Reputation Building under Observational Learning

Harry PEI*

March 30, 2021

A sequence of myopic buyers decide whether to trust a patient seller after observing previous buyers' actions and some private signals about the seller's current and past actions. With positive probability, the seller is a commitment type who plays his optimal commitment action in every period. When each buyer observes all previous buyers' actions and a bounded subset of the seller's past actions, there exist equilibria in which the patient seller receives his minmax payoff since the informativeness of buyers' actions goes to zero as the seller becomes patient. These low-payoff equilibria are robust as long as each buyer has bounded observation of the seller's past actions and can observe the buyer's action in the previous period. When each buyer can also observe an unboundedly informative private signal about the seller's current-period action, the informativeness of buyers' actions is bounded away from zero and a patient seller receives at least his optimal commitment payoff in all equilibria.

Keywords: reputation, reputation failure, speed of learning, social learning.

JEL Codes: C73, D82, D83

1 Introduction

There has been abundant evidence showing that consumers' behaviors are influenced by other consumers' choices (see for example, Cai, Chen and Fang 2009 and Conley and Udry 2010). This is the case in many developing countries where sellers' records are unavailable or incomplete due to the lack of record-keeping institutions, making it time-consuming for consumers to directly learn about sellers' past behavior. For example, a consumer needs to talk to some of the seller's previous customers in order to learn about the quality and attributes of the products they bought from the seller. Due to the time costs of these conversations, information extracted from others' choices is a useful supplement when consumers make their decisions.

I examine a seller's incentives to build good reputations when buyers have limited access to his past records and learn from previous buyers' choices. This question is relevant since the lack of provision

*Department of Economics, Northwestern University. I thank Daron Acemoglu, S. Nageeb Ali, Alp Atakan, Jie Bai, Dhruva Bhaskar, Sushil Bikhchandani, Drew Fudenberg, Olivier Gossner, Kevin He, Ju Hu, Elliot Lipnowski, George Lukyanov, Moritz Meyer-ter-Vehn, Wojciech Olszewski, Mallesh Pai, Evan Sadler, Peter Norman Sørensen, Adam Szeidl, Alex Wolitzky, and my seminar participants for helpful comments. I thank NSF Grant SES-1947021 for financial support.

of high-quality products is a key problem in many developing countries (Rahmat et al. 2016) and observational learning is ubiquitous when consumers do not have access to the seller's past records.

I introduce a new reputation model where each consumer observes *all previous consumers' choices*, *a bounded subset of the seller's past actions*, and possibly, a *private signal about the seller's current-period action*. This stands in contrast to the model of Fudenberg and Levine (1989) where consumers have unbounded observation of the seller's past actions and there is no need for them to learn from other consumers' choices. It also differs from reputation models with limited memories (e.g., Liu 2011, Liu and Skzypacz 2014) in which consumers cannot observe other consumers' choices.

I show that a patient seller's gains from reputations can be wiped out when consumers have no information about his current-period action. This is because the speed of consumer learning is too low relative to the seller's discount rate. These bad equilibria persist as long as each consumer can observe the action profile in the previous period. By contrast, the seller can guarantee high returns from building reputations when each consumer observes an *unboundedly informative private signal* about the seller's *current-period action*. This happens, for example, when a trusted third party randomly inspects a small fraction of products currently sold on the market, issues certificates to the ones with high quality, which are noticed by the current-period consumer but not necessarily by other consumers.

I study a repeated game between a long-lived player 1 (seller) and a sequence of short-lived player 2s (buyers). Players' stage-game payoffs satisfy a monotone-supermodularity assumption, with the product choice game a leading example. Player 1 is either a strategic type who maximizes his discounted payoff, or a commitment type who plays his (pure) Stackelberg action in every period.

My analysis starts from the case in which every buyer observes all previous buyers' actions and the seller's actions in the last K periods. Theorem 1 shows that when the probability of commitment type is below a cutoff, there exist equilibria where the seller receives his minmax payoff no matter how patient he is. This stands in contrast to models where buyers have unbounded observation of the seller's past actions, in which the patient seller receives at least his Stackelberg payoff in *all* equilibria regardless of the probability of commitment type and the observability of previous buyers' actions.

