $\widehat{\sigma}_1$ is canonical, then $\mathcal{H}_1(\widehat{\sigma}_1, \sigma_2|h^t) = \mathcal{H}_1(\widehat{\sigma}_1, \sigma_2'|h^t)$ for every σ_2, σ_2', and h^t. Since player 2's action depends only on player 1's actions in the last K periods, for every σ_2, there exists a canonical pure strategy $\widehat{\sigma}_1$ that best replies to σ_2. Therefore, as long as there exists a pure strategy that best replies to σ_2 and violates the no-back-loop property, there also exists a canonical pure strategy the best replies to σ_2 and violates the no-back-loop property.

Fix $\sigma_2 : \mathcal{H}_2 \to \mathcal{B}^*$. Suppose by way of contradiction that there exists a canonical pure strategy $\widehat{\sigma}_1$ that best replies to σ_2 such that there exist two histories $h^t, h^s \in \mathcal{H}_1^* \bigcap \mathcal{H}_1(\widehat{\sigma}_1, \sigma_2)$ that satisfy $h^s \in \mathcal{H}_1(\widehat{\sigma}_1, \sigma_2|h^t)$, and $\widehat{\sigma}_1(h^t) = a'$ 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 $\widehat{\sigma}_1$. Let $h^{s-1} \equiv (a_0, ..., a_{s-2}) \in \mathcal{H}_1(\widehat{\sigma}_1, \sigma_2|h^t)$. 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 $(a_{s-K-1}, ..., a_{s-2}) = (a'', a^*, ..., a^*)$ for some $a'' \neq a^*$. Let β^* and V^* denote player 2's action and player 1's continuation value at histories that belong to $\mathcal{H}_1^*$. Let β' and V' denote player 2's action and player 1's continuation value when exactly one of player 1's last K actions was a' and the other $K-1$ actions were a^*. Let β'' and V'' denote player 2's action and player 1's continuation value when exactly one of player 1's last K actions was a'' and the other $K-1$ actions were a^*. Since $\widehat{\sigma}_1$ is player 1's best reply, he prefers a^* to a' at history $(a'', a^*, ..., a^*)$, or equivalently,

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

and he prefers a' to a^* at history $(a^*, ..., a^*)$, or equivalently,

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

Since player 1's stage-game payoff function is strictly supermodular, and β^* and β'' can be ranked according

to FOSD under Assumption 2, inequalities (A.1) and (A.2) imply that $\beta^* \preceq_{FOSD} \beta''$. Let

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

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

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

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 $(a_{s-K-1}, ..., a_{s-2}) = (a'', a^*, ..., a^*)$:

- **Deviation A:** Play a' in period $s-1$, $\widehat{\sigma}_1(h^\tau)$ in period $\tau + (s-t-1)$ for every $\tau \in \{t+1, ..., s-2\}$, 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

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

This leads to the following upper bound on U:

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

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

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

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

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

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

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

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

B Proof of Theorem 2

B.1 Proof of Statement 1

I define a *state* as a sequence of player 1's actions with length K, that is, $(a_{t-K}, ..., a_{t-1})$. Let $S \equiv A^K$ be the *set of states* with a typical element denoted by $s \in S$. Let $s^* \equiv (a^*, ..., a^*)$. Fix a strategy profile $\sigma \equiv (\sigma_1, \sigma_2)$. 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 $Q(s \to s')$ 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 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. The goal is to show that $\mu(s^*)$ is close to 1 in all equilibria. I state three lemmas which apply to *all* strategy profiles and *all* stage-game payoffs.

Lemma B.1. *For every strategy profile and every $s' \in S$, we have*

$$\mu(s') = (1 - \delta)p(s') + \delta \sum_{s \in S} \mu(s)Q(s \to s'). \tag{B.1}$$

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 $q_t(s \to s')$ 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 $p_{t+1}(s) = \sum_{s' \in S} p_t(s')Q(s' \to s)$. According to Bayes rule, we have

$$\mu(s) = \sum_{t=0}^{+\infty} (1 - \delta)\delta^t p_t(s) \quad \text{and} \quad Q(s \to s') = \frac{\sum_{t=0}^{+\infty} (1 - \delta)\delta^t p_t(s) q_t(s \to s')}{\sum_{t=0}^{+\infty} (1 - \delta)\delta^t p_t(s)}.$$

This implies that

$$\sum_{s\in S}\mu(s)Q(s\to s') = \sum_{s\in S}\sum_{t=0}^{+\infty}(1-\delta)\delta^t p_t(s)q_t(s\to s') = \sum_{t=0}^{+\infty}(1-\delta)\delta^t\sum_{s\in S}p_t(s)q_t(s\to s') = \sum_{t=0}^{+\infty}(1-\delta)\delta^t p_{t+1}(s')$$

$$\frac{1}{\delta}\Big\{\mu(s')-(1-\delta)p(s')\Big\} = \frac{1}{\delta}\Big\{\sum_{t=0}^{+\infty}(1-\delta)\delta^t p_t(s')-(1-\delta)p(s')\Big\} = \sum_{t=0}^{+\infty}(1-\delta)\delta^t p_{t+1}(s').$$

