Reputation Building under Observational Learning

Harry PEI*

August 29, 2021

Abstract: A patient seller interacts with a sequence of myopic consumers. Each consumer decides whether to trust the seller after observing a bounded number of the seller's past actions and some or all previous consumers' choices. With positive probability, the seller is a type that commits to play his Stackelberg action. I show that consumers' ability to observe other consumers' choices can lead to reputation failures: There exist equilibria where every consumer imitates her predecessor with high probability and the patient seller receives his minmax payoff. Furthermore, the seller receives his minmax payoff in all equilibria where consumers do not trust him in the first period and do not trust him when the worst action profile occurred in the period before. In an extension where every consumer observes all previous consumers' choices and an unboundedly informative private signal about the seller's current-period action, the seller receives at least his Stackelberg payoff in all equilibria.

Keywords: reputation, reputation failure, imitation, observational learning.

JEL Codes: C73, D82, D83

1 Introduction

Economists have recognized that consumers' choices are influenced by other consumers' choices (Banerjee 1992, Bichandarni, Hirshleifer and Welsh 1992). This is the case in some informal markets in developing countries, where there is limited product standard enforcement and firms' records are often unavailable or incomplete due to the lack of record-keeping institutions. When a consumer has limited information about the seller's past records, she might benefit from observing other consumers' choices since those consumers may know something about the seller that she does not know.

This paper examines how consumers' observational learning affects sellers' returns from building reputations. My main result demonstrates the fragility of reputation effects when consumers can observe other consumers' choices but can only observe a limited number of the seller's past actions.

*Department of Economics, Northwestern University. I thank Daron Acemoglu, S. Nageeb Ali, Alp Atakan, Jie Bai, Dhruva Bhaskar, Sushil Bikhchandani, Joyee Deb, Eddie Dekel, Mira Frick, Drew Fudenberg, Olivier Gossner, Daniel Hauser, Kevin He, Ju Hu, Yuhta Ishii, Navin Kartik, Elliot Lipnowski, George Lukyanov, George Mailath, Moritz Meyer-ter-Vehn, Wojciech Olszewski, Mallesh Pai, Evan Sadler, Peter Norman Sørensen, Bruno Strulovici, Adam Szeidl, Caroline Thomas, Juuso Toikka, Chris Udry, Udayan Vaidya, Juuso Välimäki, Alex Wolitzky, Tomer Yehoshua-Sandak, and four anonymous referees for helpful comments. I thank NSF Grant SES-1947021 for financial support.

I study a repeated game between a patient seller and a sequence of myopic consumers, arriving one in each period and each plays the game only once. Players' stage-game payoffs satisfy a monotone-supermodularity condition. A leading example that satisfies my condition is the product choice game:[1]

seller \ consumer	**T**rust	**N**o Trust
High Effort	$1, 2$	$-c_N, 1$
Low Effort	$1 + c_T, -1$	$0, 0$

where $c_N, c_T > 0$.

I use this example to illustrate my results throughout this section. The seller observes the past actions of all players, and is either a strategic type who maximizes his discounted average payoff, or a commitment type who plays his Stackelberg action (in the example, it is *high effort*) in every period.

My modeling innovation is that every consumer observes the seller's actions in the last $K \in \mathbb{N}$ periods and consumers' actions in the last $M \in \mathbb{N} \cup \{+\infty\}$ periods. I assume that K is finite and $M \geq 1$. My assumption fits when consumers learn about the seller both via observational learning—from which they learn about previous consumers' choices, and via word-of-mouth communication with other consumers—from which they learn about the seller's actions against these consumers. I require that each consumer can observe at least her immediate predecessor's action, and can talk to at most K predecessors due to constraints on her time and attention.

My main result, Theorem 1, shows that when the probability of commitment type is below some cutoff, there exist equilibria where the patient seller receives his minmax payoff. This stands in contrast to Fudenberg and Levine (1989)'s theorem, which shows that the patient seller receives at least his Stackelberg payoff in all equilibria when consumers can observe the entire history of his actions.

Intuitively, the seller receives a low payoff when the first consumer does not trust him and every consumer imitates her predecessor with high probability. This is because the seller receives a low stage-game payoff in the first period and imitation makes consumers' actions as well as the seller's payoffs persistent. Imitation is *feasible* when consumers can observe their predecessors' choices. Imitation can be *rationalized* when each consumer observes at most a bounded number of the seller's actions.

In contrast, when consumers can observe *the seller's entire history*, imitating a predecessor who did not trust the seller is *not* incentive compatible. This is because after sufficiently many periods where consumers believe that the seller is likely to exert low effort but ends up observing high effort, consumers' posterior belief will attach probability close to one to the commitment type. After that, they will have a strict incentive to trust the seller. I also use an example to show that when consumers

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

cannot observe any other consumer's action, the seller receives at least his Stackelberg payoff in all equilibria. This implies that consumers' observational learning may hurt the seller.

The mechanism behind my result hinges on *consumers imitating their predecessors* and *consumers not trusting newly arrived sellers who have no past record*, both of which are plausible as suggested by existing empirical findings. On consumers' imitation, Cai, Chen and Fang (2009) and Zhang (2010) find that people imitate their predecessors when they decide what food to buy and whether to accept a kidney transplant. On consumers not trusting newly arrived sellers, Michelson et al. (2021) find that a large fraction of sampled farmers in Tanzania suspect that the fertilizers sold in local markets are adulterated and their pessimistic beliefs about the seller's integrity persist over time.

Although Theorem 1 only displays one equilibrium under which reputation effects fail, several lessons apply more broadly. First, I show that the seller receives his minmax payoff in *all equilibria* that satisfy the following refinement: (1) consumers do not trust him when the worst action profile (L, N) occurred in the period before, and (2) consumers do not trust him in the first period. The first requirement is rather standard, which is satisfied both by my low-payoff equilibrium and by some equilibria where the seller receives a high payoff, such as grim-trigger equilibrium. The second requirement is satisfied by my low-payoff equilibrium but is violated by grim-trigger equilibrium, since grim-trigger requires consumers *do* trust the seller in the first period. While there are applications where consumers trust newly arrived sellers, there are also situations where my *no initial trust* condition fits better.

Second, in my low-payoff equilibrium, if the seller exerts high effort in every period, (1) consumers never herd on action N, and (2) the seller receives a high *undiscounted average payoff*. I show that consumers cannot herd on N in any equilibrium under any prior belief and any discount factor. This stands in contrast to the canonical social learning results where inefficiencies are caused by herding. If each consumer observes *all* previous consumers' choices (i.e., $M = +\infty$), then in *all* equilibria under any prior belief and any discount factor, the seller's undiscounted average payoff from exerting high effort in every period is at least $\frac{K}{K+1}$ times his Stackelberg payoff plus $\frac{1}{1+K}$ times his minimal stage-game payoff. When this guaranteed undiscounted payoff is greater than the seller's minmax payoff 0, a reputation-building seller can eventually secure a strictly positive payoff. This does not contradict Theorem 1 since the time it takes for the seller to secure this positive payoff can depend on his discount factor. For example, in my low-payoff equilibrium, it takes longer for a more patient seller to switch consumers' actions from N to T. It is the prolonged process of trust building that wipes out the seller's benefit from being more patient.

Next, I consider an extension where in addition to what she observes in the baseline model, each consumer also observes an informative private signal about the seller's current-period action whose distribution satisfies a monotone likelihood ratio property. Motivated by canonical social learning models, I focus on the case where each consumer observes *all* previous consumers' choices ($M = +\infty$).

Theorem 2 shows that the seller can secure his Stackelberg payoff in *all* equilibria when consumers' private signals are *unboundedly informative* about the seller's Stackelberg action. In the product choice game, unbounded informativeness means that some signals are arbitrarily more likely to occur under high effort compared to that under low effort. This unboundedly informative private signal guarantees a positive lower bound on the informativeness of every consumer's action about the seller's current-period action. Importantly, this lower bound does not depend on the seller's discount factor. Since every consumer can observe all previous consumers' choices, the argument in Fudenberg and Levine (1992) implies that the patient seller can secure his Stackelberg payoff in all equilibria.

In contrast, Theorem 3 shows that the seller receives his minmax payoff in some equilibria when consumers' private signals are not unboundedly informative about his Stackelberg action and the prior probability of commitment type is below some cutoff.[2] Intuitively, when consumers believe that the seller is likely to exert low effort, observing the realization of a boundedly informative signal cannot convince them to trust the seller. Similar to the baseline model, consumers have incentives to imitate their predecessors when each consumer observes at most a bounded number of the seller's actions, and the seller receives a low payoff when he is not trusted in the first period.

2 Baseline Model

Time is discrete, indexed by $t = 0, 1, \ldots$ A long-lived player 1 (he, e.g., a seller) with discount factor $\delta \in (0, 1)$ interacts with an infinite sequence of short-lived player 2s (she, e.g., consumers), arriving one in each period, each plays the game only once, with 2_t denoting player 2 who arrives in period t.

In period t, player 1 chooses $a_t \in A$ and player 2_t chooses $b_t \in B$. I assume that both A and B are finite sets. Player $i \in \{1, 2\}$'s stage-game payoff is $u_i(a_t, b_t)$. Let $\mathrm{BR}_2(a) \subset B$ be player 2's best reply to a. Player 1's (pure) Stackelberg action is $\arg\max_{a \in A} \left\{ \min_{b \in \mathrm{BR}_2(a)} u_1(a, b) \right\}$.

Assumption 1. *Player* 1 *has a unique best reply to every pure action* $b \in B$. *Player* 2 *has a unique best reply to every pure action* $a \in A$. *Player* 1 *has a unique Stackelberg action.*

Since A and B are finite sets, Assumption 1 is satisfied for generic (u_1, u_2). Let a^* be player 1's Stackelberg action. I focus on games with monotone-supermodular payoffs, which have been studied in the reputation literature by Phelan (2006), Ekmekci (2011), and Liu (2011).

[2] When player 1 has three or more actions, Theorem 3 requires a more demanding condition compared to the consumers' private signals not being unboundedly informative about the seller's Stackelberg action.

Assumption 2. *Players' stage-game payoffs* (u_1, u_2) *are monotone-supermodular if there exist a complete order on* A, $\succ_A$, *and a complete order on* B, $\succ_B$, *such that:*

1. *Player* 1*'s payoff function* $u_1(a, b)$ *is strictly decreasing in* a *and is strictly increasing in* b.

2. *Player* 2*'s payoff function* $u_2(a, b)$ *has strictly increasing differences in* (a, b).

3. *Player* 1*'s Stackelberg action* a^* *is not the lowest element of* A.

The product choice game in the introduction satisfies Assumption 2 once we rank players' actions according to $H \succ_A L$ and $T \succ_B N$. This is because consumers have stronger incentives to trust the seller when the seller exerts higher effort, the seller prefers to exert low effort but benefits from consumers' trust, and the seller's Stackelberg action H differs from his lowest-cost action L.

Before choosing a_t, player 1 observes all the past actions $(a_0, ..., a_{t-1}, b_0, ..., b_{t-1})$ and his perfectly persistent type $\omega \in \{\omega_s, \omega_c\}$. Let ω_c stand for a *commitment type* who plays a^* in every period. Let ω_s stand for a *strategic type* who maximizes his discounted average payoff $\sum_{t=0}^{\infty}(1-\delta)\delta^t u_1(a_t, b_t)$. That is, player 1's payoff is normalized so that the weight on his period-t payoff is $(1-\delta)\delta^t$ and the sum of weights is 1. Let $\pi_0 \in (0, 1)$ be the prior probability of the commitment type.

My modeling innovation is on player 2's information structure. I assume that there exist $K \in \mathbb{N}$ and $M \in \mathbb{N} \cup \{+\infty\}$ such that for every $t \in \mathbb{N}$, player 2_t can observe player 1's actions in the last K periods $(a_{\max\{0,t-K\}}, ..., a_{t-1})$ and player 2's actions in the last M periods $(b_{\max\{0,t-M\}}, ..., b_{t-1})$, where $M = +\infty$ means that every player 2 can observe the entire history of her predecessors' choices.[3]

1. I assume that K is finite.[4] That is, every consumer observes a bounded number of the seller's actions. This stands in contrast to the reputation model of Fudenberg and Levine (1989) where every consumer observes the entire history of the seller's actions (i.e., $K = +\infty$).

2. I assume that $M \geq 1$. That is, every consumer can observe at least her immediate predecessor's action. This stands in contrast to existing reputation models with limited memories such as Liu (2011) and Liu and Skzypacz (2014) where consumers cannot observe other consumers' choices.

Let $\mathcal{H}_i$ be the set of player $i \in \{1, 2\}$'s private histories. Strategic-type player 1's strategy is $\sigma_1 : \mathcal{H}_1 \to \Delta(A)$. Player 2s' strategy is $\sigma_2 : \mathcal{H}_2 \to \Delta(B)$. The solution concept is Perfect Bayesian equilibrium (or PBE or equilibrium), which consists of a strategy for the strategic-type player 1, a strategy for player 2s, and a system of beliefs that satisfy the standard requirements.

[3]If $M = +\infty$, then every player 2 can infer calender time from her history. If M is finite, then player 2_t cannot infer calendar time when $t \geq M$. My main result, Theorem 1, applies (1) when $M = +\infty$, (2) when M is finite and player 2s *cannot* directly observe calender time, and (3) when M is finite but player 2s *can* directly observe calendar time.

[4]My three theorems extend to the case where $K = 0$. The proof of Theorem 2 remains the same while the constructive proofs of Theorems 1 and 3 need to be modified. The details are available upon request.

3 Main Result

Recall that a^* is player 1's Stackelberg action. Let $b^* \equiv \text{BR}_2(a^*)$. Player 1's *Stackelberg payoff* is $u_1(a^*, b^*)$. Let a' be the lowest element of A. Let $b' \equiv \text{BR}_2(a')$. The first two parts of Assumption 2 imply that $u_1(a', b')$ is player 1's *minmax payoff* in the sense of Fudenberg, Kreps and Maskin (1990). Assumption 1 and the third part of Assumption 2 imply that $a^* \neq a'$ and $u_1(a', b') < u_1(a^*, b^*)$.

Theorem 1. *Suppose players' payoff functions u_1 and u_2 satisfy Assumptions 1 and 2, then there exists a cutoff discount factor $\underline{\delta}(u_1, u_2) \in (0, 1)$.*[5] *For every $K \in \mathbb{N}$, there exists an upper bound on the prior probability of commitment type $\overline{\pi}_0 > 0$, such that for every $\pi_0 < \overline{\pi}_0$ and $\delta > \underline{\delta}(u_1, u_2)$, there exists an equilibrium in which player 1's discounted average payoff equals his minmax payoff $u_1(a', b')$.*

According to Theorem 1, there are equilibria in which the patient seller receives his minmax payoff when the prior probability of commitment type is below some cutoff, each consumer observes a limited number of the seller's actions, and can observe some or all previous consumers' choices.

The existence of low-payoff equilibria stands in contrast to the reputation result in Fudenberg and Levine (1989), which shows that the patient seller receives at least his Stackelberg payoff in *all* equilibria when every consumer observes the entire history of the seller's actions (i.e., $K = +\infty$). This applies regardless of π_0 and regardless of how many predecessors' actions each consumer can observe.

Consumers' ability to observe their predecessors' choices (i.e., $M \geq 1$) is also needed for my result. Proposition 4 in Section 3.4 shows that in the product choice game where players' actions are strategic complements, i.e., $0 < c_T < c_N$, the patient seller receives at least his Stackelberg payoff in *all* equilibria when $K = 1$ and $M = 0$. The comparison between this case and the case where $M \geq 1$ suggests that consumers' ability to observe other consumers' choices can hurt the seller.

The proof is in Appendix A. In what follows, I construct a class of equilibria in the product choice game where the seller's payoff equals his minmax payoff 0. I call them *imitation equilibria*.