In order to identify the driving forces behind those low-payoff equilibria, I also show that in *all* equilibria, buyers never herd on any action that does not best reply to the commitment action, and the seller's *undiscounted average payoff* from imitating the commitment type is at least a fraction $\frac{K}{K+1}$ of his Stackelberg payoff (plus $\frac{1}{1+K}$ times his minimal stage-game payoff). Therefore, the only plausible explanation for the seller to receive his minmax payoff in some equilibria is that the speed of reputation building is too low relative to the seller's discount rate.

The idea of slow learning is reflected in my constructive proof, which I illustrate using the following product choice game:

–	**T**rust	**N**ot Trust
High Effort	$1, 1$	$-c_N, 0$
Low Effort	$1 + c_T, -1$	$0, 0$

with $c_T, c_N > 0$

Every buyer's action depends only on the action profile in the previous period. A buyer (1) trusts the seller with zero probability when the previous buyer does not trust and the seller exerts low effort, (2) trusts the seller with probability close to zero when the previous buyer does not trust and the seller exerts high effort, (3) trusts the seller with probability less than but close to one when the previous buyer trusts and the seller exerts low effort, and (4) trusts the seller for sure when the previous buyer trusts and the seller exerts high effort. The strategic seller mixes between high and low effort unless the buyer trusts the seller and the seller exerts high effort in the previous period.

Economically, this equilibrium describes a market that has two norms: a good norm in which buyers trust the seller with high probability and a bad norm in which buyers trust the seller with low probability. The market starts from the bad norm, i.e., buyers do not trust new sellers. Transition to the good norm happens when the seller was trusted in the previous period. The transition probability increases with the seller's effort, which motivates the seller to exert high effort.

When the strategic type seller becomes more patient, he is willing to exert high effort and imitate the commitment type even when his action has very low chances of changing the future buyer's action. This slows down consumer learning from previous consumers' actions and increases the expected number of periods to build a reputation. This explains why the seller's gains from a good reputation can be entirely wiped out even when he is arbitrarily patient.

Next, suppose each buyer also observes a *private signal* about the seller's current-period action whose distribution satisfies a monotone likelihood ratio property in addition to what she observes in the previous case. Theorem 2 shows that the patient seller can secure his Stackelberg payoff in *all* equilibria if and (almost) only if the buyer's signal is *unboundedly informative*, i.e., some signal realizations are arbitrarily more likely to occur under the Stackelberg action compared to other actions.

My proof shows that when a buyer's private signal is unboundedly informative, the buyer's action is an informative signal of the seller's current-period action, with informativeness bounded away from zero. When the seller imitates the commitment type, either his stage-game payoff is close to his Stackelberg payoff (when the buyer best replies with probability close to 1), or all future buyers learn at least a certain amount of information about his type from the current-period buyer's action. This explains why the patient seller obtains his Stackelberg payoff in all equilibria.

My paper contributes to the reputation literature by relating a patient player's returns from building reputations to the speed with which his opponents learn from their predecessors' actions. My results examine how the speed of learning varies with the discount factor, and how it hinges on the private signals myopic players receive about the patient player's current-period action.

Theorem 1 identifies a new mechanism behind reputation failures, that the seller's patience endogenously lowers the speed of reputation building. This stands in contrast to most of the existing reputation models (e.g., Fudenberg and Levine 1989, 1992) in which the seller's patience only lowers the cost of reputation building but has negligible impact on its speed.

Theorem 2 shows that reputation failures caused by slow learning can be resolved when a *small* fraction of products currently sold on the market are certified, or when a *small* fraction of consumers can precisely identify the quality of products *before* purchasing. These interventions guarantee a minimal level of informativeness for each buyer's action, and a patient seller receives a high payoff when consumers observe all their predecessors' choices. My result is conceptually different from the one in Smith and Sørensen (2000) since in my model, the myopic players' payoffs depend only on players' actions but not on the patient player's type. Therefore, the myopic players asymptotically learning about the patient player's type is neither sufficient nor necessary for the latter to receive a high discounted average payoff.

In terms of applications, my results provide an explanation for instances of reputation failures and successful policy interventions in developing countries. For example, in the markets for malaria drugs (Nyqvist, et al. 2018) and watermelons (Bai 2018), consumers refuse to pay quality premiums since they believe that sellers are unlikely to supply high quality, and their pessimistic beliefs persist over time. By contrast, the sellers' reputational incentives are restored and consumers are willing to pay quality premiums after researchers randomly assign laser tag machines to a fraction of watermelon vendors (Bai 2018), or after the temporary entry of an NGO that sells high-quality drugs (Nyqvist, et al. 2018). Existing explanations, such as sellers have low discount factors, or buyers receive no information about the seller, or buyers do not purchase from the seller, either do not fit the applications well,[1] or do not provide an appealing rationale for those successful policy interventions.