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

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

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

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

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

be the *outflow* from S' to states that do not belong to S'. By definition,

$$\sum_{s'\in S'}\mu(s') = \sum_{s'\in S'}\mu(s')\underbrace{\Big(\sum_{s\in S'}Q(s'\to s)+\sum_{s\notin S'}Q(s'\to s)\Big)}_{=1} = \sum_{s'\in S'}\sum_{s\in S'}\mu(s')Q(s'\to s)+\underbrace{\sum_{s'\in S'}\sum_{s\notin S'}\mu(s')Q(s'\to s)}_{\equiv\mathcal{O}(S')}.$$

For any $S'\subset S$, by summing up the two sides of equation (B.1) for all $s'\in S'$, one can obtain that

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

These equations imply that $\mathcal{I}(S') = \mathcal{O}(S')+\sum_{s'\in S'}\frac{1-\delta}{\delta}\Big\{\mu(s')-p(s')\Big\}$. Since μ and p are probability measures on S, we have $|\sum_{s'\in S'}(\mu(s')-p(s'))|\le 1$. This leads to the following lemma:

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

$$|\mathcal{I}(S')-\mathcal{O}(S')| = \Big|\sum_{s'\in S'}\frac{1-\delta}{\delta}\big(\mu(s')-p(s')\big)\Big| \le \frac{1-\delta}{\delta}. \tag{B.4}$$

I partition the set of states according to $S\equiv S_0\cup ...\cup S_K$ so that S_k is the set of states where k of the

player 1's last K actions were not a^*. By definition, $S_0 = \{s^*\}$. I further partition every $S_k \equiv \bigcup_{j=1}^{J(k)} S_{j,k}$ according to player 2's information structure, that is, two states belong to the same partition element $S_{j,k}$ if and only if player 2 distinguish between these two states. 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 ago 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 each $S_{j,k}$ 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 ago was a^*, and $s \in S'_{j,k}$ otherwise. For any $S', S'' \subset S$ with $S' \cap S'' = \emptyset$, let

$$\mathcal{Q}(S' \to S'') \equiv \sum_{s' \in S'} \sum_{s'' \in S''} \mu(s') Q(s' \to s'') \tag{B.5}$$

be the expected flow from S' to S''. According to Bayes rule, upon observing a state that belongs to $S_{j,k}$, player 2 believes that player 1's action is a^* with probability

$$\mathcal{Q}(S_{j,k} \to S_{k-1}) + \sum_{s \in S^*_{j,k}} \sum_{s' \in S_{j,k}} \mu(s) Q(s \to s'), \tag{B.6}$$

and is not a^* with probability $\mathcal{Q}(S_{j,k} \to S_{k+1}) + \sum_{s \in S'_{j,k}} \sum_{s' \in S_k} \mu(s) Q(s \to s')$. The next lemma shows that as long as both $\mathcal{Q}(S_{k-1} \to S_{j,k})$ and $\mathcal{Q}(S_{j,k} \to S_{k-1})$ 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 player 2 believes that player 1 will play a^* with probability strictly less than $\frac{K-1}{K}$ upon observing that the current state belongs to $S_{k,j}$.

Lemma B.3. *For every $z \in \mathbb{R}_+$, there exist $y \in \mathbb{R}_+$ and $\underline{\delta} \in (0,1)$ such that when $\delta > \underline{\delta}$, for every strategy profile and every $S_{j,k}$ with $k \geq 1$. If $\max\{\mathcal{Q}(S_{j,k} \to S_{k-1}), \mathcal{Q}(S_{k-1} \to S_{j,k})\} \leq z(1-\delta)$, then either*

$$\sum_{s \in S_{j,k}} \mu(s) \leq y(1-\delta) \tag{B.7}$$

or

$$\frac{\mathcal{Q}(S_{j,k} \to S_{k-1}) + \sum_{s \in S^*_{j,k}} \sum_{s' \in S_{j,k}} \mu(s) Q(s \to s')}{\mathcal{Q}(S_{j,k} \to S_{k+1}) + \sum_{s \in S'_{j,k}} \sum_{s' \in S_k} \mu(s) Q(s \to s')} < K - 1. \tag{B.8}$$

Proof. Since $\mathcal{Q}(S_{j,k} \to S_{k-1}) = \mathcal{Q}(S'_{j,k} \to S_{k-1})$, $\mathcal{Q}(S_{j,k} \to S_{k+1}) = \mathcal{Q}(S^*_{j,k} \to S_{k+1})$, and under the

hypothesis that $\mathcal{Q}(S_{j,k} \to S_{k-1}) \leq z(1-\delta)$ and $\mathcal{Q}(S_{k-1} \to S_{j,k}) \leq z(1-\delta)$, we have:

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

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

Suppose there exists no such $y \in \mathbb{R}_+$, that is, $\frac{\sum_{s\in S_{j,k}} \mu(s)}{z(1-\delta)}$ can be arbitrarily large as $\delta \to 1$. Since the sum of $\sum_{s\in S^*_{j,k}} \mu(s) - \mathcal{Q}(S^*_{j,k} \to S_{k+1}) + z(1-\delta)$ and $\sum_{s\in S'_{j,k}} \mu(s) + \mathcal{Q}(S^*_{j,k} \to S_{k+1}) - z(1-\delta)$ equals $\sum_{s\in S_{j,k}} \mu(s)$, we know that when δ is close to 1, (B.8) is implied by:

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

or equivalently,

$$\sum_{s\in S^*_{j,k}} \mu(s) < (K-1) \sum_{s\in S'_{j,k}} \mu(s) + K\mathcal{Q}(S^*_{j,k} \to S_{k+1}). \tag{B.9}$$