Proof of Theorem 1 in the Product Choice Game: For any $q \in (0, \frac{1}{2}]$, I construct an equilibrium when $\pi_0 \leq \left(\frac{q}{2}\right)^K \left(\frac{q}{2-q}\right) \equiv \overline{\pi}_0$ and $\delta \geq \max\{\frac{c_T}{c_T+1}, \frac{c_N}{c_N+1}\}$ such that: (1) Player 2_t's action depends only on (a_{t-1}, b_{t-1}), which takes five values $\varnothing$ (i.e., $t = 0$), (H, T), (H, N), (L, T), or (L, N); (2) Player 1's action in period t depends only on (a_{t-1}, b_{t-1}) and his *reputation* π_t, which is the probability player 2_t's belief assigns to the commitment type after observing $(a_{\max\{0,t-K\}}, ..., a_{t-1})$ and $(b_{\max\{0,t-M\}}, ..., b_{t-1})$.

[5] Player 1's discount factor needs to be above some cutoff $\underline{\delta}$, which ensures that the strategic type has an incentive to play the Stackelberg action although doing so gives him a lower stage-game payoff. In the product choice game, $\underline{\delta}(u_1, u_2) = \max\{\frac{c_T}{c_T+1}, \frac{c_N}{c_N+1}\}$. In Appendix A, I discuss how large $\underline{\delta}$ needs to be in monotone-supermodular games.

1. When $t = 0$ or $(a_{t-1}, b_{t-1}) = (L, N)$, player 2_t plays N and the strategic-type player 1 plays H with probability p_t such that $\pi_t + (1 - \pi_t)p_t = q$.

2. When $(a_{t-1}, b_{t-1}) = (H, N)$, player 2_t plays T with probability $r_1 \equiv \frac{1-\delta}{\delta} c_N$ and the strategic-type player 1 plays H with probability p_t such that $\pi_t + (1 - \pi_t)p_t = 1/2$. In the last step of this proof, I verify that $\pi_t \leq q/2$, which implies that $p_t \in [q/2, 1]$.

3. When $(a_{t-1}, b_{t-1}) = (L, T)$, player 2_t plays T with probability $r_2 \equiv 1 - \frac{1-\delta}{\delta} c_T$ and the strategic-type player 1 plays H with probability p_t such that $\pi_t + (1 - \pi_t)p_t = 1/2$.

4. When $(a_{t-1}, b_{t-1}) = (H, T)$, player 2_t plays T and player 1 plays H.

Player 1's continuation value depends only on (a_{t-1}, b_{t-1}), which is denoted by $V(a_{t-1}, b_{t-1})$. One can verify that $V(H, T) = 1$, $V(L, N) = V(\varnothing) = 0$, $V(H, N) = \frac{1-\delta}{\delta} c_N$, and $V(L, T) = 1 - \frac{1-\delta}{\delta} c_T$.

Next, I verify players' incentive constraints at every (a_{t-1}, b_{t-1}). It is straightforward to check that player 2_t best replies to player 1's action at every (a_{t-1}, b_{t-1}). For player 1's incentives,

1. When $t = 0$ or $(a_{t-1}, b_{t-1}) = (L, N)$, player 1's discounted average payoff from playing L is 0 and his discounted average payoff from playing H is $(1-\delta)(-c_N)+\delta V(H, N) = 0 = V(L, N) = V(\varnothing)$.

2. When $(a_{t-1}, b_{t-1}) = (H, N)$, player 1's discounted average payoff from playing L is

$$(1 - \delta)u_1(L, r_1 T + (1 - r_1)N) + \delta\{r_1 V(L, T) + (1 - r_1)V(L, N)\} = \frac{1 - \delta}{\delta} c_N = V(H, N),$$

and his discounted average payoff from playing H is

$$(1 - \delta)u_1(H, r_1 T + (1 - r_1)N) + \delta\{r_1 V(H, T) + (1 - r_1)V(H, N)\} = \frac{1 - \delta}{\delta} c_N = V(H, N).$$

3. When $(a_{t-1}, b_{t-1}) = (L, T)$, player 1's discounted average payoff from playing L is

$$(1 - \delta)u_1(L, r_2 T + (1 - r_2)N) + \delta\{r_2 V(L, T) + (1 - r_2)V(L, N)\} = 1 - \frac{1 - \delta}{\delta} c_T = V(L, T),$$

and his discounted average payoff from playing H is

$$(1 - \delta)u_1(H, r_2 T + (1 - r_2)N) + \delta\{r_2 V(H, T) + (1 - r_2)V(H, N)\} = 1 - \frac{1 - \delta}{\delta} c_T = V(L, T).$$

4. When $(a_{t-1}, b_{t-1}) = (H, T)$, player 1's discounted average payoff from playing H is 1, and his discounted average payoff from playing L is no more than 1.

In the last step, I show that at every history where $(a_{t-1}, b_{t-1}) \neq (H, T)$, there exists $p_t \in [0, 1]$ that satisfies the requirement in my construction. For this purpose, I only need to show that $\pi_t \leq q/2$ whenever $(a_{t-1}, b_{t-1}) \neq (H, T)$, since this implies that $p_t \in [q/2, 1]$ whenever $(a_{t-1}, b_{t-1}) \neq (H, T)$.

If $a_{t-1} = L$, then player 2_t's belief attaches zero probability to the commitment type. If $(a_{t-1}, b_{t-1}) = (H, N)$, then player 2_t's belief attaches positive probability to the commitment type only when $(a_{\max\{0,t-K\}}, ..., a_{t-1}) = (H, ..., H)$ and $(b_{\max\{0,t-M\}}, ..., b_{t-1}) = (N, N, ..., N)$. This is because conditional on player 1 being the commitment type (i.e., he plays H in every period), player 2_t plays T when $b_{t-1} = T$. I use π_t^* to denote the probability player 2_t's belief assigns to the commitment type after observing $(a_{\max\{0,t-K\}}, ..., a_{t-1}) = (H, ..., H)$ and $(b_{\max\{0,t-M\}}, ..., b_{t-1}) = (N, N, ..., N)$.

I show that $\pi_t^* \leq q/2$ by induction on $t \in \mathbb{N}$. First, $\pi_0^* = \pi_0 \leq q/2$ since $\pi_0 \leq \left(\frac{q}{2}\right)^K \left(\frac{q}{2-q}\right)$. Suppose $\pi_s^* \leq q/2$ for every $s \leq t-1$. The induction hypothesis implies that in every period before t, the probability that the strategic type plays H is at least $q/2$. Let $P^{\omega_s}(\cdot)$ be the probability measure induced by the equilibrium strategy of the strategic type. Let $P^{\omega_c}(\cdot)$ be the probability measure induced by the commitment type. Let E_t be the event that $(a_{\max\{0,t-K\}}, ..., a_{t-1}) = (H, ..., H)$. Let F_t be the event that $(b_{\max\{0,t-M\}}, ..., b_{t-1}) = (N, ..., N)$. According to Bayes rule,

$$\frac{\pi_t^*}{1-\pi_t^*} \Big/ \frac{\pi_0}{1-\pi_0} = \frac{P^{\omega_c}(E_t \cap F_t)}{P^{\omega_s}(E_t \cap F_t)} = \frac{P^{\omega_c}(E_t)}{P^{\omega_s}(E_t)} \cdot \frac{P^{\omega_c}(F_t|E_t)}{P^{\omega_s}(F_t|E_t)}. \tag{3.1}$$

Since the strategic type plays H with probability at least $q/2$ in every period before t, and N occurs with lower probability under the strategy of type ω_c compared to that under type ω_s, we have

$$\frac{P^{\omega_c}(E_t)}{P^{\omega_s}(E_t)} \leq (q/2)^{-K} \quad \text{and} \quad \frac{P^{\omega_c}(F_t|E_t)}{P^{\omega_s}(F_t|E_t)} \leq 1. \tag{3.2}$$

When $\pi_0 \leq \left(\frac{q}{2}\right)^K \left(\frac{q}{2-q}\right)$, (3.1) and (3.2) imply that $\pi_t^* \leq q/2$. This verifies that the constructed strategies are feasible and incentive compatible, under which player 1's payoff is 0. □

According to the consumers' strategy in imitation equilibria, consumer-t plays N with probability 1 or close to 1 when $b_{t-1} = N$, and plays T with probability 1 or close to 1 when $b_{t-1} = T$. Hence, imitation equilibria describe situations where *the first consumer does not trust the seller* and *every subsequent consumer imitates her predecessor with high probability.*

I explain the mechanism behind Theorem 1 using the product choice game. Intuitively, allowing consumers to observe other consumers' choices has two effects. First, it provides consumers information about the seller's type, which may help the seller to build reputations. Second, it *enables consumers to imitate their predecessors*, by which I mean consumer-t using strategy $b_t = b_{t-1}$ for every $t \geq 1$. This hurts the seller's reputational incentives since consumers' imitation reduces the impact of the seller's action on future consumers' actions. Consumers' imitation can lead to reputation failures when the first consumer does not trust the seller, since the seller receives a low stage-game payoff in the first period and imitation makes consumers' actions persistent. Imitation is *feasible* as long as $M \geq 1$. I argue that consumers' imitation can be *rationalized* when K is finite no matter how large M is.

As a benchmark, imitation is *not* incentive compatible when every consumer observes the seller's entire history. This is because for every $t \in \mathbb{N}$, either consumer-t believes that $a_t = H$ with probability more than $1/2$, in which case she has a strict incentive to play T, or the probability consumer $t+1$ attaches to the commitment type after she observes $a_t = H$ is at least two times the probability consumer-t attaches to the commitment type. When the seller plays H in every period, there can be at most a finite number of consumers who have incentives to imitate a predecessor who played N.

In contrast, consumers' imitation behaviors *can* be rationalized when each consumer observes *a bounded number of the seller's actions* and the probability of commitment type is below some cutoff:

1. Consumers may not be convinced that H will be played in the future after observing H in at most K periods. This is because even when consumer-t believes that H will be played with probability less than $1/2$, consumer $t+1$'s posterior belief may not be greater than consumer-t's since she cannot observe a_{t-K}. When the seller plays H in every period, consumers after period K obtain the same information from the seller's actions. Unlike the canonical reputation models, there can be infinitely many consumers who are concerned that the seller is likely to be strategic and will play L in the future. This is reflected in the first part of (3.2), that $\frac{P^{\omega_c}(E_t)}{P^{\omega_s}(E_t)} \leq (q/2)^{-K}$.

2. Although consumers can learn from other consumers' choices, the additional information each consumer obtained from these choices *never discourages her from imitating her immediate predecessor*. In particular, when a consumer's immediate predecessor played N, observing other consumers' choices can only lower the seller's reputation, which encourages this consumer to imitate by playing N as well.[6] This is reflected in the second part of (3.2), that $\frac{P^{\omega_c}(F_t|E_t)}{P^{\omega_s}(F_t|E_t)} \leq 1$.

One technical subtlety is that the seller's payoff is 0 even when $\delta \to 1$, which suggests that an arbitrarily patient seller's payoff is sensitive to the first consumer's action. For intuition, let us take $K = 1$. When the strategic seller exerts effort, he sacrifices his current-period payoff in exchange for a higher continuation value, so he has a stronger incentive to do so when he is more patient or when consumers are less likely to imitate—since the seller can boost his continuation value only when the

[6] If each of the last M consumers played N, then the seller's reputation does not increase after the current consumer observes these M consumers' choices. If at least one of the last M consumers played T but the most recent one played N, then the current consumer can rule out the commitment type after observing these M consumers' choices.

next consumer does not imitate. Hence, it is harder for consumers to distinguish between the two types when the strategic seller is more patient. In imitation equilibria, this logic is reflected by the observation that consumers cannot distinguish between the two types when the seller's action affects the next consumer's action with probability more than $\mathcal{O}(1-\delta)$, or equivalently, when consumers imitate with probability less than $1-\mathcal{O}(1-\delta)$. Hence, the maximal probability of imitation increases with δ, so in the worst case scenario, it takes more time for a more patient seller to obtain consumers' trust. The prolonged process of trust building cancels out the positive effects of being more patient.

The key features of imitation equilibria are: *consumers not trusting newly arrived sellers who have no past record* and *consumers imitating their predecessors*, the plausibility of both are supported by empirical evidence. On consumers' imitation, Cai, Chen and Fang (2009) find that consumers imitate each other in the Chinese food market. Zhang (2010) finds that patients are more likely to refuse a kidney that has been refused by earlier patients, even conditional on the objective quality of kidneys. Cai, De Janvry and Sadoulet (2015) find that farmers in rural China are more likely to purchase weather insurance when they were told that other farmers have purchased the insurance.[7]

My *no initial trust* condition fits some informal markets in developing countries. For example, Michelson et al. (2021) find that many farmers in Tanzania suspect that the fertilizers sold in local markets are adulterated and their pessimistic beliefs about the seller's integrity persists over time. Such persistent mistrust contributes to the under-adoption of fertilizers.

Although details about farmers' information structures are not available, my result suggests a plausible explanation for the persistent mistrust between farmers and sellers. In terms of the fitness of my model, first, farmers' payoffs depend on the seller's action, namely, whether the seller has adulterated products currently sold on the market. Second, farmers are myopic, that is, they won't trust the seller if they believe that his products are adulterated and won't punish the seller if they believe that his products are authentic. Although some farmers may buy multiple times, they are unlikely to sacrifice their current-period profits since most of them have low income and may not afford to do so. Third, I require that every farmer observes the choice of her predecessor and a limited number of the seller's actions. This is plausible when farmers live close to each other—so that it is easy to observe other farmers' recent choices, and farmers have limited memories about the seller's actions. My result suggests a rationale for persistent mistrust when farmers do not trust the seller in the beginning due to a pessimistic prior, and are unwilling to trust the seller even after they observe him supplying authentic products in a bounded number of periods.

[7]Cai, De Janvry and Sadoulet (2015) write on page 82 that "*...when we told farmers about other villagers decisions, these decisions strongly influenced their own take-up choices...*", and "*...if information on other villagers decisions can be revealed in complement to the performance of the network, it can have a large impact on adoption decisions...*"

In what follows, I address several issues related to Theorem 1. Section 3.1 introduces a refinement under which the patient player receives his minmax payoff in *all* equilibria that satisfy this refinement. Sections 3.2 and 3.3 explain the connections between Theorem 1 and existing results on social learning and reputation formation. Section 3.4 uses an example to explain why consumers' ability to observe previous consumers' choices is not redundant for Theorem 1.

3.1 Equilibrium Refinement that Selects Low-Payoff Equilibria

Theorem 1 shows that the seller receives his minmax payoff in *some* equilibria. In this section, I introduce a refinement such that the seller receives his minmax payoff in *all* equilibria that satisfy this refinement. Let h_2^t be player 2_t's history. Let $A' \equiv \{a \in A \text{ s.t. } b' = \text{BR}_2(a)\}$. By definition of b', player 1's lowest action a' belongs to A'. Assumptions 1 and 2 imply that A' consists of all of player 1's actions that are below some cutoff and that player 1's Stackelberg action a^* does not belong to A'.

Proposition 1. *Suppose* (u_1, u_2) *satisfies Assumptions 1 and 2. For every* $(\delta, \pi_0) \in (0,1)^2$*, player* 1*'s payoff equals his minmax payoff* $u_1(a', b')$ *in every PBE that satisfies the following refinement:*

1. ***Punishment following bad outcome:*** *For every* $t \geq 1$, $\sigma_2(h_2^t) = b'$ *when* $b_{t-1} = b'$ *and* $a_{t-1} \in A'$.

2. ***No initial trust:*** *Player* 2 *plays* b' *in period* 0.

The proof is in Appendix B. In the product choice game, my refinement requires that (1) consumer-t does not trust the seller when the worst action profile (L, N) occurred in period $t-1$, and (2) consumers do not trust a seller who newly arrives and has no past record.