Section 2 sets up the model. Section 3 states Theorems 1 and 2. Section 4 discusses the robustness of my reputation failure result under alternative specifications of the buyers' information structure, and explains the role of my modeling assumptions. Section 5 concludes and discusses the connections between my paper and the literature on reputation formation and social learning.

[1] For example, Bai (2018)'s structural estimation results in Table 8 show that watermelon vendors' discount factor across interactions is about 0.98. Even before the intervention, residents purchase from the local vendor and exchange information about their experiences, despite they refuse to pay premiums for those so-called high-quality melons.

2 Baseline Model

Time is discrete, indexed by $t = 0, 1....$ 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 the short-lived player in period t.

In period t, player 1 chooses $a_t \in A$ (e.g., his effort or product quality), and then player 2_t chooses $b_t \in B$ (e.g., how many units to buy). Both A and B are finite sets. Player $i \in \{1, 2\}$'s stage-game payoff is $u_i(a_t, b_t)$. Let $\mathrm{BR}_2(a)$ be player 2's best reply to a. Player 1's (pure) Stackelberg action is $\arg\max_{a \in A} \Big\{ \min_{b \in \mathrm{BR}_2(a)} u_1(a, b) \Big\}$.

Assumption 1. *Player 1 has a unique best reply to every $b \in B$. Player 2 has a unique best reply to every $a \in A$. 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 a primary focus of the reputation literature, and fit applications to business transactions (Mailath and Samuelson 2001, Ekmekci 2011, Liu 2011), capital taxation (Phelan 2006), and monetary policy (Barro 1986).

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

1. *Player 1's payoff $u_1(a, b)$ is strictly decreasing in a and is strictly increasing in b.*
2. *Player 2's payoff $u_2(a, b)$ has strictly increasing differences in (a, b).*
3. *Player 1's Stackelberg action a^* is not the lowest element of A.*

In the product choice game, rank the seller's actions according to the quality he supplies and rank the buyer's actions according to the extent to which she trusts the seller (e.g., the number of units she buys), monotone-supermodularity implies that (1) buyers have stronger incentives to trust when the seller supplies higher quality, (2) the seller finds it costly to supply high quality, but strictly benefits from buyers' trust, (3) supplying the lowest quality is not the seller's optimal commitment.

My modeling innovation is on player 2's information structure. I analyze two cases, which differ only in terms of whether player 2_t observes an informative signal about a_t before choosing b_t.

1. **Without Contemporaneous Information:** Player 2_t observes all of her predecessors' actions $(b_0, ..., b_{t-1})$ and player 1's actions in the last K periods $(a_{\max\{0,t-K\}}, ..., a_{t-1})$, where $K \in \mathbb{N}$ is a parameter. Since player 2_t does not observe a_t before choosing b_t, the stage game is equivalent

to one with simultaneous-move. In Section 4, I discuss an alternative setting in which player 2_t observes $(a_{\max\{0,t-K\}}, ..., a_{t-1})$ and $(b_{\max\{0,t-M\}}, ..., b_{t-1})$ where K and M are finite integers.

2. **With Contemporaneous Information:** In addition to observing $(a_{\max\{0,t-K\}}, ..., a_{t-1})$ and $(b_0, ..., b_{t-1})$, player 2_t also privately observes $s_t \in S$, drawn according to $f(\cdot|a_t) \in \Delta(S)$, with S being a countable set. Let $f(s|a)$ be the probability of signal s when player 1's action is a.

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 mechanically plays the Stackelberg action a^* in every period. Let ω_s stand for a *strategic type* who maximizes his discounted average payoff $\sum_{t=0}^{\infty}(1-\delta)\delta^t u_1(a_t, b_t)$, i.e., payoffs are normalized so that the weight on period t is $(1-\delta)\delta^t$. Player 2's prior belief attaches probability $\pi_0 \in (0,1)$ to the commitment type. Let $\mathcal{H}_i$ be the set of player $i \in \{1,2\}$'s histories. Strategic-type player 1's strategy is $\sigma_1 : \mathcal{H}_1 \to \Delta(A)$. Player 2's strategy is $\sigma_2 : \mathcal{H}_2 \to \Delta(B)$. The solution concept is Perfect Bayesian equilibrium (or *equilibrium*).