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

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

According to Lemma B.2, we have $\mathcal{Q}(S^*_{j,k} \to S_{k+1}) \geq \mathcal{I}(S_{j,k}) - \mathcal{Q}(S_{j,k} \to S_k \backslash S_{j,k}) - \frac{(1-\delta)(1+z\delta)}{\delta}$. 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 $\mathcal{Q}(S^*_{j,k} \to S_k \backslash S_{j,k}) = 0$. This implies that $\mathcal{Q}(S_{j,k} \to S_k \backslash S_{j,k}) =$

$\mathcal{Q}(S'_{j,k} \to S_k \backslash S_{j,k})$. Since $\mathcal{I}(S_{j,k}) = \mathcal{I}(S^*_{j,k}) + \mathcal{I}(S'_{j,k}) - \mathcal{Q}(S^*_{j,k} \to S'_{j,k}) - \mathcal{Q}(S'_{j,k} \to S^*_{j,k})$,

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

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

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

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

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

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

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

Hence, when δ is close to 1, inequality (B.9) is implied by

$$\begin{aligned}
\sum_{s \in S^*_{j,k}} \mu(s) \quad < \quad & (K-1)\Big\{\sum_{s \in S'_{j,k}} \mu(s) - \mathcal{Q}(S'_{j,k} \to S_k \backslash S_{j,k})\Big\} \\
& + (K-1)\Big\{\mathcal{I}(S^*_{j,k}) - \mathcal{Q}(S'_{j,k} \to S^*_{j,k})\Big\}. \qquad \text{(B.10)}
\end{aligned}$$

I say that a sequence of states $\{s_1, ..., s_j\} \subset S_{j,k}$ form a *connected sequence* if for every $i \in \{1, 2, ..., i-1\}$, 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 $\{s_1, ..., s_m\} \subset S_{j,k}$ with length m at least one such that (i) for every $i \in \{1, 2, ..., m-1\}$, 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 $\{s_2, ..., s_m\} \subset S^*_{j,k}$, 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

$$\begin{aligned}\mu(s_2) &\leq \mu(s_1)Q(s_1 \to s_2) + \mathcal{Q}(S\backslash S_{j,k} \to \{s_2\}) + \frac{1-\delta}{\delta} \\ &= \mu(s_1) - \mathcal{Q}(\{s_1\} \to S_k\backslash S_{j,k}) + \mathcal{Q}(S\backslash S_{j,k} \to \{s_2\}) + \frac{1-\delta}{\delta}, \end{aligned} \tag{B.11}$$

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

$$\begin{aligned}\mu(s_{i+1}) &\leq \mu(s_i)Q(s_i \to s_{i+1}) + \mathcal{Q}(S\backslash S_{j,k} \to \{s_{i+1}\}) + \frac{1-\delta}{\delta} \\ &= \mu(s_i) + \mathcal{Q}(S\backslash S_{j,k} \to \{s_{i+1}\}) + \frac{1-\delta}{\delta}. \end{aligned} \tag{B.12}$$

Iteratively apply (B.12) and (B.11) for every $i \geq 2$, we obtain:

$$\mu(s_i) \leq \mu(s_1) + \mathcal{Q}(S\backslash S_{j,k} \to \{s_2, ..., s_i\}) - \mathcal{Q}(\{s_1\} \to S_k\backslash S_{j,k}) + \frac{(1-\delta)(i-1)}{\delta}. \tag{B.13}$$

Summing up inequality (B.13) for $i \in \{2, ..., m\}$, we obtain:

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

One can obtain (B.10) by summing up the above equation for every $s_1 \in S'_{j,k}$ and taking $\delta \to 1$. This is because the left-hand-side of this sum equals $\sum_{s \in S^*_{j,k}} \mu(s)$. Therefore, after ignoring the last term that vanishes to 0 as $\delta \to 1$, the additive property of the operator $\mathcal{Q}$ implies that the right-hand-side equals

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

This establishes inequality (B.10) and leads to the conclusion of Lemma B.3 □

The next two steps make use of players' incentive constraints. First, I use Lemma B.2 and the no-back-loop lemma to show that both the inflow to $S_0 \equiv \{s^*\}$ and the outflow from S_0 are negligible as $\delta \to 1$.

Lemma B.4. *For every δ and in every Nash equilibrium under δ, $\mathcal{O}(S_0) \leq \frac{2(1-\delta)}{\delta}$ and $\mathcal{I}(S_0) \leq \frac{1-\delta}{\delta}$.*

Proof. Let S' denote a subset of states such that $s \in S'$ if and only if (i) $s \neq s^*$, and (ii) there exists a best reply $\widehat{\sigma}_1$ such that s^* is reached within a finite number of periods when the initial state is s. 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, $\mathcal{O}(\{s^*\}) \leq \frac{2(1-\delta)}{\delta}$. □

Lemma B.4 implies that the inflow to S_0 and the outflow from S_0 are both negligible. Lemma B.3 implies that for every $S_{j,1}$, either the occupation measure of states in $S_{j,1}$ is negligible, or player 2 has no incentive to play b^* at $S_{j,1}$. The next lemma shows that it *cannot* be the case that states in S_1 have significant occupation measure yet player 2 has no incentive to play b^* at $S_{j,1}$. This conclusion generalizes to every $S_{j,k}$, provided that the flow from S_{k-1} to $S_{j,k}$ and that from $S_{j,k}$ to S_{k-1} are both negligible. Iteratively apply Lemma B.3 and Lemma B.5, all states except for s^* (i.e., the unique state in S_0) have negligible occupation measure.