Among the two conditions in my refinement, *punishment following bad outcome* is rather standard. It is satisfied not only by imitation equilibria but is also satisfied by some equilibria where the seller receives a high payoff, such as equilibria where consumers use grim-trigger strategies.

My *no initial trust* condition is satisfied by imitation equilibria but is violated by grim-trigger equilibria, since grim-trigger requires consumers *do* trust the seller in the first period. The extent to which consumers trust newly arrived sellers depends on the application. While there are situations where grim-trigger fits better, there are also situations where most consumers *do not* trust newly arrived sellers or trust newly arrived sellers with low probability, as suggested by the evidence I cited earlier. Hence, *no initial trust* or more generally *low initial trust* is plausible.

In the online appendix, I examine the robustness of my refinement result. I show that in every equilibrium that satisfies *punishment following bad outcome*, player 1's discounted average payoff is no more than $u_1(a', b')$ conditional on b' being played in period 0, and his discounted average payoff cannot significantly exceed $u_1(a^*, b^*)$ when δ is close to 1 and π_0 is close to 0. Hence, there exists a real-valued function $\gamma(\pi_0, \delta) \in [0, 1]$ that converges to 0 when δ is close to 1 and π_0 is close to 0, such that for every $\varepsilon \in [0, 1]$ and every equilibrium that satisfies *punishment following bad outcome* and player 2 playing b' with probability $1 - \varepsilon$ in period 0, player 1's equilibrium payoff is no more than

$$(1-\varepsilon)u_1(a', b') + \varepsilon\Big\{\gamma(\pi_0, \delta) \max_{(a,b)\in A\times B} u_1(a, b) + (1 - \gamma(\pi_0, \delta))u_1(a^*, b^*)\Big\}.^8 \tag{3.3}$$

In the product choice game, the payoff upper bound in (3.3) implies that in every equilibrium that satisfies *punishment following bad outcome*, (1) the seller's payoff must be close to his minmax payoff if he is trusted with probability close to 0 in the first period, and (2) when the seller is patient and the prior probability of commitment type is below some cutoff, he may receive his Stackelberg payoff only if he is trusted with probability close to one in the first period.

3.2 Connections with Canonical Social Learning Models

The imitation equilibria constructed in the proof of Theorem 1 are reminiscent of the canonical results on social learning. In Banerjee (1992), Bichandarni, Hirshleifer and Welsh (1992), and Smith and Sørensen (2000), a sequence of myopic players chooses their actions sequentially after observing all predecessors' actions and a private signal of an exogenous state. Inefficiencies take the form of *herding* in the sense that myopic players ignore their private signals and imitate their immediate predecessors.

My model is analogous once we view $(a_{\max\{0,t-K\}}, ..., a_{t-1})$ as player 2_t's private signal. The conceptual difference is that in my model, the myopic players' payoffs depend only on the patient player's endogenous actions but not on the patient player's type. The myopic players never herd on action N in imitation equilibria since their actions are responsive to the seller's action in the period before. Proposition 2 shows that the *no bad herd* conclusion applies more generally. Formally, I say that player 2s herd on action b at $h^t \equiv (a_s, b_s)_{s \leq t-1}$ if player 2s play b at h^t and at every successor of h^t. Let $\pi(h^t) \in [0, 1]$ be the probability player 2's belief at h^t assigns to the commitment type.

Proposition 2. *Suppose players' payoffs satisfy Assumption 1, then for every $(\delta, \pi_0) \in (0, 1)^2$, every $b \neq b^*$, and every equilibrium, player 2s cannot herd on b at any history h^t with $\pi(h^t) > 0$.*

The proof is in Appendix C. Proposition 2 implies that as long as player 1 imitates the commitment

[8] When π_0 is below some cutoff and δ is arbitrarily close to 1, there exist equilibria that satisfy *punishment following bad outcome* and ε *initial trust* where player 1's payoff is arbitrarily close to (3.3). This implies that my payoff upper bound is tight in the sense that it can be approximately attained by some equilibria.

type, player 2s can never herd on any action that does not best reply to a^* regardless of player 1's discount factor, player 2's prior belief, and the equilibrium we focus on. This implies that reputation failure cannot be caused by myopic players herding on actions that give the patient player a low payoff.

For a heuristic explanation, once player 2s herd on action $b \neq b^*$, the strategic-type player 1 cannot affect player 2s' future actions, so he has no intertemporal incentive. As a result, the strategic-type player 1 will not play a^* when a^* is not a best reply to b in the stage game.[9] This implies that player 2 will learn that player 1 is the commitment type upon observing a^*, and hence, will have a strict incentive to play b^*. This contradicts the presumption that player 2s herd on action b, that is not b^*.

3.3 Connections with Canonical Reputation Models

Fudenberg and Levine (1992) show that a patient player can secure his Stackelberg payoff in all equilibria if (1) with positive probability, he is a commitment type who plays his Stackelberg action in every period, and (2) every short-run player can observe the entire history of some noisy signal that can statistically identify the patient player's action. An elegant proof of their result is provided by Gossner (2011). The key is to show that for any $\delta \in (0,1)$ and any Bayes Nash equilibrium (σ_1, σ_2),

$$\mathbb{E}^{(a^*,\sigma_2)}\Big[\sum_{t=0}^{\infty} d\Big(y_t(\cdot|a^*)\Big|\Big|y_t(\cdot)\Big)\Big] \leq -\log \pi_0, \tag{3.4}$$

where $y_t(\cdot)$ is the equilibrium distribution of player 2's signals about a_t, $y_t(\cdot|a^*)$ is the distribution of player 2's signals about a_t conditional on player 1 being the commitment type, $d(\cdot||\cdot)$ is the Kullback-Leibler divergence between two distributions, and $\mathbb{E}^{(a^*,\sigma_2)}[\cdot]$ is the expectation operator when player 1 plays a^* in every period and player 2 plays σ_2. When player 2's signals can identify player 1's actions, $d\big(y_t(\cdot|a^*)\big|\big|y_t(\cdot)\big)$ is bounded away from 0 whenever player 2_t does not have a strict incentive to play b^*. Inequality (3.4) implies that in expectation, there can be at most a bounded number of periods in which player 2s do not have strict incentives to play b^*. Importantly, this upper bound does not depend on δ. This explains why player 1's equilibrium payoff is at least $u_1(a^*, b^*)$ when $\delta \to 1$.

Fudenberg and Levine (1992)'s model is analogous to mine when $M = +\infty$, i.e., every consumer can observe the entire history of her predecessors' actions. This is because each consumer's action can be viewed as an informative signal about the seller's past actions, so observing the entire history of consumers' choices can be viewed as observing the entire history of some noisy signal about the seller's actions. Inequality (3.4) applies to imitation equilibria of my model once we take $y_t(\cdot)$ as the

[9] In the case where a^* is player 1's myopic best reply to b, both types of player 1 play a^* in equilibrium after player 2s herd on action b. When both types of player 1 play a^*, player 2 has a strict incentive to play b^*, which is not b.

equilibrium distribution of b_{t+1} and $y_t(\cdot|a^*)$ as the distribution of b_{t+1} conditional on player 1 being the commitment type. Consumer-$t+1$'s action can statistically identify the seller's action in period t, so $d\big(y_t(\cdot|a^*)\big|\big|y_t(\cdot)\big) > 0$ when player 2_t does not have a strict incentive to play b^*.

However, the distribution of b_{t+1} in imitation equilibria is such that $d\big(y_t(\cdot|a^*)\big|\big|y_t(\cdot)\big) \to 0$ as $\delta \to 1$. This stands in contrast to Fudenberg and Levine (1992)'s model in which $d\big(y_t(\cdot|a^*)\big|\big|y_t(\cdot)\big)$ is bounded away from zero whenever player 2_t does not have a strict incentive to play b^*. As a result, inequality (3.4) cannot rule out situations where the expected number of periods in which player 2 has no incentive to play b^* grows without bound as $\delta \to 1$. This is indeed the case in imitation equilibria, where the prolonged process of reputation building cancels out the positive effects of increased patience.

The above discussion unveils an interesting feature of imitation equilibria: Although the patient player can eventually guarantee a high continuation value by exerting high effort in every period, his *discounted average payoff* equals his minmax payoff. Intuitively, each player 2's action is informative about her observations of player 1's past actions, and every player 2 can observe the entire history of player 2s' actions. As a result, either player 2_t strictly prefers to play b^*, or all future player 2s learn something about player 1's type from b_t. The arguments in Gossner (2011) imply that there exist at most a finite number of periods where player 2 does not have a strict incentive to play b^*. Therefore, the patient player 1 can eventually secure a high continuation value by playing a^* in every period. This logic generalizes to all equilibria when every consumer can observe all of her predecessors' choices.

Proposition 3. *Suppose $M = +\infty$ and players' payoffs satisfy Assumptions 1 and 2.*[10] *For every $(\delta, \pi_0) \in (0,1)^2$ and every strategy profile (σ_1, σ_2) that is part of a PBE, we have*

$$\liminf_{t\to\infty} \frac{1}{t}\mathbb{E}^{(a^*,\sigma_2)}\Big[\sum_{s=0}^{t-1} u_1(a_s,b_s)\Big] \geq \frac{K}{K+1}u_1(a^*,b^*) + \frac{1}{K+1}u_1(a^*,b'). \tag{3.5}$$

When π_0 is small and δ is large, there exists an equilibrium such that (3.5) holds with equality.

According to Proposition 3, the informed player's undiscounted average payoff from playing the Stackelberg action is at least a fraction $\frac{K}{K+1}$ of his Stackelberg payoff plus a fraction $\frac{1}{K+1}$ of some low payoff $u_1(a^*, b')$. This is true for all equilibria, all discount factors, and all prior beliefs. This lower bound is tight in the sense that it can be attained by some equilibria when π_0 is small and δ is large.

When the right-hand-side of (3.5) is strictly greater than $u_1(a', b')$, the patient player 1 can guarantee an asymptotic payoff that is strictly greater than his minmax payoff by playing a^* in every period. The only way to reconcile this conclusion and Theorem 1 is that when player 1 plays a^* in

[10] I show in the online appendix that Proposition 3 is not true when M is finite, in the sense that there exist equilibria where player 1's undiscounted average payoff from imitating the commitment type equals his minmax payoff $u_1(a', b')$.

every period, it takes more time for him to secure this high asymptotic payoff when δ is larger. It is the prolonged process of reputation building that cancels out the direct effects of increased δ.

The proof of Proposition 3 is in Appendix D. For a heuristic explanation, Assumption 2 implies that a^* is suboptimal for player 1 in the stage game. Therefore, for every $t \in \mathbb{N}$, either the strategic type has no incentive to play a^* in period t, or $(b_{t+1}, ..., b_{t+K})$ is informative about a_t. In the first case, players 2_{t+1} to 2_{t+K} learn that player 1 is committed after observing $a_t = a^*$. By playing a^* in every period, player 1's average payoff from period t to $t+K$ is at least a fraction $\frac{K}{K+1}$ of his Stackelberg payoff plus $\frac{1}{K+1}$ times his minimal stage-game payoff. In the second case, all future player 2s observe an informative signal about a_t given that $M = +\infty$. According to the arguments in Fudenberg and Levine (1992) and Gossner (2011), player 2s' posterior beliefs attach probability close to 1 to the commitment type after a finite number of periods with learning. The two parts together imply that player 1's asymptotic payoff is no less than the right-hand-side of (3.5).

3.4 Detrimental Effects of Consumers' Observational Learning

I focus on the product choice game where $0 < c_T < c_N$, i.e., players' actions are strategic complements. I show that the patient seller's payoff is arbitrarily close to his Stackelberg payoff 1 in *all* equilibria when every consumer can only observe the seller's action in the period before but cannot observe any other consumer's action, i.e., $(K, M) = (1, 0)$.[11] This stands in contrast to Theorem 1, which shows that the seller receives his minmax payoff 0 in some equilibria when every consumer can also observe her immediate predecessor's action (i.e., $M \geq 1$). The comparison between these two conclusions suggests that consumers' ability to observe other consumers' choices can hurt the patient seller.

Since $M = 0$, consumers may not know calendar time t. Each consumer has a prior belief about t, observes the seller's action in the period before, and updates her belief about t using Bayes Rule. For example, if a consumer observes $a_{t-1} = \varnothing$, then she knows that $t = 0$; if she observes $a_{t-1} = H$ or $a_{t-1} = L$, then her posterior belief about t is not degenerate. I provide an interpretation of the seller's discount factor δ in order to make consumers' prior belief about calendar time well-defined.

1. Let $\delta_1 \in (0, 1)$ be the seller's *survival rate*, namely, the seller survives in the next period with probability δ_1, and exits the market with probability $1 - \delta_1$ after which the game ends.

2. Let $\delta_2 \in (0, 1)$ be the seller's *time preference*, namely, the seller is indifferent between one unit of payoff in period t and δ_2 unit of payoff in period $t - 1$.

[11] Liu and Skrzypacz (2014) study the case where the seller's cost is greater when consumers trust him, i.e., $0 < c_N < c_T$. Section 4 of Sperisen (2018) studies the case where $c_N = c_T$.

By definition, $\delta = \delta_1\delta_2$. Under this interpretation of the seller's discount factor, consumers' prior belief attaches probability $(1-\delta_1)\delta_1^t$ to the calendar time being t.[12]

Proposition 4. *Suppose in the product choice game, $0 < c_T < c_N$ and $(K, M) = (1, 0)$. For every $\pi_0 > 0$, there exists $\underline{\delta} \in (0, 1)$, such that when $\delta > \underline{\delta}$, the seller's payoff is at least $\delta - (1-\delta)c_N$ in every PBE.*

The proof is in Appendix E. In the online appendix, I generalize this result to a class of games where players' actions are strategic complements. Intuitively, $M = 0$ implies that b_t only affects the seller's stage-game payoff in period t, $K = 1$ implies that the seller has an incentive to play $a_{t-1} = H$ only if it increases consumer-t's probability of playing T, and $c_N > c_T$ implies that the seller has a stronger incentive to play $a_t = H$ if consumer-t plays T with higher probability. I consider two cases:

1. When playing H does not increase the probability of being trusted in the next period, the strategic seller has no incentive to play H. This implies that consumers will be convinced that the seller is the commitment type after observing action H, and will have a strict incentive to play T. Hence, the strategic seller can secure his Stackelberg payoff by playing H in every period.

2. When playing H increases the probability of being trusted in the next period, I show in Appendix E that when δ_1 is close to 1, either consumer-t has a strict incentive to play T when $a_{t-1} = H$, or the seller has an incentive to play H in period t when $a_{t-1} = L$.[13] The seller obtains at least his Stackelberg payoff in the first case. In the second case, since the seller's stage-game payoff function is strictly supermodular and consumer-t plays T with higher probability when $a_{t-1} = H$, the seller must have a strict incentive to play H when $a_{t-1} = H$ as long as he has a weak incentive to play H when $a_{t-1} = L$. If the strategic-type seller has a strict incentive to play H when $a_{t-1} = H$, then consumer-t has a strict incentive to play T when $a_{t-1} = H$.

The two cases together imply that the patient seller receives payoff at least 1 in all equilibria.

In contrast, allowing consumers to observe their predecessors' choices leads to a richer set of feasible strategies for the consumers and as a result, a larger set of equilibria. Allowing both players to observe previous consumers' choices also weakens the implication of supermodular stage-game payoffs, since players can coordination their continuation plays on these commonly observed actions. Hence, it is not

[12] When $\delta_1 = 1$, consumers have an improper uniform prior belief about calendar time. When $\delta_1 < 1$, consumers' prior belief about calendar time is well-defined. Fixing δ, the values of δ_1 and δ_2 do not affect Theorem 1.