3 Results

Theorem 1 shows that when player 2 receives no contemporaneous information, player 1 receives his minmax payoff in some equilibria no matter how patient he is. Theorem 2 shows that a patient player 1 can secure his Stackelberg payoff in all equilibria when each player 2 observes an unboundedly informative private signal about player 1's current-period action and all of player 2s' past actions. Section 3.3 provides a unified explanation for these theorems and the existing reputation results.

3.1 Reputation Failure without Contemporaneous Information

Let a' be the lowest element of A. Let $b' \equiv \mathrm{BR}_2(a')$. Let $b^* \equiv \mathrm{BR}_2(a^*)$. Player 1's Stackelberg payoff is $u_1(a^*, b^*)$. The first two requirements of MSM imply that $u_1(a', b')$ is player 1's minmax payoff.[2] The third requirement implies that $a^* \neq a'$ and $u_1(a', b') < u_1(a^*, b^*)$.

Theorem 1. *Suppose u_1 and u_2 satisfy Assumptions 1 and 2, then there exists $\underline{\delta}(u_1, u_2) \in (0,1)$.*[3] *For every $K \in \mathbb{N}$, there exists $\overline{\pi}_0 > 0$, such that for every $\pi_0 < \overline{\pi}_0$ and $\delta > \underline{\delta}(u_1, u_2)$, there exists an equilibrium in which the strategic type player 1's discounted average payoff equals $u_1(a', b')$.*

[2] I adopt the notion of minmax payoff in Fudenberg, Kreps and Maskin (1990) which requires player 2 to play a best reply against some $\alpha \in \Delta(A)$ when she minmaxes player 1.

[3] The cutoff discount factor $\underline{\delta}(u_1, u_2)$ depends only on players' stage-game payoffs and is strictly between 0 and 1. In the product choice game, $\underline{\delta}(u_1, u_2) \equiv \max\{\frac{c_T}{c_T+1}, \frac{c_N}{c_N+1}\}$, i.e., it is large enough such that player 1 has an incentive to play H regardless of player 2's current-period action if doing so increases his continuation value by 1.

Theorem 1 suggests that when the initial trust level between players is low (i.e., π_0 is small) and player 2 has bounded observation of player 1's past actions, player 1 can receive his minmax payoff no matter how patient he is. As will become clear later in the proof, the low-payoff equilibrium is robust to player 2's information structure in the sense that it remains to be an equilibrium as long as player 2_t can observe (a_{t-1}, b_{t-1}), i.e., player 2s do not need to observe calendar time (i.e., when the game started) and unbounded observation of player 2s' past actions is allowed but not required.

The existence of low-payoff equilibria stands in contrast to the reputation result in Fudenberg and Levine (1989), which shows that when player 2s have unbounded observation of player 1's past actions, the patient player receives payoff at least $u_1(a^*, b^*)$ in every equilibrium by playing a^* in every period. This applies regardless of π_0 and whether player 2s' past actions are observed or not.

In my model, unbounded observation of player 1's past actions is replaced by unbounded observation of player 2's past actions together with bounded observation of player 1's past actions. Since the history of player 1's actions is revealed once player 2s aggregate their private information, one may wonder why player 2 observing all her predecessors' actions is not sufficient for player 1 to secure his Stackelberg payoff, or at least some payoff in between his Stackelberg payoff and his minmax payoff.

I argue in two steps that player 1 receiving his minmax payoff from building his reputation is not because his opponents herd on actions lower than b^*, or the signals his opponents receive are uninformative, or his payoff is low in the long run, but instead, it is because the *speed* with which he can build his reputation (or the informativeness of his opponents' actions about his past actions) is endogenous and vanishes to zero as his discount factor approaches unity.

First, I show that in *every* equilibrium, as long as player 1 imitates the commitment type, player 2s cannot herd on any action other than b^* and player 1's *undiscounted average payoff* is at least a fraction $\frac{K}{K+1}$ of his Stackelberg payoff. Formally, let $\pi(h^t)$ be the probability of commitment type at history h^t. For every $b \in B$, player 2s herd on b at h^t if they play b at every $h^s \succeq h^t$. Let $\mathbb{E}^{(a^*,\sigma_2)}[\cdot]$ be the expectation when player 1 plays a^* in every period and player 2 uses strategy σ_2.