Lemma B.5. *Suppose u_2 is such that 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 exist $z > 0$ and $\underline{\delta} \in (0,1)$ such that for every $\delta > \underline{\delta}$, every Nash equilibrium under δ, and every $k \in \{1, 2, ..., K\}$. If $\max\{\mathcal{Q}(S_{k-1} \to S_k), \mathcal{Q}(S_k \to S_{k-1})\} < y(1-\delta)$, then $\sum_{s \in S_k} \mu(s) \leq z(1-\delta)$.*

The proof can be found by the end of this section. The intuition is that when both $\mathcal{Q}(S_{k-1} \to S_k)$ and $\mathcal{Q}(S_k \to S_{k-1})$ are negligible, $\mathcal{Q}(S_{k-1} \to S_{j,k})$ and $\mathcal{Q}(S_{j,k} \to S_{k-1})$ are also negligible for every j. Suppose by way of contradiction that $\sum_{s \in S_k} \mu(s)$ is bounded away from 0, then Lemma B.3 implies that at every $S_{j,k}$ where states in $S_{j,k}$ occur with occupation measure bounded above 0, player 2 has no incentive to play b^* at $S_{j,k}$. Since the flow from S_k to S_{k-1} is negligible, the flow from S_k to S_{k+1} must be bounded above 0. Lemma B.2 then implies that the flow from S_{k+1} to S_k is also bounded above 0. This implies that in

equilibrium, player 1 will take the most costly action a^* at some states in S_{k+1} in order to reach states in $S_{j,k}$ where player 2 has no incentive to play b^*, i.e., player 1's payoff is bounded below $u_1(a^*, b^*)$ in states that belong to $S_{j,k}$. This is suboptimal for player 1 since he can secure payoff $u_1(a^*, b^*)$ by playing a^* in every period, which implies that for every $k \geq 1$, $\sum_{s \in S_k} \mu(s) \approx 0$ in all equilibria.

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 δ, and $k \geq 1$, such that in this equilibrium, $\max\{\mathcal{Q}(S_{k-1} \to S_{j,k}), \mathcal{Q}(S_k \to S_{k-1})\} < y(1-\delta)$ but $\sum_{s \in S_k} \mu(s) > z(1-\delta)$. Pick a large enough z, Lemma B.3 implies that for every $S_{j,k} \subset S_k$, either $\sum_{s \in S_{j,k}} \mu(s) < \frac{z}{2^K}(1-\delta)$, 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 $\mathcal{Q}(S'_k \to S_{k-1})$. Let $V(s)$ be player 1's continuation value in state s and let $\overline{V} \equiv \max_{s \in S} V(s)$. Let $\underline{v}$ be player 1's lowest stage-game payoff. Let $v' \equiv \max_{a \in A, b \prec b^*} u_1(a, b)$ and $v^* \equiv u_1(a^*, b^*)$. Assumptions 1 and inequality (3.5) together imply that $v^* > v' > \underline{v}$. Since player 1 can reach any state within K periods, we have $V(s) \geq (1-\delta^K)\underline{v} + \delta^K \overline{V}$ for every $s \in S$. Theorem 1 suggests that player 1's continuation value at s^* is at least $u_1(a^*, b^*)$. Therefore, $\overline{V} \geq v^*$. Let M be the largest integer m such that

$$(1-\delta^m)v' + \delta^m \overline{V} \geq (1-\delta^K)\underline{v} + \delta^K \overline{V}. \tag{B.14}$$

Applying the L'Hospital Rule, (B.14) implies that when δ is close to 1, we have $M \leq K \frac{\overline{V}-\underline{v}}{\overline{V}-v'}$. 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 $\tau \in \{t+1, ..., t+M\}$ such that when player 1 uses this pure-strategy best reply, the state in period τ does not belong to S'_k. Therefore, $\frac{\sum_{s \in S'_k} \mu(s)}{\mathcal{O}(S'_k)} \leq \frac{1-\delta^M}{\delta^M(1-\delta)}$. When $\delta \to 1$, the RHS of the above inequality converges to M, which implies that

$$\sum_{s \in S'_k} \mu(s) \leq K \cdot \frac{\overline{V}-\underline{v}}{\overline{V}-v'} \cdot \mathcal{O}(S'_k). \tag{B.15}$$

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 $\mathcal{Q}(S_{k-1} \to S'_k) \geq \frac{\sum_{s \in S'_k} \mu(s)}{2M}$. This implies that there exists $s \in S_{k+1}$ and a canonical pure best reply $\widehat{\sigma}_1$ 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 $\widehat{\sigma}_1$ and $\widehat{\sigma}_1$ 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 $\widehat{\sigma}_1$. 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 $(1-\delta^K)\underline{v} + \delta^K v^*$. 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 $\widehat{\sigma}_1$. Starting from state s, this strategy and $\widehat{\sigma}_1$ 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 $\widehat{\sigma}_1$ is player 1's best reply to player 2's equilibrium strategy. □