[13] Intuitively, when the strategic seller has no incentive to play H in period t when $a_{t-1} = L$, he will never play H if he has played L before. When δ_1 is close to 1, consumers' beliefs attach probability close to 1 to the calendar time t being large. If $a_{t-1} = H$ for some large enough t, then either the seller is the commitment type, or he is the strategic type who plays H in the long run. In both cases, consumers have strict incentives to play T after observing action H.

necessarily the case that the seller has a stronger incentive to exert high effort when consumers trust him with higher probability in the current period. This leads to equilibria where the seller receives a low payoff, such as the imitation equilibria constructed in the proof of Theorem 1.

4 Extension: Reputation with Contemporaneous Information

Motivated by the social learning models of Banerjee (1992), Bichandarni, Hirshleifer and Welsh (1992), and Smith and Sørensen (2000), I study an extension where each player 2 observes player 1's actions in the last K periods, the entire history of player 2s' actions (i.e., $M = +\infty$), and a private signal s_t about player 1's current-period action a_t. Whether player 1 can observe s_t is irrelevant for my results. Let $s_t \in S$ where S is a countable set. Let $f(s_t|a_t)$ be the probability of s_t when player 1's action is a_t. I restrict attention to signal distributions that satisfy a *monotone likelihood ratio property* (MLRP).

MLRP. *The distribution of player 2's private signal satisfies MLRP if there exists a complete order on* S, $\succ_S$, *such that* $f(s|a)f(s'|a') \geq f(s'|a)f(s|a')$ *for every* $a \succ_A a'$ *and* $s \succ_S s'$.

I replace $\succ_A$, $\succ_B$, and $\succ_S$ with $\succ$ in order to simplify notation. Whether player 1 can guarantee his Stackelberg payoff in all equilibria depends on whether player 2's private signal is unboundedly informative about player 1's Stackelberg action a^*.

Unbounded Informativeness. *Player 2's private signal is unboundedly informative about* a^* *if for every* $L > 0$, *there exists* $s \in S$, *such that* $f(s|a^*) > Lf(s|a)$ *for every* $a \neq a^*$.

My notion of unbounded informativeness is similar to that in Smith and Sørensen (2000).[14] When S is a finite set, unbounded informativeness requires the existence of $s^* \in S$ such that $f(s^*|a) > 0$ if and only if $a = a^*$. When S is countably infinite, $f(\cdot|a)$ can have full support for every $a \in A$, as long as there exists a sequence $\{s_n\}_{n \in \mathbb{N}} \subset S$ such that $\lim_{n \to +\infty} \frac{f(s_n|a^*)}{f(s_n|a)} = +\infty$ for every $a \neq a^*$.

For an interpretation of s_t, consider a regulator who only has the budget to inspect an ε fraction of sellers in every period and can issue certificates to the ones that are being inspected. The certificate can be modeled as the private signal s_t when the current-period consumer can notice the certificate before deciding what to buy. MLRP implies that the seller is more likely to obtain a good certificate when he exerts higher effort. If S is a finite set, then consumers' private signal is unboundedly informative about a^* when the seller can obtain a good certificate only if he plays a^*. This is the case, for example, when the certificate reveals the seller's action with probability $\varepsilon > 0$.

[14]First, when S is infinite, I allow for, but does not require, signal realizations that can perfectly rule out some of player 1's actions, while Smith and Sørensen (2000) require the signal distribution to have full support conditional on every state. Second, I restrict attention to S that is countable while Smith and Sørensen (2000) allow S to be uncountable.

Theorem 2. *Suppose players' payoffs satisfy Assumptions 1 and 2, every player* 2 *can observe all previous player 2s' choices, player* 2*'s private signal satisfies MLRP, and is unboundedly informative about* a^**. Then for every prior belief* $\pi_0 > 0$ *and constant* $\varepsilon > 0$*, there exists* $\delta^* \in (0,1)$ *such that player* 1*'s payoff is at least* $u_1(a^*, b^*) - \varepsilon$ *in all equilibria when* $\delta > \delta^*$*.*

For every $a \neq a^*$, a^* is *not strongly separable* from a if there exists $\varepsilon > 0$ such that $f(s|a) \geq \varepsilon f(s|a^*)$ for every $s \in S$. If player 2's private signal is unboundedly informative about a^*, then there exists no $a \neq a^*$ such that a^* is not strongly separable from a. However, player 2's private signal not being unboundedly informative about a^* does not imply that a^* is not strongly separable from some $a \neq a^*$. Theorem 3 shows a partial converse of Theorem 2 when player 2s' private signals cannot strongly separate the Stackelberg action a^* from the lowest action a'.

Theorem 3. *Suppose players' payoffs satisfy Assumptions 1 and 2, every player* 2 *can observe all previous player 2s' choices, player 2's private signal satisfies MLRP, and* a^* *is not strongly separable from* a'*. For every* $K \in \mathbb{N}$*, there exists* $\overline{\pi}_0 \in (0,1)$ *such that for every* $\pi_0 < \overline{\pi}_0$ *and* δ *large enough, there exists an equilibrium where player* 1*'s payoff is* $u_1(a', b')$*.*

The proofs are in Appendix F. Theorem 2 implies that the patient player can guarantee his S-tackelberg payoff in all equilibria when each of his opponents can observe the entire history of their predecessors' choices and an unboundedly informative private signal about the patient player's current-period action.[15] Theorem 3 extends the reputation failure result of Theorem 1 to situations where K is finite, M is infinite, and player 2_t observes a private signal about a_t before choosing b_t.

When $|A| = 2$, every signal distribution satisfies MLRP. Since Assumption 2 requires that $a^* \neq a'$, we have $A = \{a^*, a'\}$. The private signal is not unboundedly informative about a^* *if and only if* a^* is not strongly separable from a'. Hence, the private signal being unboundedly informative about a^* is both necessary and sufficient for player 1 to secure his Stackelberg payoff in all equilibria.[16]

The above conclusion is reminiscent of a well-known result in Bichandarni, Hirshleifer and Welsh (1992) and Smith and Sørensen (2000). They show that in canonical social learning models where there are two states, every myopic player has a finite number of actions, and all players share the same payoff function, the myopic players' actions are asymptotically efficient if and only if their private signals are unboundedly informative about the payoff-relevant state.[17]

---

[15] Theorem 2 only establishes a common property of all equilibria but does not establish the existence of equilibrium. When S is infinite, the existence of equilibrium does not follow from the canonical result of Fudenberg and Levine (1983). I provide a constructive proof for the existence of equilibrium in Appendix F.2.

[16] When $|A| \geq 3$, MLRP cannot be dropped and the condition in Theorem 3 cannot be replaced by "*the private signal not being unboundedly informative about* a^*", or "a^* *is not strongly separable from* $a^\dagger$ *for some* $a^\dagger \notin \{a^*, a'\}$".

[17] Lee (1993) shows that asymptotic efficiency can be achieved under boundedly informative signals when players have

My model differs from Smith and Sørensen (2000) since the myopic players' payoffs depend only on the action profile but do not depend on the persistent state—which is player 1's type in my model. My analysis focuses on the patient player's *discounted average payoff* instead of his *asymptotic payoff* or the *asymptotic outcome*. In fact, the myopic players asymptotically learning about the persistent state is *neither necessary nor sufficient* for the patient player to receive a high discounted average payoff. It is not necessary since player 2's payoff depends only on the action profile but not on player 1's type. For example, suppose player 2s believe that the strategic-type player 1 plays a^* in every period, they cannot learn about player 1's type but player 1 can receive his Stackelberg payoff $u_1(a^*, b^*)$ by playing a^* in every period. It is not sufficient since in imitation equilibria, player 1's asymptotic payoff is $u_1(a^*, b^*)$ but his discounted average payoff is $u_1(a', b')$ no matter how patient he is.

I sketch the proof of Theorem 2 in the case where S is finite, under which there exists $s^* \in S$ such that $f(s^*|a) > 0$ if and only if $a = a^*$.[18]

A rough intuition is that player 2_t observing an unboundedly informative private signal s_t about a_t guarantees a positive lower bound on the informativeness of player 2_t's action b_t about a_t. Unlike imitation equilibria where the informativeness of b_t about a_{t-1} vanishes to 0 as δ goes to 1, the informativeness of b_t about a_t is bounded away from zero for all $\delta \in (0, 1)$. Since every player 2 can observe all of her predecessors' actions, the arguments in Fudenberg and Levine (1992) and Gossner (2011) imply that the patient player receives at least his Stackelberg payoff in all equilibria.

A more detailed explanation proceeds in two steps, which highlights the role of unbounded informativeness and MLRP. First, I examine whether player 2_t's action is informative about her private signal s_t. Intuitively, b_t can be uninformative about s_t for two reasons: (1) player 2_t is unwilling to play b^* no matter which s_t she observes, and (2) player 2_t is willing to play b^* no matter which s_t she observes. Since s_t is unboundedly informative about a^*, player 2 has a strict incentive to play b^* when she observes $s_t = s^*$. This rules out the first possibility. When player 2_t is willing to play b^* no matter which s_t she observes, player 1's stage-game payoff is $u_1(a^*, b^*)$ when he plays a^* in period t.

Second, I examine whether player 2_t's action is informative about player 1's type. When player 1's action choice is binary, i.e., $A \equiv \{a^*, a'\}$, player 2_t is willing to play b^* if and only if $\frac{f(s_t|a^*)}{f(s_t|a')}$ is above some cutoff. This implies that $\Pr(b_t = b^*|a_t = a^*) - \Pr(b_t = b^*|a_t = a') \geq 0$. Since player 2_t plays b^*

a rich set of actions (e.g., a continuum). When the states, actions, and signals can be ordered such that players' payoffs satisfy single-crossing differences, Kartik, Lee and Rappoport (2021) show that asymptotic efficiency can be achieved as long as the signal distribution satisfies directionally unbounded beliefs, which is weaker than unbounded informativeness.

[18] When S is infinite and the signal is unboundedly informative about a^*, there exists a nonempty subset $S(\pi) \subset S$ for every $\pi \in (0, 1)$ such that when the prior probability of commitment type is at least π before player 2_t observes s_t, she has a strict incentive to play b^* after observing any $s_t \in S(\pi)$. See Lemma F.1 in Appendix F for details.

after observing s^* which occurs if and only if player 1 plays a^*, there exists $c > 0$ such that

$$\Pr(b_t = b^*|a_t = a^*) - \Pr(b_t = b^*|a_t = a') \geq c(1 - \Pr(b_t = b^*|a_t = a^*)), \tag{4.1}$$

i.e., the informativeness of b_t about a_t is bounded from below by some positive function of $1 - \Pr(b_t = b^*|a_t = a^*)$. For every $\nu \in (0, 1)$, when $\Pr(b_t = b^*|a_t = a^*) \leq 1 - \nu$, the strategic type plays a^* with probability bounded away from 1, so the informativeness of b_t about player 1's type is bounded from below by a strictly positive function of ν.

When player 1 has three or more actions, player 2_t's incentive to play b^* can no longer be summarized by a likelihood ratio. As a result, player 2_t's action can be uninformative about player 1's type even when the private signal is unboundedly informative about a^* and b_t is informative about s_t. I provide a counterexample in Section 4.1. Nevertheless, when the private signal satisfies MLRP, b_t is informative about player 1's type in every period where $\Pr(b_t = b^*|a_t = a^*) \neq 1$.

Formally, for every $\alpha \in \Delta(A)$ and $\beta : S \to \Delta(B)$, let $\gamma(\alpha, \beta) \in \Delta(B)$ be the distribution of b induced by (α, β). I show in Lemma F.2 of Appendix F that there exists $c > 0$ such that for every $\nu \in (0, 1)$, every $\alpha \in \Delta(A)$ such that a^* belongs to the support of α, and every β that best replies to α, if the probability of b^* under $\gamma(a^*, \beta)$ is less than $1 - \nu$, then the Kullback-Leibler divergence between $\gamma(\alpha, \beta)$ and $\gamma(a^*, \beta)$ is at least $c\nu^2$. This implies that when player 1 imitates the commitment type, either b^* occurs with probability at least $1 - \nu$ under (a^*, β), or the informativeness of b_t about player 1's type, measured by the Kullback-Leibler divergence between the distribution induced by the equilibrium strategy and the distribution induced by the commitment type, is bounded away from 0.

Back to the discussion on the connections between my results and the canonical reputation results in Section 3.3. Inequality (3.4) also applies to my model with contemporaneous private signals once we view $y_t(\cdot)$ as the equilibrium distribution of b_t and $y_t(\cdot|a^*)$ as the distribution of b_t conditional on player 1 being the commitment type. The above discussion implies that when player 2's private signal is unboundedly informative about a^* and satisfies MLRP, there exists a strictly increasing function $g : [0, 1] \to \mathbb{R}_+$ such that $g(0) = 0$ and $d\big(y_t(\cdot|a^*)\big|\big|y_t(\cdot)\big) > g(\nu)$ when player 2_t plays b^* with probability less than $1 - \nu$. Inequality (3.4) implies that for every $\nu \in (0, 1)$, the expected number of periods where $\Pr(b_t = b^*|a_t = a^*) < 1 - \nu$ is bounded from above and this upper bound depends only on ν and is independent of δ. Hence, for every $\nu > 0$, there exists $\underline{\delta} \in (0, 1)$ such that when $\delta > \underline{\delta}$, player 1 receives at least a fraction $1 - \nu$ of $u_1(a^*, b^*)$ when he plays a^* in every period.

4.1 Conditions in Theorems 2 and 3

Bounded Informativeness: I use an example to explain why "a^* is not strongly separable from a'" in Theorem 3 cannot be replaced by a weaker condition that "consumers' private signal s_t is not unboundedly informative above a^*". Suppose players' stage-game payoffs are

-	b^*	b'
$\overline{a}$	$1,4$	$-2,0$
a^*	$2,1$	$-1,0$
$\underline{a}$	$3,-2$	$0,0$

Let $S \equiv \{\overline{s}, s^*, \underline{s}\}$, with $f(\overline{s}|\overline{a}) = 2/3$, $f(s^*|\overline{a}) = 1/3$, $f(\overline{s}|a^*) = 1/3$, $f(s^*|a^*) = 2/3$, and $f(\underline{s}|\underline{a}) = 1$. One can verify that players' stage-game payoffs are monotone-supermodular when player 1's actions are ranked according to $\overline{a} \succ a^* \succ \underline{a}$, and player 2's actions are ranked according to $b^* \succ b'$. When signal realizations are ranked according to $\overline{s} \succ s^* \succ \underline{s}$, the signal distribution satisfies MLRP, and is not unboundedly informative about a^*. Player 1's payoff is at least 2 in every equilibrium. This is because when he plays a^*, player 2 observes either s^* or $\overline{s}$, and has a strict incentive to play b^*.

Not Strongly Separable from Other Actions: I use an example to explain why in Theorem 3, "a^* is not strongly separable from a'" cannot be replaced by "a^* is not strongly separable from $a^\dagger$ for some $a^\dagger \notin \{a^*, a'\}$". Suppose player 1's stage-game payoff is given by the following matrix:

-	b^*	$b^\dagger$	b''	b'
a^*	5	-2	-3	-4
$a^\dagger$	6	-1	-2	-3
a''	7	2	1	-1
a'	8	3	2	0

Player 2's stage-game payoff function is such that b^* is a strict best reply to a^*, $b^\dagger$ is a strict best reply to $a^\dagger$, b'' is a strict best reply to a'', and b' is a strict best reply to a'.

Suppose $S \equiv \{s^*, s'', s'\}$ such that $f(s'|a') = 1$ and $f(s'|a) = 0$ for every $a \neq a'$. For every $s \in \{s^*, s''\}$ and $a \in \{a^*, a^\dagger, a''\}$, we have $f(s|a) > 0$, and $\frac{f(s^*|a)}{f(s''|a)}$ is strictly increasing in a.