Claim 1. *Suppose payoffs satisfy Assumptions 1 and 2, then in every equilibrium* (σ_1, σ_2)

1. *For every $b \neq b^*$, player 2s cannot herd on b at any history h^t that occurs with positive probability under (σ_1, σ_2) satisfying $\pi(h^t) > 0$.*

2. $$\liminf_{t\to\infty} \frac{1}{t}\mathbb{E}^{(a^*,\sigma_2)}\Big[\sum_{s=0}^{t-1} u_1(a_s, b_s)\Big] \geq \frac{K}{K+1}u_1(a^*, b^*) + \frac{1}{K+1}u_1(a^*, b'). \tag{3.1}$$

   When π_0 is small and δ is large, there exists an equilibrium such that (3.1) holds with equality.

The proof is in Appendix C. Intuitively, once player 2s herd on $b \neq b^*$, the strategic type has no intertemporal incentive and hence, will not play a^*. As a result, player 2 knows that player 1 is the commitment type after observing a^*, and will have a strict incentive to play b^*. This contradicts the presumption that they herd on $b \neq b^*$. For the second statement, since a^* is suboptimal for player 1 in the stage game, we know that 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. In the second case, all future player 2s observe an informative signal about a_t, and in expectation, their posterior belief attaches probability close to 1 to the commitment type after a finite number of periods with learning.

However, the above argument does not imply that imitating the commitment type leads to a high *discounted average payoff* since the informativeness of $(b_{t+1}, ..., b_{t+K})$ about a_t is endogenous and may depend on δ. As a result, the expected number of periods needed to establish a reputation may grow with δ and the payoff consequences of these periods cannot be neglected no matter how large δ is.

This stands in contrast to reputation models with imperfect monitoring (e.g., Fudenberg and Levine 1992) where the informativeness of player 2s' signals is bounded away from zero. The logic behind my result also differs from that in Ely and Välimäki (2003) and Ely, et al. (2008). In their models, player 2 can take a non-participating action under which future player 2s receive no informative signal about a_t and the patient player receives a low payoff in the long run, which cannot happen in my model.

A constructive proof of Theorem 1 is in Appendix A. I present a snapshot of my argument and illustrate the economic forces using the product choice game in Section 1.

Proof of Theorem 1: Product Choice Game. Consider the following strategy profile in which player 2_t's strategy depends only on (a_{t-1}, b_{t-1}), which is either $\varnothing$ (i.e., when $t = 0$), (H, T), (H, N), (L, T), or (L, N), and the strategic type player 1's action depends on (a_{t-1}, b_{t-1}) and player 2's belief about the commitment type π_t. Let q be any constant in $(0, 1/2)$.

1. When $(a_{t-1}, b_{t-1}) = (L, N)$ or $\varnothing$, player 2_t plays N for sure and the strategic type player 1 mixes between H and L such that the unconditional probability of H is q, i.e., the probability that the strategic type plays H, denoted by p_t, satisfies $\pi_t + (1 - \pi_t)p_t = q$.

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 N with complementary probability, i.e., $\pi_t + (1 - \pi_t)p_t = 1/2$. The strategic type player 1 mixes between H and L such that the unconditional probability of H is $1/2$, i.e., $\pi_t + (1 - \pi_t)p_t = 1/2$.

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 N with complementary probability. The strategic type player 1 mixes between H and L such that the unconditional probability of H is 1/2, i.e., $\pi_t + (1-\pi_t)p_t = 1/2$.

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

The strategic type player 1's discounted average payoff under this strategy profile is 0 and his continuation value is a function of (a_{t-1}, b_{t-1}), denoted by $V(a_{t-1}, b_{t-1})$, which is given by

$$V(H,T) = 1, \quad V(L,N) = V(\varnothing) = 0, \quad V(H,N) = \frac{1-\delta}{\delta} c_N, \quad V(L,T) = 1 - \frac{1-\delta}{\delta} c_T.$$

I verify players' incentive constraints for every (a_{t-1}, b_{t-1}). Player 2s best reply to their posterior beliefs about player 1's stage-game action, which is the case for every (a_{t-1}, b_{t-1}). For player 1,

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

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

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

and his payoff from playing H is

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

4. When $(a_{t-1}, b_{t-1}) = (H, T)$, player 1's payoff from playing H is 1 and his payoff from playing L is $(1-\delta)(1+c_T) + \delta V(L,T) = 1$.