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

B.2 Constructing Equilibria where a^* Occurs with Low Frequency

Case 1: I consider the case where (u_1, u_2) satisfies (3.5) but $K \geq \overline{K}$. Since $K \geq \overline{K}$, b^* best replies to $\frac{K-1}{K}a^* + \frac{1}{K}a'$ for some $a' \neq a^*$. Inequality (3.5) implies that every best reply to a' is strictly lower than b^*. Hence, $K \geq 2$ and there exists $\alpha \in (0, \frac{K-1}{K})$ such that $\{b^*, b'\} \subset \text{BR}_2(\alpha a^* + (1-\alpha)a')$ for some $b' \prec_B b^*$. Let b'' be player 2's lowest best reply to a'. Since $u_2(a,b)$ has strictly increasing differences, we know that $b'' \preceq_B b' \prec_B b^*$. Let $\mathcal{H}_1^{**}$ be the set of histories such that player 2 observes at most one a' and does not observe any action other than a^* and a', which will contain the set of histories that occur with positive probability. Player 2's belief is derived from Bayes rule at every history that belongs to $\mathcal{H}_1^{**}$. For player 2's belief at histories that occur with zero probability,

1. If player 2_t observes two or more a' and observes no action other than a^* and a', then she believes that $(a_{t-2}, a_{t-1}) = (a', a')$.

2. If player 2 observes $a'' \notin \{a^*, a'\}$, then she believes that the action in the period before is a''.

Then I describe player 1's equilibrium strategy. At every history that belongs to $\mathcal{H}_1^{**}$, player 1 plays a' in period t if $t = K-1$ or $t \geq K$ and $(a_{\min\{0,t-K+1\}}, ...a_{t-1}) = (a^*, ..., a^*)$. Player 1 plays a^* in period t at other histories that belong to $\mathcal{H}_1^{**}$. Histories that do not belong to $\mathcal{H}_1^{**}$ occur with zero probability, at which player 1's behavior is given by:

1. Player 1 plays a^* if $(a_{t-2}, a_{t-1}) \neq (a', a')$ and the last $\min\{K, t\}$ actions are either a^* or a'.

2. Player 1 plays a^* with probability α and plays a' with probability $1-\alpha$ if $(a_{t-2}, a_{t-1}) = (a', a')$ and actions in the last $\min\{K, t\}$ periods are either a^* or a'.

3. Player 1 plays a' in period t if actions other than a^* and a' occurred in period $t-1$.

Player 2 plays b^* at every history that belongs to $\mathcal{H}_1^{**}$. At every history that (i) does not belong to $\mathcal{H}_1^{**}$, and (ii) actions other than a^* and a' do not occur in the last K periods, player 2 plays b^* with probability β and plays b' with probability $1-\beta$, where

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

$$+\delta^{K-1} \underbrace{\frac{u_1(a', b^*) + (\delta + \delta^2 + ... + \delta^{K-1}) u_1(a^*, b^*)}{1 + \delta + ... + \delta^{K-1}}}_{\equiv V_K}. \tag{B.16}$$

Since $u_1(a', b^*) > u_1(a^*, b^*) > u_1(a', b') > u_1(a^*, b')$, β is strictly between 0 and 1. At every history that does not belong to $\mathcal{H}_1^{**}$ and actions other than a^* and a' occurred in the last K periods, player 2 plays b''.

Player 2's incentive constraint at every history that occurs with zero probability is satisfied under her belief since (i) she mixes between b^* and b' whenever she believes that player 1 plays $\alpha a^* + (1-\alpha)a'$, and (ii) she plays b'' whenever she believes that player 1 plays a'. At histories that occur with positive probability, player 2 believes that a' is played with probability $1-\pi_0$ and a^* is played with probability π_0 in period 0, so she plays a best reply to this mixed action. From period 1 to $K-1$, player 2 believes that a^* is played by both types, so she plays her best reply b^*. After period K, player 2's belief assigns probability 1 to the commitment type upon observing any history where player 1's last K actions were a^*, and therefore, she has a strict incentive to play b^*. For player 2's incentive constraints at histories where a' occurred only once, she

believes that $(a_{t-K}, ..., a_{t-1}) = (a', a^*, ..., a^*)$ with probability

$$\frac{\delta^{K-1}}{1 + \delta + ... + \delta^{K-1}} \tag{B.17}$$

and $(a_{t-K}, ..., a_{t-1}) \neq (a', a^*, ..., a^*)$ with complementary probability. When $\delta \to 1$, expression (B.17) is less than but converges to $\frac{1}{K}$. Since player 1 plays a' when $(a_{t-K}, ..., a_{t-1}) = (a', a^*, ..., a^*)$ and plays a^* at other histories where a' occurred once, player 2 believes that player 1's current period action is a' with probability less than $\frac{1}{K}$ and is a^* with probability more than $\frac{K-1}{K}$. Hence, there exists $\underline{\delta} \in (0, 1)$ such that when $\delta > \underline{\delta}$, player 2s have a strict incentive to play b^* if a' occurred only once in the last K periods.

I verify player 1's incentive constraint: (i) he has no incentive to reach any history that occurs with zero probability starting from any history that occurs with positive probability, and (ii) he has an incentive to play a' when $(a_{t-2}, a_{t-1}) = (a', a')$ or when $a_{t-1} \notin \{a', a^*\}$. When $(a_{t-2}, a_{t-1}) = (a', a')$, player 1 is indifferent between playing a' and a^* in period t.