Players' stage-game payoffs satisfy Assumptions 1 and 2 once we rank player 1's actions according to $a^* \succ a^\dagger \succ a'' \succ a'$ and player 2's actions according to $b^* \succ b^\dagger \succ b'' \succ b'$. The signal distribution satisfies MLRP once we rank the signal realizations according to $s^* \succ s'' \succ s'$. Player 1's Stackelberg action is a^*, which is not strongly separable from $a^\dagger$.

Player 1's commitment payoff from $a^\dagger$ is -1, which is strictly less than his minmax payoff 0. Hence, there exists no equilibrium in which player 1's payoff equals his commitment payoff from $a^\dagger$.

Next, I show there is no equilibrium where player 1's payoff equals his minmax payoff 0. Since player 1 is the commitment type with positive probability, both s^* and s'' occur with positive probability in period 0. Since both s^* and s'' occur on the equilibrium path, player 2's action is supported in $\{b^*, b^\dagger, b''\}$ after she observes s^* or s''. Suppose player 1 plays a'' in period 0, player 2_0 observes either s^* or s'', so her action is supported in $\{b^*, b^\dagger, b''\}$. This implies that player 1's stage-game payoff in period 0 is at least 1 and his expected continuation value after playing a'' is at least 0 in any PBE. Hence, player 1's discounted average payoff is strictly greater than his minmax payoff 0 in all PBEs.

One can obtain a higher payoff lower bound under the following refinement of PBE: For every history h^t no matter whether it is on-path or off-path, player 2_t's posterior belief about a_t after observing s_t is supported in $A(s_t) \equiv \big\{a \in A \big| f(s_t|a) > 0\big\}$. In every PBE that satisfies this refinement, suppose player 1 deviates and plays a'' in every period, then at every history, player 2 must be playing some mixed action supported in $\{b^*, b^\dagger, b''\}$. Hence, player 1's discounted average payoff from playing a'' in every period is at least 1, so his equilibrium payoff in every refined PBE must be no less than 1.

MLRP: In order to demonstrate that MLRP is not redundant, consider the following game:

-	b^*	b'
$\overline{a}$	$1,4$	$-2,0$
a^*	$2,1$	$-1,0$
$\underline{a}$	$3,-2$	$0,0$

Let $S \equiv \{\overline{s}, s^*, \underline{s}\}$, with $f(s^*|a^*) = 2/3$, $f(\underline{s}|a^*) = 1/3$, $f(\overline{s}|\overline{a}) = 1$, $f(\overline{s}|\underline{a}) = 1/3$, and $f(\underline{s}|\underline{a}) = 2/3$.

Players' payoffs satisfy Assumptions 1 and 2 when player 1's actions are ranked according to $\overline{a} \succ a^* \succ \underline{a}$ and player 2's actions are ranked according to $b^* \succ b'$. Player 1's Stackelberg action is a^*, his Stackelberg payoff is 2, s_t is unboundedly informative about a^*. However, MLRP is violated.

I construct an equilibrium where player 1's payoff is 1, which is bounded below his Stackelberg payoff 2. The strategic-type player 1 plays a mixed action that depends only on player 2's posterior belief about his type. If player 2's posterior belief assigns probability π to the commitment type, then the strategic-type player 1 plays $\alpha(\pi) \in \Delta(A)$ such that $(1-\pi)\cdot\alpha(\pi)+\pi\cdot a^* = 0.5\cdot a^*+0.25\cdot\overline{a}+0.25\cdot\underline{a}$. Player 2_t plays b^* if $s_t \in \{s^*, \overline{s}\}$, and plays b' if $s_t = \underline{s}$.

This strategy profile is an equilibrium since player 1's expected stage-game payoff is 1 no matter which action he plays, and his continuation value is independent of his current-period action. Player 2 has a strict incentive to play b^* after observing $\overline{s}$ or s^*, and has an incentive to play b' after observing $\underline{s}$. Regardless of player 1's type, the probability with which player 2 plays b^* in each period is 2/3.

In the above example, b_t is uninformative about player 1's type despite the probability of $b_t = b^*$ is bounded away from 1. As a result, even when player 1 builds a reputation for playing a^*, player 2 can still play b' with significant probability in unbounded number of periods. This explains why the patient player's equilibrium payoff is bounded below his Stackelberg payoff in some equilibria.

5 Concluding Remarks

I examine a patient seller's returns from building reputations when consumers have limited access to his past records and can learn from other consumers' choices.

My main result shows that consumers' observational learning can lead to reputation failures. This is because observing other consumers' choices enables consumers to imitate their predecessors, and consumers' imitation behaviors can be rationalized when each of them observes at most a bounded number of the seller's actions. When every consumer imitates her predecessor with high probability, the seller receives a low payoff as long as he receives a low stage-game payoff in the first period. I also show that the seller receives his minmax payoff in all equilibria where consumers do not trust him when he first arrives and do not trust him when the worst action profile occurred in the period before. In contrast, the seller receives at least his Stackelberg payoff in all equilibria when each consumer also observes a unboundedly informative private signal about his current-period action. I conclude by reviewing the related literature on social learning and reputation formation.

Social Learning: In the case where $M = +\infty$, my model is analogous to a social learning model where a sequence of myopic players observes their predecessors' choices and some private signals (e.g., the long-run player's actions in the last K periods) in order to forecast the long-run player's current-period action. This stands in contrast to the social learning models in Banerjee (1992), Bichandarni, Hirshleifer and Welsh (1992), Lee (1993), Smith and Sørensen (2000), Bose, Orosel, Ottaviani and Vesterlund (2006), and Kartik, Lee and Rappoport (2021) in which a sequence of myopic players learns about an exogenous payoff-relevant state, rather than some endogenous actions.

Due to differences in the object to learn, the myopic players asymptotically learn about the patient player's type is neither sufficient nor necessary for the patient player to receive a high *discounted* average payoff in my model. The differences in the object to learn also leads to different forms of inefficiencies. In canonical social learning models, inefficiencies arise when myopic players ignore their private signals and herd on some inefficient action. In contrast, the myopic players can never herd on any action other than b^* in any equilibrium of my baseline model.[19]

In terms of research question, I examine the effects of social learning on a patient player's discounted average payoff. This stands in contrast to existing results that focus on players' asymptotic beliefs, players' asymptotic rates of learning (e.g., Gale and Kariv 2003, Hann-Caruthers, Martynov and Tamuz 2018, Harel, Mossel, Strack and Tamuz 2021), and players' asymptotic payoffs (e.g., Rosenberg and Vieille 2019).[20] As demonstrated by the imitation equilibria in the constructive proof of Theorem 1, the patient player's discounted average payoff can be low even though his asymptotic payoff is high.

My paper is also related to social learning models with bounded memories. Drakopoulos, Ozdaglar and Tsitsiklis (2012) study a model where a sequence of myopic players learns about an exogenous state. Every player observes a private signal and the actions of her last M predecessors. They show that learning is possible when $M \geq 2$ but not when $M = 1$. In contrast, the myopic players in my model are learning about the endogenous behaviors of a strategic long-run player instead of an exogenous state. As a result, the informativeness of their private signal (which is the patient player's actions in the last K periods in my model) is also endogenous. In contrast to their conclusion which highlights the distinction between the case where $M = 1$ and the case where $M \geq 2$, the values of K and M do not play an important role here as long as K is finite and M is at least one.

Reputation Failure: Theorem 1 is related to the literature on reputation failures. Schmidt (1993), Cripps and Thomas (1997), and Chan (2000) assume that the uninformed player is forward-looking. They show that reputation fails in the sense that there *exist equilibria* in which the informed player receives a low payoff. The takeaway from their analysis is that the informed player's patience helps reputation building while the uninformed player's patience hurts reputation building.

In contrast, my analysis highlights another effect, that the informed player's patience makes it hard for his opponents to distinguish between the commitment type and the strategic type. This effect does not affect the patient player's payoff when his opponents observe his entire history, but plays an important role when each of his opponents only observes a bounded number of his actions. When each uninformed player receives limited information, there is a rationale for her to imitate her

[19] Logina, Lukyanov and Shamruk (2019) study a social learning model in which every myopic player observes a private signal about a patient player's action. They show that the patient player exerts high effort only when the myopic players' beliefs are intermediate. Their logic is similar to the one in Banerjee (1992) and Bichandarni, Hirshleifer and Welsh (1992). Board and Meyer-ter-Vehn (2020) study a model of innovation adoption in which players learn about a persistent exogenous state, and characterize the rate of learning under different network structures.

[20] Rosenberg and Vieille (2019) bound the discounted sum of a sequence of myopic players' payoffs when they learn about an exogenous state. Che and Hörner (2018) and Smith, Sørensen and Tian (2021) characterize mechanisms that maximize a sequence of myopic players' discounted average payoff when they learn about an exogenous payoff-relevant state. None of these papers examine what happens when players learn about the endogenous actions of a patient player.

predecessor, and her imitation behaviors wipe out the seller's returns from building reputations.

Ely and Välimäki (2003), Ely, Fudenberg and Levine (2008), and Deb, Mitchell and Pai (2021) focus on *participation games* where the uninformed player(s) can take an action under which the informed player receives his minmax payoff and future uninformed players cannot learn about his current-period action.[21] This lack-of-identification problem leads to equilibria with low payoffs. Deb and Ishii (2021) show that lack-of-identification occurs when uninformed players do not know the monitoring structure. In contrast, the uninformed players cannot shut down learning in my model and consumers' actions in imitation equilibria can statistically identify the seller's past actions.

Bai (2021) studies a model where the seller is either a low-cost type who may exert effort or a high-cost type who never exerts effort. Every consumer observes a noisy signal of the seller's effort and communicates the realized signal to all future consumers. She shows that the low-cost type has no incentive to exert effort when (1) δ is low, (2) consumers' prior belief attaches low enough probability to the low-cost type, and (3) the fixed cost of establishing a reputation is high enough. In contrast, reputation effects fail in my model since consumers have limited observation of the seller's past actions and can observe previous consumers' choices. I introduce a refinement and show that the seller's payoff equals his minmax payoff in all equilibria that satisfy this refinement, no matter how patient he is.

Reputation Models with Limited Memory: Liu (2011) and Liu and Skrzypacz (2014) study reputation models where consumers observe a bounded number of the seller's actions but *cannot* observe other consumers' choices.[22] I show that consumers' ability to observe other consumers' choices can lead to qualitatively different predictions. First, my reputation failure result needs consumers to observe other consumers' choices, as demonstrated by the comparison between Theorem 1 and Proposition 4. Second, in terms of players' equilibrium behaviors, the reputation cycles in Liu and Skrzypacz (2014) cannot arise in my model due to consumers' observational learning.

In Kaya and Roy (2020), a long-lived seller has persistent private information about his quality and decides whether to accept a myopic consumer's offer in every period. Quality affects both the seller's production cost and consumers' valuations. That is, values are interdependent in their model. When each consumer observes a bounded number of the seller's past actions but *cannot* observe previous consumers' price offers, they show that longer records can hurt the high-quality seller due to the low-quality seller's incentive to imitate. In contrast, the consumers' payoffs in my model do

[21]Levine (2021) studies a model where signals are less informative when the uninformed players do not participate.

[22]Heller and Mohlin (2018) and Bhaskar and Thomas (2019) study repeated games with random matching in which players cannot observe their opponents' actions taken more than K periods ago. They show that cooperation is sustainable in repeated prisoner's dilemma when payoffs are supermodular but not when payoffs are submordular, and that cooperation is sustainable in games with one-sided moral hazard when observations of opponents' past play are noisy.

not directly depend on the seller's type and each consumer can observe at least one other consumer's action in addition to a bounded number of the seller's actions. In contrast to their conclusion that longer memories of the seller's actions may hurt the high-quality seller, I show that consumers' ability to observe other consumers' choices can also hurt the seller.

A Proof of Theorem 1

Since $a^* \neq a'$, a' is the lowest action, and $u_1(a,b)$ is strictly decreasing in a, we know that $u_1(a',b') < u_1(a^*,b^*)$. I normalize player 1's payoff function by setting $u_1(a',b') = 0$ and $u_1(a^*,b^*) = 1$. Assumption 2 implies that $u_1(a,b') < 0$ for every $a \succ a'$ and $u_1(a,b^*) > 1$ for every $a \prec a^*$.

Let $\underline{q}$ be the largest $q \in [0,1]$ such that b' is not player 2's strict best reply to mixed action $qa^* + (1-q)a'$. Let $\overline{q}$ be the smallest $q \in [0,1]$ such that b^* is not player 2's strict best reply to mixed action $qa^* + (1-q)a'$. Assumption 1 implies that b^* is a strict best reply to a^* and b' is a strict best reply to a'. Hence $0 < \underline{q} < \overline{q} < 1$ and there exist $b^{**} \neq b'$ and $b'' \neq b^*$ such that $\{b^{**}, b'\} \subset \text{BR}_2(\underline{q}a^* + (1-\underline{q})a')$ and $\{b^*, b''\} \subset \text{BR}_2(\overline{q}a^* + (1-\overline{q})a')$. Assumption 2 implies that $b^* \succ b''$, $b^{**} \succ b'$, and $b^* \succ b'$. I consider the following three cases separately.

Case 1: $b^* = b^{**}$ and $b' = b''$.

Case 2: $b^* \succ b'' \succ b^{**} \succ b'$.

Case 3: $b^* \succ b'' = b^{**} \succ b'$.

First, I construct equilibria in which (1) player 1's ex ante payoff is 0, (2) player 2_t's action depends only on (a_{t-1}, b_{t-1}), (3) player 1's action in period t depends only on (a_{t-1}, b_{t-1}) and player 2's posterior belief about player 1's type, (4) player 1 plays either a^* or a' on the equilibrium path, and (5) if $a_{t-1} \notin \{a', a^*\}$, then the continuation play proceeds as if $(a_{t-1}, b_{t-1}) = (a', b_{t-1})$. Since $u_1(a,b)$ is strictly in a and a' is player 1's lowest action, the strategic-type player 1 strictly prefers a' to actions other than a^* and a' at any private history. I comment on $\underline{\delta}(u_1, u_2)$ by the end of this section.

Case 1: $b^* = b^{**}$ **and** $b' = b''$ In this case, $\underline{q} = \overline{q} \equiv q$. The construction resembles that in the product choice game after replacing H with a^*, L with a', T with b^*, and N with b'.

1. When $(a_{t-1}, b_{t-1}) = (a', b')$ or $\varnothing$. Player 2 plays b'. The strategic type player 1 mixes between a^* and a'. His probability of playing a^*, denoted by p_t, satisfies $\pi_t + (1-\pi_t)p_t = q$.

2. When $(a_{t-1}, b_{t-1}) = (a^*, b')$. Player 2 plays b^* with probability $-\frac{1-\delta}{\delta} u_1(a^*, b')$ and plays b' with complementary probability. The strategic type player 1 mixes between a^* and a'. His probability of playing a^*, denoted by p_t, satisfies $\pi_t + (1-\pi_t)p_t = q$.

3. When $(a_{t-1}, b_{t-1}) = (a', b^*)$. Player 2 plays b^* with probability $\frac{1-(1-\delta)u_1(a',b^*)}{\delta}$ and plays b' with complementary probability. The strategic type player 1 mixes between a^* and a'. His probability of playing a^*, denoted by p_t, satisfies $\pi_t + (1-\pi_t)p_t = q$.

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

Suppose $\pi_0 \leq \left(\frac{q}{2}\right)^{-K-1}$. Verifying players' incentive constraints and that player 2's posterior belief attaches probability less than $q/2$ to the commitment type at every history where $(a_{t-1}, b_{t-1}) \neq (a^*, b^*)$ follows from the same steps as in the product choice game, which I omit in order to avoid repetition.

Case 2: $b^* \succ b'' \succ b^{**} \succ b'$ Consider the following strategy profile, which is parameterized by $r(a^*, b')$, $r(a^*, b'')$, $r(a', b^*)$, and $r(a', b^{**})$, all of them belong to $(0, 1)$ and will be specified later on. Recall that π_t is player 2_t's belief about the commitment type.