I show that when $\pi_0 \leq \left(\frac{q}{2}\right)^K \left(\frac{q}{2-q}\right)$, player 2's posterior belief attaches probability no more than

$q/2$ to the commitment type at every history where $(a_{t-1}, b_{t-1}) \neq (H, T)$. This implies that player 1's mixed strategy is well-defined since the strategic type needs to mix only when $(a_{t-1}, b_{t-1}) \neq (H, T)$.

If $a_{t-1} = L$ or $\varnothing$, then player 2's belief attaches probability either 0 or π_0 to the commitment type, which is less than $q/2$. If $(a_{t-1}, b_{t-1}) = (H, N)$, then player 2's belief attaches positive probability to the commitment type if and only if $(a_{\max\{0,t-K\}}, ..., a_{t-1}) = (H, ..., H)$ and $(b_0, ..., b_{t-1}) = (N, N, ..., N)$. Let π_t be the posterior probability of commitment type at such a history.

I show that $\pi_t \leq q/2$ by induction on calender time t. The conclusion true when $t = 0$ since $\pi_0 \leq \left(\frac{q}{2}\right)^K \left(\frac{q}{2-q}\right)$ and $K \geq 1$. Suppose $\pi_s \leq q/2$ for every $s \leq t-1$. Since the unconditional probability with which player 1 plays H is at least $q/2$ and according to the induction hypothesis, $\pi_s \leq q/2$ for every $s \leq t-1$. Therefore, the probability with which the strategic type plays H at each of those histories 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 $(b_0, ..., b_{t-1}) = (N, ..., N)$. According to Bayes rule,

$$\frac{\pi_t}{1-\pi_t} \Big/ \frac{\pi_0}{1-\pi_0} = \frac{P^{\omega_c}(E_t \cap F_t)}{P^{\omega_s}(E_t \cap F_t)} = \frac{P^{\omega_c}(E_t)}{P^{\omega_s}(E_t)} \cdot \frac{P^{\omega_c}(F_t|E_t)}{P^{\omega_s}(F_t|E_t)}. \tag{3.2}$$

Since the strategic type plays H with probability at least $1/4$ in every period before t and N occurs with lower probability under the strategy of the commitment type compared to that of the strategic type, we have

$$\frac{P^{\omega_c}(E_t)}{P^{\omega_s}(E_t)} \leq (q/2)^{-K} \quad \text{and} \quad \frac{P^{\omega_c}(F_t|E_t)}{P^{\omega_s}(F_t|E_t)} \leq 1. \tag{3.3}$$

Since $\pi_0 \leq \left(\frac{q}{2}\right)^K \left(\frac{q}{2-q}\right)$, inequalities (3.2) and (3.3) lead to the conclusion that $\pi_t \leq q/2$. $\square$

The above equilibrium has several attractive features. First, it is robust to alternative specifications of buyers' information structure, which includes for example, each buyer only observes buyers' actions in the last $M \geq 1$ periods and does not directly observe t, i.e., she does not know when the game started. Second, buyers' equilibrium strategy is simple and intuitive, making it plausible to be played in practice. In particular, every buyer's action depends only on players' actions in the previous period but not on more complicated aspects of her private history, and her equilibrium strategy is monotone in the sense that the probability with which she trusts the seller is greater when the previous buyer trusted the seller and the seller exerted high effort.

This equilibrium also exhibits interesting dynamics and highlights the interactions between reputation building, bounded memory, and social learning. Despite b_{t+1} is informative about a_t whenever player 1 receives a low stage-game payoff, its informativeness goes to zero as δ approaches 1. As a result, it takes more time for a more patient player to establish his reputation.

For an economic interpretation of the equilibrium dynamics, consider a market with two social norms: a good norm in which buyers trust the seller with high probability, and a bad one in which buyers trust the seller with low probability. The market initially gets stuck in the bad norm where buyers are unwilling to trust a new seller, and are willing to trust with high probability only when at least one buyer has trusted before. The seller has an incentive to exert high effort since doing so may affect the buyer's future actions after which play transits to a good norm. When the seller is patient, he is willing to invest in his reputation even when doing so affects the buyer's future actions with low probability. This endogenously lowers the informativeness of the buyer's action about the seller's past actions, and slows down the rate at which future player 2s learn about player 1's type. As a result, player 2s remain pessimistic about player 1's type despite observing the latter exerted high effort, and therefore, have no incentive to break away from the bad norm.