Next, I show that player 1 has no incentive to play a' at every history that belongs to $\mathcal{H}_1^{**}$ where

$$(a_{t-K+1}, ..., a_{t-1}) \neq (a^*, ..., a^*).$$

For every $m \in \{1, 2, ..., K\}$, let s_m be the state where $a_{t-m} = a'$ and all actions that belong to $\{a_{t-K}, .., a_{t-1}\} \backslash \{a_{t-m}\}$ are a^*. Let V_m player 1's continuation value in state s_m. For every $m \in \{1, 2, ..., K-1\}$, player 1 prefers a^* to a' in state s_m if

$$V_m > (1-\delta)u_1(a', b^*) + \delta(1 - \delta^{K-m})u_1(a^*, \beta) + (\delta^{K-m+1} - \delta^K)u_1(a^*, b^*) + \delta^K V_K. \tag{B.18}$$

Since

$$V_m = (1-\delta)^{K-m}u_1(a^*, b^*) + \delta^{K-m}V_K \text{ for every } 1 \leq m \leq K,$$

we have:

$$\begin{aligned}
&(1-\delta)(1-\delta^{K-m})u_1(a^*, b^*) + \delta(1-\delta^{K-m})\Big(u_1(a^*, b^*) - u_1(a^*, \beta)\Big) \\
&-(1-\delta)u_1(a', b^*) + \delta^{K-m}(1-\delta^m)V_k - (\delta^{K-m+1} - \delta^K)u_1(a^*, b^*) > 0.
\end{aligned} \tag{B.19}$$

Dividing the above expression by $1 - \delta$, and then taking the limit where $\delta \to 1$, we obtain that inequality (B.19) is true when δ is close to 1 if

$$(K-m)\Big(u_1(a^*, b^*) - u_1(a^*, \beta)\Big) + mV_K - (m-1)u_1(a^*, b^*) - u_1(a', b^*) > 0,$$

or equivalently,

$$\underbrace{(K-m)}_{\geq 1}(1-\beta)\Big(u_1(a^*,b^*)-u_1(a^*,b')\Big) > (m-1)\underbrace{\Big(u_1(a^*,b^*)-V_K\Big)}_{<0}+\Big(u_1(a',b^*)-V_K\Big). \quad \text{(B.20)}$$

When δ is close to 1,

$$V_K \approx \frac{1}{K}u_1(a',b^*)+\frac{K-1}{K}u_1(a^*,b^*),$$

and therefore,

$$1-\beta=\frac{u_1(a',b^*)-V_K}{u_1(a',b^*)-u_1(a',b')} \approx \frac{K-1}{K}\cdot\frac{u_1(a',b^*)-u_1(a^*,b^*)}{u_1(a',b^*)-u_1(a',b')} > \frac{K-1}{K}\cdot\frac{u_1(a',b^*)-u_1(a^*,b^*)}{u_1(a^*,b^*)-u_1(a^*,b')},$$

where the last inequality follows from $u_1(a,b)$ having strictly increasing differences. This implies that

$$(1-\beta)\Big(u_1(a^*,b^*)-u_1(a^*,b')\Big) > u_1(a',b^*)-V_K \approx \frac{K-1}{K}\Big(u_1(a',b^*)-u_1(a^*,b^*)\Big).$$

Inequality (B.20) is true when δ is close to 1 since $K-m \geq 1$ and $u_1(a',b^*)-V_K$ converges to $\frac{K-1}{K}(u_1(a',b^*)-u_1(a^*,b^*))$ as $\delta \to 1$.

Next, I show that player 1 has no incentive to play actions other than a' and a^* at every history that belongs to $\mathcal{H}_1^{**}$. It is straightforward to show that he has no incentive to play any action that does not belong to $\{a^*,a',\underline{a}\}$, since playing $\underline{a}$ leads to a strictly higher stage-game payoff for player 1 while not lowering his continuation value. Hence, I only need to show that when $a' \neq \underline{a}$, player 1 has no incentive to play $\underline{a}$ at any history that belongs to $\mathcal{H}_1^{**}$. This is because his payoff at any history that occurs with positive probability is bounded from below by $V_1 \approx \frac{1}{K}u_1(a',b^*)+\frac{K-1}{K}u_1(a^*,b^*) > u_1(a^*,b^*)$. Inequality (3.5) implies that player 1's stage-game payoff is strictly less than $u_1(a^*,b^*)$ when player 2's action is strictly lower than b^*. Since player 2 plays b'' when actions other than a^* and a' occurred in the last K periods, player 1's continuation value when he plays $\underline{a}$ is at most:

$$(1-\delta)u_1(\underline{a},b^*)+(\delta-\delta^2)u_1(a',b')+(\delta^2-\delta^{K+1})u_1(a^*,b')+\delta^{K+1}V_K$$

Since $V_1 < V_2 < ... < V_K$, it is sufficient to show that

$$V_1=(1-\delta^{K-1})u_1(a^*,b^*)+\delta^{K-1}V_K > (1-\delta)u_1(\underline{a},b^*)+(\delta-\delta^2)u_1(a',b')+(\delta^2-\delta^{K+1})u_1(a^*,b')+\delta^{K+1}V_K.$$

or equivalently,

$$(1-\delta^{K-1})u_1(a^*,b^*)+(\delta^{K-1}-\delta^{K+1})V_K-(1-\delta)u_1(\underline{a},b^*)-(\delta-\delta^2)u_1(a',b')-(\delta^2-\delta^{K+1})u_1(a^*,b')>0.$$