1. When $(a_{t-1}, b_{t-1}) = (a', b')$ or (a', b'') or $\varnothing$. Player 2 plays b'. The strategic type player 1 mixes between a^* and a'. He plays a^* with probability p_t such that $\pi_t + (1 - \pi_t)p_t = \underline{q}$.

2. When $(a_{t-1}, b_{t-1}) = (a^*, b')$. Player 2 plays b^{**} with probability $r(a^*, b')$ and b' with complementary probability. The strategic type player 1 mixes between a^* and a'. He plays a^* with probability p_t such that $\pi_t + (1 - \pi_t)p_t = \underline{q}$.

3. When $(a_{t-1}, b_{t-1}) = (a^*, b'')$. Player 2 plays b^{**} with probability $r(a^*, b'')$ and b' with complementary probability. The strategic type player 1 mixes between a^* and a'. He plays a^* with probability p_t such that $\pi_t + (1 - \pi_t)p_t = \underline{q}$.

4. When $(a_{t-1}, b_{t-1}) = (a', b^*)$. Player 2 plays b^* with probability $r(a', b^*)$ and b'' with complementary probability. The strategic type player 1 mixes between a^* and a'. He plays a^* with probability p_t such that $\pi_t + (1 - \pi_t)p_t = \overline{q}$.

5. When $(a_{t-1}, b_{t-1}) = (a', b^{**})$. Player 2 plays b^* with probability $r(a', b^{**})$ and b'' with complementary probability. The strategic type player 1 mixes between a^* and a'. He plays a^* with probability p_t such that $\pi_t + (1 - \pi_t)p_t = \overline{q}$.

6. When $(a_{t-1}, b_{t-1}) = (a^*, b^*)$ or (a^*, b^{**}). Player 2 plays b^* and player 1 plays a^*.

Player 2's incentive constraint at every history is satisfied. Next, I compute player 1's continuation value in period t for every (a_{t-1}, b_{t-1}), which I denote by $V(a_{t-1}, b_{t-1})$. Then I verify player 1's incentive constraints. From the descriptions of players' strategies from (1) to (6), we know that $V(\varnothing) = V(a', b') = V(a', b'') = 0$ and $V(a^*, b^{**}) = V(a^*, b^*) = 1$. Player 1's indifference at $(a_{t-1}, b_{t-1}) = (a', b')$ implies that

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

Since $(1-\delta)u_1(a^*, b') + \delta V(a^*, b') = (1-\delta)u_1(a', b') + \delta u_1(a', b') = 0$, player 1 is indifferent when $(a_{t-1}, b_{t-1}) \in \{(a', b''), (a^*, b'), (a^*, b'')\}$ if and only if

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

which implies that

$$V(a', b^{**}) = 1 - \frac{1-\delta}{\delta}\Big(\underbrace{u_1(a', b^{**}) - u_1(a^*, b^{**})}_{>0}\Big). \tag{A.3}$$

Let $V(a', b^*)$ be such that player 1 is indifferent when $(a_{t-1}, b_{t-1}) = (a^*, b^*)$. This yields:

$$V(a', b^*) = \frac{1-(1-\delta)u_1(a', b^*)}{\delta}. \tag{A.4}$$

According to (A.4), player 1 is indifferent when $(a_{t-1}, b_{t-1}) \in \{(a^*, b^{**}), (a', b^*), (a', b^{**})\}$ if and only if

$$(1-\delta)u_1(a^*, b'') + \delta V(a^*, b'') = (1-\delta)u_1(a', b'') + \delta V(a', b'') = (1-\delta)u_1(a', b''). \tag{A.5}$$

This yields:

$$V(a^*, b'') = \frac{1-\delta}{\delta}\Big(\underbrace{u_1(a', b'') - u_1(a^*, b'')}_{>0}\Big). \tag{A.6}$$

Next, I pin down variables $r(a^*, b')$, $r(a^*, b'')$, $r(a', b^*)$, and $r(a', b^{**})$.

1. $r(a^*, b')$ is pinned down by:

$$\underbrace{V(a^*, b')}_{\text{positive but close to 0}} = r(a^*, b')\Big((1-\delta)u_1(a^*, b^{**}) + \delta \underbrace{V(a^*, b^{**})}_{=1}\Big).$$

   Such $r \in [0, 1]$ exists since $0 < V(a^*, b') < (1-\delta)u_1(a^*, b^{**}) + \delta V(a^*, b^{**})$ when δ is large enough.

2. $r(a^*, b'')$ is pinned down by:

$$\underbrace{V(a^*, b'')}_{\text{positive but close to 0}} = r(a^*, b'')\Big((1-\delta)u_1(a^*, b^{**}) + \delta V(a^*, b^{**})\Big).$$

   Such $r \in [0, 1]$ exists since $0 < V(a^*, b'') < (1-\delta)u_1(a^*, b^{**}) + \delta V(a^*, b^{**})$ when δ is large enough.

3. $r(a', b^*)$ is pinned down by:

$$\underbrace{V(a', b^*)}_{\text{less than but close to 1}} = r(a', b^*) + (1 - r(a', b^*))\Big((1-\delta)u_1(a^*, b'') + \delta \underbrace{V(a^*, b'')}_{\text{positive but close to 0}}\Big).$$

Such $r \in [0, 1]$ exists since $(1-\delta)u_1(a^*, b'') + \delta V(a^*, b'') < V(a', b^*) < 1$ when δ is large enough.

4. $r(a', b^{**})$ is pinned down by:

$$\underbrace{V(a', b^{**})}_{\text{less than but close to 1}} = r(a', b^{**}) + (1 - r(a', b^{**}))\Big((1-\delta)u_1(a^*, b'') + \delta \underbrace{V(a^*, b'')}_{\text{positive but close to 0}}\Big).$$

Such $r \in [0, 1]$ exists since $(1-\delta)u_1(a^*, b'') + \delta V(a^*, b'') < V(a', b^*) < 1$ when δ is large enough.

When the prior probability of commitment type is less than $\overline{\pi}_0$ where $\overline{\pi}_0$ is given by

$$\frac{\overline{\pi}_0}{1 - \overline{\pi}_0} = \Big(\frac{\underline{q}}{2}\Big)^{K+1}, \tag{A.7}$$

player 2's posterior belief attaches probability less than $\underline{q}/2$ to the commitment type at every history where $(a_{t-1}, b_{t-1}) \notin \{(a^*, b^*), (a^*, b^{**})\}$. This implies that the strategic type player 1 plays a^* with probability at least $\underline{q}/2$ at every history, and that his mixed action at every history is well-defined.

Case 3: $b^* \succ b'' = b^{**} \succ b'$ I write b'' instead of b^{**}. Consider the following strategy profile, parameterized by $s(a^*, b')$, $s(a^*, b'')$, $s(a', b^*)$, and $s(a', b^{**})$.

1. When $(a_{t-1}, b_{t-1}) = (a', b')$ or $\varnothing$. Player 2 plays b'. The strategic type player 1 mixes between a^* and a'. He plays a^* with probability p_t such that $\pi_t + (1 - \pi_t)p_t = \underline{q}$.

2. When $(a_{t-1}, b_{t-1}) = (a^*, b')$. Player 2 plays b'' with probability $s(a^*, b')$ and b' with complementary probability. The strategic type player 1 mixes between a^* and a'. He plays a^* with probability p_t such that $\pi_t + (1 - \pi_t)p_t = \underline{q}$.

3. When $(a_{t-1}, b_{t-1}) = (a', b'')$. Player 2 plays b'' with probability $s(a', b'')$ and b' with complementary probability. The strategic type player 1 mixes between a^* and a'. He plays a^* with probability p_t such that $\pi_t + (1 - \pi_t)p_t = \underline{q}$.

4. When $(a_{t-1}, b_{t-1}) = (a^*, b'')$. Player 2 plays b^* with probability $s(a^*, b'')$ and b'' with complementary probability. The strategic type player 1 mixes between a^* and a'. He plays a^* with probability p_t such that $\pi_t + (1 - \pi_t)p_t = \overline{q}$.

5. When $(a_{t-1}, b_{t-1}) = (a', b^*)$. Player 2 plays b^* with probability $s(a', b^{**})$ and b'' with complementary probability. The strategic type player 1 mixes between a^* and a'. He plays a^* with probability p_t such that $\pi_t + (1-\pi_t)p_t = \overline{q}$.

6. When $(a_{t-1}, b_{t-1}) = (a^*, b^*)$. Player 2 plays b^* and player 1 plays a^*.

According to (1) and (6), $V(\varnothing) = V(a', b') = 0$ and $V(a^*, b^*) = 1$. Player 1's indifference at (a', b') implies that $V(a^*, b') = -\frac{1-\delta}{\delta}u_1(a^*, b')$. Let $V(a', b^*) = \frac{1-(1-\delta)u_1(a', b^*)}{\delta}$, under which player 1 is indifferent between a^* and a' when $(a_{t-1}, b_{t-1}) = (a^*, b^*)$.

Since $(1-\delta)u_1(a^*, b') + \delta V(a^*, b') = (1-\delta)u_1(a', b') + \delta V(a', b')$ and $(1-\delta)u_1(a^*, b^*) + \delta V(a^*, b^*) = (1-\delta)u_1(a', b^*) + \delta V(a', b^*)$ under these continuation values, the strategic type of player 1 is indifferent at (a^*, b'), (a', b''), (a^*, b''), and (a', b^*) if and only if

$$(1-\delta)u_1(a^*, b'') + \delta V(a^*, b'') = (1-\delta)u_1(a', b'') + \delta V(a', b''). \tag{A.8}$$

Assumption 2 implies that $u_1(a', b'') > u_1(a^*, b'')$, $u_1(a^*, b'') < u_1(a^*, b^*)$ and $u_1(a', b'') > u_1(a', b')$.

Lemma A.1. *There exists $\gamma \in (0,1) \cap (u_1(a^*, b''), u_1(a', b''))$ such that*

$$\gamma(1 - u_1(a^*, b'')) \geq (1-\gamma)u_1(a', b''). \tag{A.9}$$

Proof. Consider two cases separately. First, suppose $u_1(a',, b'') \leq 1$. By setting $\gamma = u_1(a', b'')$,

$$\gamma(1 - u_1(a^*, b'')) = u_1(a', b'')(1 - u_1(a^*, b'')) > u_1(a', b'')(1 - u_1(a', b'')).$$

The intermediate value theorem implies that (A.9) holds for some γ that is strictly less than $u_1(a', b'')$ but is strictly greater than $u_1(a^*, b'')$. Second, suppose $u_1(a', b'') > 1$. By setting $\gamma = 1$, the left-hand-side of (A.9) is strictly positive while the right-hand-side of (A.9) is 0. The intermediate value theorem implies that (A.9) holds for some γ that is strictly less than 1 but is strictly greater than $u_1(a^*, b'')$ □

Pick $\gamma \in (0,1) \cap (u_1(a^*, b''), u_1(a', b''))$ that satisfies (A.9) and set player 1's continuation values at (a^*, b'') and (a', b'') to be

$$V(a^*, b'') = \frac{1}{\delta}\big(\gamma - (1-\delta)u_1(a^*, b'')\big) \tag{A.10}$$

and

$$V(a', b'') = \frac{1}{\delta}\big(\gamma - (1-\delta)u_1(a', b'')\big). \tag{A.11}$$

These continuation values satisfy player 1's incentive constraint (A.8), and moreover,

$$V(a^*,b'') > (1-\delta)u_1(a^*,b'') + \delta V(a^*,b'') = \gamma = (1-\delta)u_1(a',b'') + \delta V(a',b'') > V(a',b'').$$

When δ is close to 1, both $V(a^*,b'')$ and $V(a',b'')$ are bounded away from 0 and 1, and moreover, $V(a',b'') < u_1(a',b'')$ and $V(a^*,b'') > u_1(a^*,b'')$.

Next, I pin down the values of $s(a^*,b')$, $s(a^*,b'')$, $s(a',b^*)$, and $s(a',b'')$ so that player 1 receives these continuation values. Recall that $V(a^*,b') = -\frac{1-\delta}{\delta}u_1(a^*,b')$ and $V(a',b^*) = \frac{1-(1-\delta)u_1(a',b^*)}{\delta}$, and the values of $V(a^*,b'')$ and $V(a',b'')$ are given by (A.10) and (A.11).

1. $s(a^*,b')$ is pinned down by:

$$\underbrace{V(a^*,b')}_{\text{positive but close to 0}} = s(a^*,b')\Big((1-\delta)u_1(a^*,b'') + \delta \underbrace{V(a^*,b'')}_{\text{bounded away from 0}}\Big).$$

   Such $s \in [0,1]$ exists since $0 < V(a^*,b') < (1-\delta)u_1(a^*,b'') + \delta V(a^*,b'')$ when δ is large enough.

2. $s(a',b'')$ is pinned down by:

$$V(a',b'') = s(a',b'')\Big((1-\delta)u_1(a',b'') + \delta V(a',b'')\Big).$$

   Such $s \in [0,1]$ exists since $0 < V(a',b'') < (1-\delta)u_1(a',b'') + \delta V(a',b'')$ when δ is large enough.

3. $s(a^*,b'')$ is pinned down by:

$$V(a^*,b'') = s(a^*,b'') + (1-s(a^*,b''))\Big((1-\delta)u_1(a^*,b'') + \delta V(a^*,b'')\Big).$$

   Such $s \in [0,1]$ exists since $(1-\delta)u_1(a^*,b'') + \delta V(a^*,b'') < V(a^*,b'') < 1$ when δ is large enough.

4. $s(a',b^*)$ is pinned down by:

$$\underbrace{V(a',b^*)}_{\text{close to but less than 1}} = s(a',b^*) + (1-s(a',b^*))\Big((1-\delta)u_1(a^*,b'') + \delta \underbrace{V(a^*,b'')}_{\text{bounded away from 1}}\Big)$$

   Such $s \in [0,1]$ exists since $(1-\delta)u_1(a^*,b'') + \delta V(a^*,b'') < V(a',b^*) < 1$ when δ is large enough.

Next, I show that player 2's posterior belief attaches probability less than $\underline{q}/2$ to the commitment type at every history where $(a_{t-1},b_{t-1}) \neq (a^*,b^*)$. The key step is Lemma A.2.

Lemma A.2. *If γ satisfies (A.9), then $s(a',b'') + s(a^*,b'') \geq 1$.*

Proof. According to the expressions of player 1's continuation value, we have

$$s(a^*, b'') = \frac{V(a^*, b'') - \gamma}{1 - \gamma} \quad \text{and} \quad s(a', b'') = \frac{V(a', b'')}{\gamma}. \tag{A.12}$$

Therefore, $s(a', b'') + s(a^*, b'') \geq 1$ if and only if

$$\frac{V(a^*, b'') - \gamma}{1 - \gamma} + \frac{V(a', b'')}{\gamma} \geq 1$$

which is equivalent to $(1-\gamma)V(a', b'') \geq \gamma(1-V(a^*, b''))$. Plugging in (A.10) and (A.11), this inequality is equivalent to $\gamma(1 - u_1(a^*, b'')) \geq (1 - \gamma)u_1(a', b'')$, which is (A.9). □

Since player 2 plays b'' with probability $1 - s(a^*, b'')$ when $(a_{t-1}, b_{t-1}) = (a^*, b'')$ and plays b'' with probability $s(a', b'')$ when $(a_{t-1}, b_{t-1}) = (a', b'')$, Lemma A.2 implies that

$$\Pr(b_{t+1} = b'' | b_t = b'', a_t = a') \geq \Pr(b_{t+1} = b'' | b_t = b'', a_t = a^*). \tag{A.13}$$

Therefore, the likelihood ratio between the commitment type and the strategic type does not increase when player 2 observes $b_{t+1} = b''$ conditional on $b_t = b''$. Back to the proof of $\pi_t \leq \underline{q}/2$ whenever $(a_{t-1}, b_{t-1}) \neq (a^*, b^*)$, we only need to consider histories such that $a_{t-1} = a^*$. Assume $\pi_0 < \overline{\pi}_0$ where $\overline{\pi}_0$ is given by

$$\frac{\overline{\pi}_0}{1 - \overline{\pi}_0} = \left(\frac{\underline{q}}{2}\right)^{K+1} \frac{\underline{q}}{2 - \underline{q}}. \tag{A.14}$$

1. At histories where $(a_{t-1}, b_{t-1}) = (a^*, b')$, then the same argument as that in Section 3 implies that when π_0 is no more than $\overline{\pi}_0$ defined in (A.14), player 2's posterior belief attaches probability less than $\underline{q}/2$ at every such history.

2. At histories where $(a_{t-1}, b_{t-1}) = (a^*, b'')$, then player 2's posterior belief about the commitment type is strictly positive only if $(a_{t-K}, ..., a_{t-1}) = (a^*, ..., a^*)$ and there exists $s \leq t - 1$ such that $b_\tau = b'$ for every $\tau < s$ and $b_\tau = b''$ for every $t - 1 \geq \tau \geq s$. Let E_t be the event that $(a_{t-K}, ..., a_{t-1}) = (a^*, ..., a^*)$, let $F_{s,t}$ be the event that $(b_0, ..., b_{t-1}) = (b', ..., b', b'', b'', ..., b'')$ where the first b'' occurs in period s. Let $\pi^*_{s,t}$ be the posterior probability of commitment type conditional on $E_t \cap F_t$. According to Bayes rule,

$$\frac{\pi^*_{s,t}}{1 - \pi^*_{s,t}} \bigg/ \frac{\pi_0}{1 - \pi_0} = \frac{P^{\omega_c}(E_t \cap F_t)}{P^{\omega_s}(E_t \cap F_t)} = \frac{P^{\omega_c}(E_t)}{P^{\omega_s}(E_t)} \cdot \frac{P^{\omega_c}(F_t | E_t)}{P^{\omega_s}(F_t | E_t)}. \tag{A.15}$$

The first term on the right-hand-side of (A.15) is no more than $(\underline{q}/2)^{-K}$. For every $n < s$, let

$$l_n \equiv \frac{P^{\omega_c}(a_n = a'|E_t, (b_0, ..., b_{n-1}) = (b', ..., b'))}{P^{\omega_s}(a_n = a'|E_t, (b_0, ..., b_{n-1}) = (b', ..., b'))} \tag{A.16}$$

and for every $n \geq s$, let

$$l_n \equiv \frac{P^{\omega_c}(a_n = a''|E_t, (b_0, ..., b_{n-1}) = (b', ..., b'))}{P^{\omega_s}(a_n = a''|E_t, (b_0, ..., b_{n-1}) = (b', ..., b'))} \tag{A.17}$$

According to Bayes rule, the second term on the right-hand-side of (A.15) equals $\Pi_{i=0}^{t-1} l_i$. According to Lemma A.2, $l_n \leq 1$ for every $n \neq s$. Since $\pi_0 \leq \overline{\pi}_0$, we have $\pi^*_{s,t} \leq \underline{q}/2$ for every $t \leq s$. Since $\pi_t \leq \max_{s \leq t} \pi^*_{s,t}$, we have $\pi_t \leq \underline{q}/2$ for every $t \leq s$. Since the unconditional probability with which player 1 plays a^* is at least $\underline{q}$ in every period and $\pi^*_{s,s} \leq \underline{q}/2$, we have $l_s \leq (\underline{q}/2)^{-1}$. This implies that $\pi_t \leq \underline{q}/2$ for every $t \in \mathbb{N}$, which concludes the proof.

Remark: I provide sufficient conditions for the cutoff discount factor $\underline{\delta}(u_1, u_2)$. Recall we adopt the normalization that $u_1(a^*, b^*) = 1$ and $u_1(a', b') = 0$. In Case 1, the cutoff discount factor is:

$$\underline{\delta}(u_1, u_2) = \max\Big\{\frac{-u_1(a^*, b')}{1 - u_1(a^*, b')}, 1 - \frac{1}{u_1(a', b^*)}\Big\}.$$

In Case 2, the cutoff discount factor is pinned down by $V(a^*, b') \leq (1-\delta)u_1(a^*, b^{**})+\delta$, $V(a^*, b'') \leq (1-\delta)u_1(a^*, b^{**})+\delta$, $(1-\delta)u_1(a^*, b'')+\delta V(a^*, b'') \leq V(a', b^*)$, and $(1-\delta)u_1(a^*, b'')+\delta V(a^*, b'') \leq V(a', b^*)$, where $V(a^*, b')$, $V(a', b^{**})$, $V(a', b^*)$ and $V(a^*, b'')$ are given by (A.1), (A.3), (A.4), and (A.6). In Case 3, the cutoff discount factor is pinned down by $V(a^*, b') \leq (1-\delta)u_1(a^*, b'') + \delta V(a^*, b'')$, $V(a', b'') \leq (1-\delta)u_1(a', b'')+\delta V(a', b'')$, $(1-\delta)u_1(a^*, b'')+\delta V(a^*, b'') \leq V(a^*, b'')$, and $(1-\delta)u_1(a^*, b'')+\delta V(a^*, b'') \leq V(a', b^*)$, where $V(a^*, b') = -\frac{1-\delta}{\delta}u_1(a^*, b')$, $V(a', b^*) = \frac{1-(1-\delta)u_1(a', b^*)}{\delta}$, and the values of $V(a^*, b'')$ and $V(a', b'')$ are given by (A.10) and (A.11).

B Proof of Proposition 1

For every $t \geq 1$, let $\widehat{\mathcal{H}}^t$ be the set of period-t histories where b' and actions in A' were played from period 0 to period $t-1$. In every PBE that satisfies *no initial trust*, player 2 plays b' in period 0. Since $M \geq 1$, player 2_0 knows that calendar time is 0 from her history (since she observes no action at all). Hence, player 2_0's incentive to play b' implies that the strategic-type player 1 must be playing some action in A' with positive probability in period 0.

I show that for every $t \geq 1$ and at every $h^t \in \widehat{\mathcal{H}}^t$, the strategic-type player 1 has an incentive to play some action in A' at h^t. Since $h^t \in \widehat{\mathcal{H}}^t$, we have $a_{t-1} \in A'$ and $b_{t-1} = b'$. Punishing following bad outcome implies that player 2_t plays b' at h^t.

Suppose by way of contradiction that at some $h^t \in \widehat{\mathcal{H}}^t$, the strategic-type player 1 has no incentive to play any action in A', then player 2_t's incentive to play b' at h^t (due to punishment following bad outcome) implies that there exists $\widetilde{h}^s$ such that (1) player 2_t *cannot* distinguish between h^t and $\widetilde{h}^s$, and (2) the strategic-type player 1 plays some action in A' with positive probability at $\widetilde{h}^s$. In another word, some action in A' is player 1's best reply at $\widetilde{h}^s$ but not at h^t. Since player 1's best replies at h^t and $\widetilde{h}^s$ are different, it must be the case that there exists $\tau \geq t$, such that player 2_τ *can* distinguish between h^t and $\widetilde{h}^s$. Since $\tau \geq t$, this contradicts the presumption that player 2_t *cannot* distinguish between h^t and $\widetilde{h}^s$.

Hence, there exists a best reply for the strategic-type player 1, denoted by σ_1^*, that plays some pure action in A' in period 0 and plays some pure action in A' at every $h^t \in \widehat{\mathcal{H}}^t$ for every $t \in \mathbb{N}$, from which he obtains his equilibrium payoff when player 2s play their equilibrium strategy. Since player 2 plays b' in period 0, *punishment following bad outcome* implies that player 2 plays b' in every period if player 1 plays according to σ_1^*. Since $u_1(a, b)$ is strictly decreasing in a, player 1's discounted average payoff from playing his best reply σ_1^* is no more than $u_1(a', b')$.

C Proof of Proposition 2

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

Next, I establish the result when M is finite and is at least one. Suppose by way of contradiction that player 2s herd on $b \neq b^*$ at h^t. Since K and M are both finite, player 2_t's action must be measurable with respect to $(a_{\max\{0,t-K\}}, ..., a_{t-1}, b_{\max\{0,t-M\}}, ..., b_{t-1})$. For every $t \geq \max\{M, K\}$, $(a_{t-K}, ..., a_{t-1}, b_{t-M}, ..., b_{t-1})$, and $h^t \equiv (a_s, b_s)_{s \leq t-1}$, there exists $h^T \succ h^t$ such that player 2's action at h^T coincides with her action at $(a_{t-K}, ..., a_{t-1}, b_{t-M}, ..., b_{t-1})$. Therefore, player 2s herding on $b \neq b^*$ at any $h^t \equiv (a_s, b_s)_{s \leq t-1}$ implies that they play b in every period after $\max\{K, M\}$. Hence, the strategic-type player 1 has no intertemporal incentive after period T. Consider two cases. Suppose $\mathrm{BR}_1(b) \neq \{a^*\}$, then the strategic type has no incentive to play a^*, so player 2 attaches probability 1 to the commitment type after observing a^*, which means that player 2 has a strict incentive to play b^*. This contradicts the presumption that they herd on action $b \neq b^*$. Suppose $\mathrm{BR}_1(b) \neq \{a^*\}$, then in equilibrium, the strategic type has no incentive to play a^* and player 2 has a strict incentive to play b^*. This contradicts the presumption that they herd on action $b \neq b^*$.

D Proof of Proposition 3

I establish the lower bound on player 1's undiscounted average payoff in Section D.1. I construct an equilibrium in which player 1's asymptotic payoff equals the right-hand-side of (3.5) in Section D.2.

D.1 Lower Bound on Undiscounted Average Payoff

Consider the strategic-type's payoff when he deviates and imitates the commitment type. For every $\beta \in \Delta(B)$ and $a \prec a^*$, Assumption 2 implies that $u_1(a^*, \beta) < u_1(a, \beta)$. Let $h^t \equiv \{a_s, b_s\}_{s=0}^{t-1}$. For every $t \in \mathbb{N}$ and $a \in A$, let $E_t(a, b^t)$ be the event that (1) player 1 plays a in period t, (2) player 1 has played a^* from period $t-K+1$ to $t-1$, (3) player 1 plays according to σ_1 starting from period $t+1$, and (4) the history of player 2's actions until period t is $b^t \equiv (b_0, ..., b_{t-1})$. For every $\tau \in \{1, 2, ..., K\}$ and $h^t \equiv (a^*, ..., a^*, b^t)$, let $y_t^\tau(\cdot|a, h^t) \in \Delta(B)$ be the distribution of $b_{t+\tau}$ conditional on event $E_t(a, b^t)$, and let $y_t(\cdot|a, h^t) \in \Delta(B^K)$ be the distribution of $(b_{t+1}, ..., b_{t+K})$ conditional on event $E_t(a, b^t)$. Let $\overline{u}_1$ and $\underline{u}_1$ be player 1's highest and lowest feasible stage-game payoffs, respectively, and let $||\cdot||$ be the total variation norm. If

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

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

2_{t+K}. Let

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

Since

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

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

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

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

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

$$\frac{1}{K+1}\mathbb{E}^{(a^*,\sigma_2)}\Big[\sum_{s=t}^{t+K} u_1(a_s,b_s)\Big|h^t\Big] \geq \frac{K}{K+1}u_1(a^*,b^*)+\frac{1}{K+1}u_1(a^*,b'). \tag{D.3}$$

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

$$d\Big(y_t(\cdot|\alpha,h^t)\Big\|y_t(\cdot|a^*,h^t)\Big) \geq 2\gamma^2\Delta^2. \tag{D.4}$$

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

$$\mathbb{E}^{(a^*,\sigma_2)}\Big[\sum_{s=0}^{\infty} d\Big(y_{s(K+1)+\tau}(\cdot|\sigma_1(h^{s(K+1)+\tau}),h^{s(K+1)+\tau})\Big\|y_{s(K+1)+\tau}(\cdot|a^*,h^{s(K+1)+\tau})\Big)\Big] \leq -\log\pi_0. \tag{D.5}$$

Inequalities (D.4) and (D.5) together imply that:

$$\mathbb{E}^{(a^*,\sigma_2)}\Big[\sum_{s=0}^{\infty}\mathbf{1}\Big\{h^{s(K+1)+\tau}\in\mathcal{H}_1^{(a^*,\sigma_2)} \text{ and } \sigma_2(h^{s(K+1)+\tau})\prec b^*\Big\}\Big] \leq -\frac{\log\pi_0}{2\gamma^2\Delta^2} \tag{D.6}$$

I derive a lower bound for $\liminf_{t\to\infty}\frac{1}{t}\mathbb{E}^{(a^*,\sigma_2)}\Big[\sum_{s=0}^{t-1}u_1(a_s,b_s)\Big]$ using inequalities (D.3) and (D.6). For every $\tau\in\{0,1,...,K\}$, let

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

let

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

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

$$\mathcal{H}_1^\tau\cap\mathcal{H}_1^{\tau'}=\{\varnothing\}\text{ and }\mathcal{H}_0^\tau\cap\mathcal{H}_0^{\tau'}=\{\varnothing\}.\tag{D.7}$$

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

For every $\tau\in\{0,1,...,K\}$, inequality (D.3) implies that player 1's expected average payoff at histories in $\mathcal{H}_0^\tau$ is at least the right-hand-side of (3.5). Since $\mathcal{H}_0^\tau\cap\mathcal{H}_0^{\tau'}=\{\varnothing\}$ for every $\tau\neq\tau'$, it implies that player 1's expected average payoff at histories in $\bigcup_{\tau=0}^K\mathcal{H}_0^\tau$ is at least the right-hand-side of (3.5). For every $\tau\in\{0,1,...,K\}$, (D.6) implies that player 1's expected average payoff at histories belonging to set $\mathcal{H}_1^\tau\Big\backslash\bigcup_{s=0}^K\mathcal{H}_0^s$ is at least $u_1(a^*,b^*)$. Since $\mathcal{H}_1^\tau\cap\mathcal{H}_1^{\tau'}=\{\varnothing\}$ for every $\tau\neq\tau'$, it implies that player 1's expected average payoff at histories in $\bigcup_{s=0}^K\mathcal{H}_1^s\Big\backslash\bigcup_{s=0}^K\mathcal{H}_0^s$ is at least $u_1(a^*,b^*)$. The two parts imply that

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

D.2 Tightness of Lower Bound

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

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

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

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

- (1) there exists no $r<t$, such that $b_r \neq b^*$ and r is not divisible by $K+1$; (2) there exists no $s<t$ such that $b_s \neq b'$ and s is divisible by $K+1$; (3) player 2 observes player 1 playing an off-path action in period $t-1$, then players play (a^*, b^*) if t is divisible by $K+1$, and play (a', b') if t is not divisible by $K+1$.

- (1) there exists no $r<t$, such that $b_r \neq b^*$ and r is not divisible by $K+1$, but (2) there exists $s<t$ such that $b_s \neq b'$ and s is divisible by $K+1$. If $t-1$ is divisible by $K+1$, $b_{t-1}=b^*$ while $a_{t-1} \neq a^*$, then play (a', b') in period t. If $t-1$ is divisible by $K+1$, $b_{t-1}=b^*$ while $a_{t-1}=a^*$, then play (a^*, b^*) in period t if and only if $\xi_t > 1/2$ and play (a', b') in period t otherwise. If $t-1$ is not divisible by $K+1$, or $b_{t-1} \neq b^*$, then play (a^*, b^*) if t is not divisible by $K+1$ and play (a', b') if t is divisible by $K+1$.