Dividing the left-hand-side of the above inequality by $1-\delta$ and then taking the $\delta \to 1$ limit, we know that the above inequality is true when δ is close to 1 if

$$(K-1)u_1(a^*,b^*)+2V_K \geq u_1(\underline{a},b^*)+u_1(a',b')+(K-1)u_1(a^*,b'). \tag{B.21}$$

Since u_1 has strictly increasing differences, we have:

$$u_1(\underline{a},b^*) < u_1(a',b^*)-u_1(a',b')+u_1(\underline{a},b') \leq u_1(a',b^*)-u_1(a',b')+u_1(a^*,b^*)$$

$$< u_1(a^*,b^*)+u_1(a',b')-u_1(a^*,b')-u_1(a',b')+u_1(a^*,b^*) = 2u_1(a^*,b^*)-u_1(a^*,b').$$

So the right-hand-side of (B.21) is bounded from above by $2u_1(a^*,b^*)+(K-2)u_1(a^*,b')+u_1(a',b')$, which is strictly less than $(K+1)u_1(a^*,b^*)$. Since $V_K > u_1(a^*,b^*)$, the left-hand-side of (B.21) is strictly greater than $(K+1)u_1(a^*,b^*)$. This establishes inequality (B.21).

In the last step, I show that if any action other than a^* and a' occurred in period $t-1$, player 1 has an incentive to play a' in period t. Since $a_{t-1} \notin \{a^*,a'\}$, player 2's actions from period t to period $t+K-1$ are b'' regardless of player 1's behavior in those periods, and moreover, player 2 has an incentive to play actions greater than b'' in period $s(\geq t+K)$ only if player 1 has played a^* at least $K-1$ times and a' at least once after the last time they played actions other than a^* and a'. Since $V_K > V_{K-1} > ... > V_1$, player 1's continuation value in period t is bounded from above by:

$$(1-\delta)u_1(a',b'')+(\delta-\delta^K)u_1(a^*,b'')+\delta^K V_K.$$

This upper bound is attained when player 1 plays a' in period t and plays a^* in the next $K-1$ periods, after which play reaches state s_K and player 1's continuation value is V_K. This verifies his incentive to play a' when his previous period action was neither a^* nor a'.

Case 2: I consider the case where (u_1,u_2) violates (3.5). Then there exist $a' \neq a^*$ and b' (notice that b' may equal b^*) such that b' best replies to a' and $u_1(a',b') \geq \max_{a\in A}\max_{b\in \text{BR}_2(a)} u_1(a,b)$. By definition, $u_1(a',b') \geq u_1(a^*,b^*)$. I construct equilibria where the rational-type of player 1 plays a' in every period and

for every $t \geq 1$, player 2_t plays b^* if a^* was played in each of the last K periods and plays b' if the former is not the case and no action except for a^* and a' appeared in the last $\min\{t, K\}$ periods. The rest of the construction considers two subcases separately.

If a' is player 1's lowest action, then at every history that occurs with positive probability, player 1 plays a' and player 2 plays b'. Obviously, player 1 has no incentive to play actions other than a' at any history and player 2's strategy is also optimal given her belief, which verifies that this is an equilibrium.

If a' is not player 1's lowest action, then let a'' be player 1's lowest action. By definition, there exists $\phi \in (0,1)$ such that b' as well as an action strictly lower than b', denoted by b'', are both best replies to $\alpha \equiv \phi a' + (1-\phi)a''$. Upon observing any history that occurs with zero probability, if there exists any action that is neither a' nor a^*, player 2 believes that it occurred in the period before. At every history that occurs with zero probability, player 1 plays a' if $a_{t-1} \in \{a^*, a'\}$ and plays α if $a_{t-1} \notin \{a^*, a'\}$. Since player 2 assigns probability 1 to $a_{t-1} \notin \{a', a^*\}$ when she observes at least one action that is not a' and a^*, she has an incentive to mix between b' and b'', where his probability of playing b' is denoted by β and is given by

$$\beta u_1(a'', b') + (1-\beta)u_1(a'', b'') = (1-\delta^K)\Big(\beta u_1(a', b') + (1-\beta)u_1(a', b'')\Big) + \delta^K u_1(a', b'). \quad \text{(B.22)}$$

This implies that at a history where $a_{t-1} \notin \{a^*, a'\}$, player 1 is indifferent between playing a' and a'', and given that $u_1(a^*, b^*) \leq u_1(a', b')$, he strictly prefers a'' to any action that is not a' or a''. What remains to be verified is that at a history where the last K actions were a', player 1 has no incentive to play actions other than a'. First, playing a^* is suboptimal given that $u_1(a^*, b^*) \leq u_1(a', b')$ and playing actions other than a^* and a' is strictly dominated by playing a''. Hence, I only need to verify that player 1 has no incentive to play a''. Player 1's payoff when he plays a'' at history $(a_{t-K}, ..., a_{t-1}) = (a', ..., a')$ is

$$(1-\delta)u_1(a'', b') + \delta(1-\delta^K)\Big(\beta u_1(a', b') + (1-\beta)u_1(a', b'')\Big) + \delta^{K+1}u_1(a', b'), \quad \text{(B.23)}$$

which by the definition of β in (B.22) as well as the assumption that u_1 has strictly increasing differences, implies that (B.23) is strictly smaller than $u_1(a', b')$. This verifies that player 1 has no incentive to deviate.

References

[1] Acemoglu, Daron, Ali Makhdoum, Azarakhsh Malekian and Asu Ozdaglar (2022) "Learning from Reviews: The Selection Effect and the Speed of Learning," *Econometrica*, 90, 2857-2899.