- there exists $r<t$, such that $b_r \neq b^*$ and r is not divisible by $K+1$, then play (a', b') in all subsequent periods.

Player 1's undiscounted time-average payoff from playing a^* in every period equals the right-hand-side of (3.5). I verify players' incentive constraints. Since b^* best replies to a^* and b' best replies to a', player 2's incentive constraints are satisfied. I verify player 1's incentives. At every on-path h^t,

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

$$V \equiv \frac{u_1(a', b') + \delta u_1(a^*, b^*) + \delta^2 u_1(a^*, b^*) + \ldots + \delta^K u_1(a^*, b^*)}{1+\delta+\ldots+\delta^K}, \tag{D.8}$$

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

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

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

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

E Proof of Proposition 4

Player 2's strategy is represented by a triple $(r_\varnothing, r_H, r_L)$, where r_x is the probability with which she plays T when $a_{t-1} = x$ for $x \in \{\varnothing, H, L\}$. First, I show that $r_H > r_L$. Suppose by way of contradiction that $r_H \leq r_L$, then the strategic-type player 1 has no incentive to play H. After player 2 observes $a_{t-1} = H$, she infers that player 1 is the commitment type for sure and has a strict incentive to play T, which implies that $r_H = 1$. Since $r_H \leq r_L$, we have $r_L = 1$ as well. However, since player 2_t knows that player 1 is the strategic type after observing $a_{t-1} = L$ and the strategic-type player 1 has no incentive to play H, we know that $r_L = 0$. This contradicts the previous conclusion that $r_L = 1$.

Since player 2_t's strategy depends only on a_{t-1}, starting from period 1, player 1's continuation value depends only on whether $a_{t-1} = L$ or $a_{t-1} = H$. Let $V(L)$ and $V(H)$ be these continuation values, respectively. Player 1 has an incentive to play H when $a_{t-1} = H$ if and only if $(1-\delta)(r_H + (1-r_H)(-c_N)) + \delta V(H) - (1-\delta)(1+c_T)r_H - \delta V(L) \geq 0$, or equivalently,

$$\frac{\delta}{1-\delta}(V(H) - V(L)) \geq c_T r_H + c_N(1-r_H). \tag{E.1}$$

Similarly, player 1 has an incentive to play H when $a_{t-1} = L$ if and only if

$$\frac{\delta}{1-\delta}(V(H) - V(L)) \geq c_T r_L + c_N(1-r_L). \tag{E.2}$$

Since $r_H > r_L$ and $c_N > c_T$, the right-hand-side of (E.2) is strictly greater than the right-hand-side of (E.1), which implies the following two statements:

- If player 1 is indifferent between H and L when $a_{t-1} = L$, then player 1 has a strict incentive to play H when $a_{t-1} = H$.
- If player 1 is indifferent between H and L when $a_{t-1} = H$, then player 1 has a strict incentive to play L when $a_{t-1} = L$.

I consider several cases separately. First, suppose player 1 has a strict incentive to play L when $a_{t-1} = H$, then he also has a strict incentive to play L when $a_{t-1} = L$. Then by observing $a_{t-1} = H$, player 2 infers that player 1 is the commitment type and has a strict incentive to play T, which implies that $r_H = 1$. A strategic type player 1 can guarantee discounted average payoff at least $\delta - (1-\delta)c_N$ by playing H in every period.

Next, suppose player 1 has a strict incentive to play H when $a_{t-1} = H$, then after player 2 observes $a_{t-1} = H$, she knows that player 1 will play H regardless of his type and will have a strict incentive to play T. As a result, $r_H = 1$. A strategic type player 1 can guarantee discounted average payoff at least $\delta - (1-\delta)c_N$ by playing H in every period.

The above reasoning implies that in every equilibrium where the strategic-type player 1 receives a payoff strictly less than $\delta - (1-\delta)c_N$, the strategic-type player 1 is indifferent when $a_{t-1} = H$ and strictly prefers to play L when $a_{t-1} = L$, and moreover, $r_H < 1$. I show that there is no such equilibria when δ_1 is close to 1, which is the case when δ is close to 1. Let p_t be the probability of the event:

$$E_t \equiv \Big\{\text{Player 1 is the strategic type and plays } H \text{ in period } t\Big\}.$$

Since the strategic type strictly prefers to play L in period t when $a_{t-1} = L$, we have $1-\pi_0 \geq p_0 \geq p_1 \geq p_2 \geq ...$ Since player 2's prior belief attaches probability π_0 to the commitment type and probability $\delta_1^t(1-\delta_1)$ to the calendar time being t, she prefers N to T after observing $a_{t-1} = H$ only if

$$\sum_{t=1}^{+\infty}(1-\delta_1)\delta_1^t(\pi_0 + 2p_t - p_{t-1}) \leq 0. \tag{E.3}$$

Since $\pi_0 + 2p_t - p_{t-1} \leq \frac{\pi_0}{2}$ only if $p_{t-1} - p_t \geq p_t + \frac{\pi_0}{2} \geq \frac{\pi_0}{2}$, there can be at most $T \equiv \left\lceil \frac{2(1-\pi_0)}{\pi_0} \right\rceil$ periods where $\pi_0 + 2p_t - p_{t-1} \leq \frac{\pi_0}{2}$. Since $\pi_0 + 2p_t - p_{t-1} \geq -1$, we have

$$\sum_{t=1}^{+\infty}(1-\delta_1)\delta_1^t(\pi_0 + 2p_t - p_{t-1}) \geq -(\delta_1 - \delta_1^{T+1}) + \delta_1^{T+1}\frac{\pi_0}{2}. \tag{E.4}$$

The right-hand-side of (E.4) is strictly positive when δ_1 is close to 1, which contradicts (E.3). Since $\delta < \delta_1$, the above contradiction implies that such equilibria do not exist when δ is close to 1.

F Proofs in Section 4

Appendix F.1 shows Theorem 2. Appendix F.2 establishes the existence of equilibrium when the private signal is unboundedly informative, $M = +\infty$, and δ is large. Appendix F.3 shows Theorem 3.

F.1 Proof of Theorem 2

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

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

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

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

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

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

Lemma F.2. *Suppose the private signal is unboundedly informative about a^*, and satisfies MLRP. For every $\pi^* \in (0,1)$, there exists $c > 0$ such that for every $\nu \in (0,1)$, $\alpha \in \Delta(A)$ with $\alpha(a^*) > \pi^*$, and $\beta : S \to \Delta(B)$ that best replies to α. If $\gamma(a^*,\beta)[b^*] < 1-\nu$, then $d\big(\gamma(\alpha,\beta)\big\|\gamma(a^*,\beta)\big) > 2c\nu^2$.*

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

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

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

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

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

and

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

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

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

Since the distribution over private signals satisfies MLRP,

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

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

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

and

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

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

$$\eta(1-\frac{1}{M(\pi^*)})-\lambda \sum_{s\in[\underline{s},\underline{s}^*)} f(s|\overline{\alpha})-(1-\lambda) \sum_{s\in(\overline{s}^*,\overline{s}]} f(s|\underline{\alpha}) \leq D\Big\{1+\frac{\lambda}{D+\sum_{s\succ\overline{s}} f(s|a^*)}+\frac{1-\lambda}{D+\sum_{s\prec\underline{s}} f(s|a^*)}\Big\}. \tag{F.6}$$

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

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

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

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

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

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

As a result,

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

F.2 Existence of Equilibrium

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

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

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

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

- At any other history, player 2_t believes that $a_t=a(s_t)$ upon receiving any $s_t\in S$, and plays $b(s_t)$. Player 1 plays $\arg\max_{a\in A}v(a)$ in period t if there exists $\tau<t$ such that $b_\tau\neq b^*$. At histories where there exists no $\tau<t$ such that $b_\tau\neq b^*$ but $a_{t-1}\neq a^*$, player 1 plays a^* if

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

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

and plays the following action if inequality (F.13) is violated:

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

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

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

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

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

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

When $a_{t-1}\neq a^*$ but there is no $\tau<t$ such that $b_\tau\neq b^*$, notice that the left-hand-side of (F.13)

is player 1's continuation value from playing a^*, and the right-hand-side is his continuation value from playing $\widetilde{a} \neq a^*$. This verifies his incentive constraint. When there exists $\tau < t$ such that $b_\tau \neq b^*$, player 2 plays $b(s)$ upon observing s, and it is optimal for player 1 to play $\arg\max_{a \in A} v(a)$.

F.3 Proof of Theorem 3

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

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

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

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

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

1. There exists $\varepsilon > 0$ that depends only on the signal distribution such that $f_0(\beta) \geq \varepsilon f_1(\beta)$ for every monotone β, and $f_0(\beta) \leq (1-\varepsilon) f_1(\beta)$ for every monotone β satisfying $f_0(\beta) < \varepsilon$.[23]

2. For every $f_0 \in [0,1]$, there exists $f_1 \in [0,1]$ such that $(f_0, f_1) \in F$.

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

3. There exists a continuous and strictly increasing function $g : [0,1] \to [0,1]$ with $g(0) = 0$ and $g(1) = 1$ such that $(x, g(x)) \in F$ for every $x \in [0,1]$.

4. There exists $\underline{q} > 0$ such that when player 1 plays $\underline{q}a^* + (1-\underline{q})a'$, $\beta(s) = b'$ for all $s \in S$ is player 2's best reply.

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

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

1. $\alpha(\varnothing) = \underline{q}a^* + (1-\underline{q})a'$, $\phi(\varnothing) = \underline{\beta}$, and $V(\varnothing) = 0$.

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

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

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

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

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

and

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Second, consider the case in which

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

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

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

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

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

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

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

I bound the value of the following expression from below

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

which is at least

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

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

Next, I specify players' strategies at off-path histories and verify that player 1 has no incentive to play any action other than a^* and a'. For every $s_t \notin S'$, player 2 believes that player 1's action is

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

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

which is a strictly increasing function of $f^\dagger(\beta)$. Since $V(h^t, b', a^*) \le -\frac{1-\delta}{\delta} u_1(a^*, b')$, we have $f_1(\beta) \le V(h^t, a)$. Therefore, (F.20) is at most $f_1(\beta)(1-\delta)(u_1(a^\dagger, b^*) - u_1(a^*, b'))$, which is no more than $V(h^t, a)(1-\delta)(u_1(a^\dagger, b^*) - u_1(a^*, b'))$, which is strictly less than $V(h^t, a)$ when δ is close to 1. This implies that player 1 has no incentive to play actions other than a^* and a'.

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

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

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

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

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

References

[1] Bai, Jie (2021) "Melons as Lemons: Asymmetric Information, Consumer Learning and Quality Provision," Working Paper.

[2] Banerjee, Abhijit (1992) "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, 107, 797-817.

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

[4] Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch (1992) "A Theory of Fads, Fashion, Custom, and Cultural Change as Information Cascades," *Journal of Political Economy*, 100, 992-1026.

[5] Board, Simon and Moritz Meyer-ter-Vehn (2020) "Learning Dynamics in Social Networks," Working Paper.

[6] Bose, Subir, Gerhard Orosel, Marco Ottaviani and Lise Vesterlund (2006) "Dynamic Monopoly Pricing and Herding," *RAND Journal of Economics*, 37, 910-928.

[7] Cai, Hongbin, Yuyu Chen and Hanming Fang (2009) "Observational Learning: Evidence from a Randomized Natural Field Experiment," *American Economic Review*, 99, 864-882.

[8] Cai, Jing, Alain De Janvry and Elisabeth Sadoulet (2015) "Social Networks and the Decision to Insure," *American Economic Journal: Applied Economics*, 7, 81-108.

[9] Chan, Jimmy (2000) "On the Non-Existence of Reputation Effects in Two-Person Infinitely-Repeated Games," Working Paper.

[10] Che, Yeon-Koo and Johannes Hörner (2018) "Recommender Systems as Mechanisms for Social Learning," *Quarterly Journal of Economics*, 133, 871-925.

[11] Cripps, Martin and Jonathan Thomas (1997) "Reputation and Perfection in Repeated Common Interest Games," *Games and Economic Behavior*, 28, 433-462.

[12] Deb, Joyee and Yuhta Ishii (2021) "Reputation Building under Uncertain Monitoring," Working Paper.

[13] Deb, Rahul, Matthew Mitchell, and Mallesh Pai (2021) "(Bad) Reputation in Relational Contracting," *Theoretical Economics*, forthcoming.

[14] Drakopoulos, Kimon, Asuman Ozdaglar and John Tsitsiklis (2012) "On Learning with Finite Memory," *IEEE Journal of Selected Topics in Signal Processing*, 6, 780-794.

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

[16] Ely, Jeffrey, Drew Fudenberg and David Levine (2008) "When is Reputation Bad?" *Games and Economic Behavior*, 63, 498-526.

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

[18] Fudenberg, Drew, David Kreps and Eric Maskin (1990) "Repeated Games with Long-Run and Short-Run Players," *Review of Economic Studies*, 57, 555-573.

[19] Fudenberg, Drew and David Levine (1983) "Subgame Perfect Equilibria of Finite and Infinite Horizon Games," *Journal of Economic Theory*, 31, 251-268.

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

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

[22] Gale, Douglas and Shachar Kariv (2003) "Bayesian Learning in Social Networks," *Games and Economic Behavior*, 45, 329-346.

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

[24] Hann-Caruthers, Wade, Vadim Martynov and Omer Tamuz (2018) "The Speed of Sequential Asymptotic Learning," *Journal of Economic Theory*, 173, 383-409.

[25] Harel, Matan, Elchanan Mossel, Philipp Strack and Omer Tamuz (2021) "Rational Groupthink," *Quarterly Journal of Economics*, 136, 621668.

[26] Heller, Yuval and Erik Mohlin (2018) "Observations on Cooperation," *Review of Economic Studies*, 85(4), 2253-2282.

[27] Kartik, Navin, SangMok Lee and Daniel Rappoport (2021) "Observational Learning with Ordered States," Working Paper.

[28] Kaya, Ayça and Santanu Roy (2020) "Market Screening with Limited Records," Working Paper.

[29] Lee, In Ho (1993) "On the Convergence of Information Cascades," *Journal of Economic Theory*, 61, 395-411.

[30] Levine, David (2021) "The Reputation Trap," *Econometrica*, forthcoming.

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

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

[33] Logina, Ekaterina, Georgy Lukyanov and Konstantin Shamruk (2019) "Reputation and Social Learning," Working Paper.

[34] Mailath, George and Larry Samuelson (2015) "Reputations in Repeated Games," *Handbook of Game Theory with Economic Applications*, 165-238.

[35] Michelson, Hope, Anna Fairbairn, Brenna Ellison, Annemie Maertens and Victor Manyong (2021) "Misperceived Quality: Fertilizer in Tanzania," *Journal of Development Economics*, 148, 102579.

[36] Phelan, Christopher (2006) "Public Trust and Government Betrayal," *Journal of Economic Theory*, 130, 27-43.

[37] Rosenberg, Dinah, and Nicolas Vieille (2019) "On the Efficiency of Social Learning," *Econometrica*, 87, 2141-2168.

[38] Schmidt, Klaus (1993) "Reputation and Equilibrium Characterization in Repeated Games with Conflicting Interests," *Econometrica*, 61, 325-351.

[39] Smith, Lones and Peter Norman Sørensen (2000) "Pathological Outcomes of Observational Learning," *Econometrica*, 68, 371-398.

[40] Smith, Lones, Peter Norman Sørensen and Jianrong Tian (2020) "Informational Herding, Optimal Experimentation, and Contrarianism," *Review of Economic Studies*, forthcoming.

[41] Sperisen, Benjamin (2018) "Bounded Memory and Incomplete Information," *Games and Economic Behavior*, 109, 382-400.

[42] Zhang, Juanjuan (2010) "The Sound of Silence: Observational Learning in the U.S. Kidney Market," *Management Science*, 29, 315-335.