[2] Acemoglu, Daron and Alexander Wolitzky (2014) "Cycles of Conflict: An Economic Model," *American Economic Review*, 104, 1350-1367.

[3] Bhaskar, V. and Caroline Thomas (2019) "Community Enforcement of Trust with Bounded Memory," *Review of Economic Studies*, 86, 1010-1032.

[4] Board, Simon and Moritz Meyer-ter-Vehn (2013) "Reputation for Quality," *Econometrica*, 81(6), 2381-2462.

[5] Cai, Hongbin, Ginger Zhe Jin, Chong Liu and Li-an Zhou (2014) "Seller Reputation: From Word-of-Mouth to Centralized Feedback," *International Journal of Industrial Organization*, 34, 51-65.

[6] Clark, Daniel, Drew Fudenberg and Alexander Wolitzky (2021) "Record-Keeping and Cooperation in Large Societies," *Review of Economic Studies*, 88, 2179-2209.

[7] Cripps, Martin, George Mailath and Larry Samuelson (2004) "Imperfect Monitoring and Impermanent Reputations," *Econometrica*, 72, 407-432.

[8] Cripps, Martin and Caroline Thomas (2019) "Strategic Experimentation in Queues," *Theoretical Economics*, 14, 647-708.

[9] Deb, Joyee (2020) "Cooperation and Community Responsibility," *Journal of Political Economy*, 128, 1976-2009.

[10] Deb, Joyee, Takuo Sugaya and Alexander Wolitzky (2020) "The Folk Theorem in Repeated Games With Anonymous Random Matching," *Econometrica*, 88, 917-964.

[11] Dellarocas, Chrysanthos (2006) "Reputation Mechanisms" Handbook on Information Systems and Economics, T. Hendershott (ed.), Elsevier Publishing, 629-660.

[12] Ekmekci, Mehmet (2011) "Sustainable Reputations with Rating Systems," *Journal of Economic Theory*, 146, 479-503.

[13] Ekmekci, Mehmet, Olivier Gossner and Andrea Wilson (2012) "Impermanent Types and Permanent Reputations," *Journal of Economic Theory*, 147, 162-178.

[14] Ekmekci, Mehmet and Lucas Maestri (2022) "Wait or Act Now? Learning Dynamics in Stopping Games," *Journal of Economic Theory*, 205, 105541.

[15] Ellison, Glenn (1994) "Cooperation in the Prisoner's Dilemma with Anonymous Random Matching," *Review of Economic Studies*, 61, 567-588.

[16] Ely, Jeffrey and Juuso Välimäki (2003) "Bad Reputation," *Quarterly Journal of Economics*, 118, 785-814.

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

[18] Fudenberg, Drew and David Levine (1992) "Maintaining a Reputation when Strategies are Imperfectly Observed," *Review of Economic Studies*, 59, 561-579.

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

[20] Gossner, Olivier (2011) "Simple Bounds on the Value of a Reputation," *Econometrica*, 79, 1627-1641.

[21] Heller, Yuval and Erik Mohlin (2018) "Observations on Cooperation," *Review of Economic Studies*, 88, 1892-1935.

[22] Hu, Ju (2020) "On the Existence of the Ex Post Symmetric Random Entry Model," *Journal of Mathematical Economics*, 90, 42-47.

[23] Jehiel, Philippe and Larry Samuelson (2012) "Reputation with Analogical Reasoning," *Quarterly Journal of Economics*, 127(4), 1927-1970.

[24] Kandori, Michihiro (1992) "Social Norms and Community Enforcement," *Review of Economic Studies*, 59, 63-80.

[25] Kaya, Ayça and Santanu Roy (2022) "Market Screening with Limited Records," *Games and Economic Behavior*, 132, 106-132.

[26] Levine, David (2021) "The Reputation Trap," *Econometrica*, 89, 2659-2678.

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

[28] Liu, Qingmin (2011) "Information Acquisition and Reputation Dynamics," *Review of Economic Studies*, 78, 1400-1425.

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

[30] Mailath, George and Larry Samuelson (2001) "Who Wants a Good Reputation?" *Review of Economic Studies*, 68, 415-441.

[31] Pei, Harry (2020) "Reputation Effects under Interdependent Values," *Econometrica*, 88(5), 2175-2202.

[32] Pei, Harry (2023) "Reputation Building under Observational Learning," *Review of Economic Studies*, 90(3), 1441-1469.

[33] Quah, John and Bruno Strulovici (2012) "Aggregating the Single Crossing Property," *Econometrica*, 80, 2333-2348.

[34] Renault, Jérôme, Eilon Solan and Nicolas Vieille (2013) "Dynamic Sender-Receiver Games," *Journal of Economic Theory*, 148, 502-534.

[35] Sorin, Sylvain (1999) "Merging, Reputation, and Repeated Games with Incomplete Information," *Games and Economic Behavior*, 29, 274-308.

[36] 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.

[37] Tadelis, Steven (2016) "Reputation and Feedback Systems in Online Platform Markets," *Annual Review of Economics*, 8, 321-340.

[38] Takahashi, Satoru (2010) "Community Enforcement When Players Observe Partners' Past Play," *Journal of Economic Theory*, 145, 42-62.

[39] Vong, Allen (2022) "Certification for Consistent Quality Provision," Working Paper.