3.2 Reputation Result with Contemporaneous Information

I establish a reputation result when player 2_t observes a private signal s_t, distributed according to $f(\cdot|a_t) \in \Delta(S)$, in addition to $(b_0, ..., b_{t-1})$ and $(a_{\max\{0,t-K\}}, ..., a_{t-1})$, before choosing b_t. I restrict attention to signal distributions that satisfy a standard *monotone likelihood ratio property* (MLRP).

MLRP. $\mathbf{f} \equiv \{f(\cdot|a)\}_{a \in A}$ *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* $s \succ_S s'$.

In applications to retail markets where $a \in A$ stands for the quality the seller supplies, MLRP implies that each buyer is more likely to receive a better signal (i.e., a higher s) when the seller supplies higher quality. Whether the patient player can guarantee his Stackelberg payoff in all equilibria hinges on whether player 2's private signal is unboundedly informative about the Stackelberg action a^*.

Unbounded Informativeness. $\mathbf{f}$ *is unboundedly informative about* a^* *if for every* $M > 0$, *there exists* $s \in S$, *such that* $f(s|a^*) > Mf(s|a)$ *for every* $a \neq a^*$.

My notion of unbounded informativeness is similar to that in Smith and Sørensen (2000).[4] When S is a finite set, unbounded informativeness requires the existence of $s^* \in S$ such that $f(s^*|a) > 0$ if

[4] 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.

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 $a \neq a^*$.

Theorem 2 shows that under MLRP, $\mathbf{f}$ being unboundedly informative about a^* is sufficient and almost necessary for player 1 to secure his Stackelberg payoff in all equilibria.

Theorem 2. *Suppose payoffs satisfy Assumptions 1 and 2, and* $\mathbf{f}$ *satisfies MLRP.*

1. *If* $\mathbf{f}$ *is unboundedly informative about* a^**, then for every* $\pi_0 > 0$ *and* $\varepsilon > 0$*, there exists* $\delta^* \in (0,1)$ *such that when* $\delta > \delta^*$*, player* 1*'s payoff is at least* $u_1(a^*, b^*) - \varepsilon$ *in all equilibria.*

2. *If there exists* $\varepsilon > 0$ *such that* $f(s|a') \geq \varepsilon f(s|a^*)$ *for every* $s \in S$*, then 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 in which player* 1*'s payoff is* $u_1(a', b')$*.*

The proof is in Appendix B.1, and a constructive proof for the existence of equilibrium is in Appendix B.2.[5] In games where $|A| = 2$, every $\mathbf{f}$ satisfies MLRP. Since MSM requires that $a^* \neq a'$, when $|A| = 2$ and $\mathbf{f}$ is not unboundedly informative about a^*, there exists $\varepsilon > 0$ such that $f(s|a') \geq \varepsilon f(s|a^*)$ for every $s \in S$. According to Theorem 2, $\mathbf{f}$ being unboundedly informative about a^* is *both necessary and sufficient* for player 1 to secure his Stackelberg payoff in all equilibria. When $|A| \geq 3$, MLRP cannot be dropped and the condition in statement 2 cannot be replaced by $\mathbf{f}$ not being unboundedly informative about a^*. I present a counterexample in Section 4.

The requirement of unboundedly informative private signal in Theorem 2 is reminiscent of the well-known results of Bikhchandani, Hirshleifer and Welch (1992) and Smith and Sørensen (2000), that in canonical social learning models, myopic players' actions are asymptotically efficient if and only if their private signals are unboundedly informative about a persistent exogenous state.

My reputation result is conceptually different from their social learning results since the short-run players asymptotically learn about the persistent state is *neither necessary nor sufficient* for player 1 to receive a high discounted average payoff.[6] It is not sufficient since converging to a high-payoff outcome asymptotically does not imply that player 1 receives a high discounted average payoff. This is demonstrated by my constructive proof of Theorem 1 in which player 1's asymptotic payoff is his Stackelberg payoff while his discounted average payoff equals his minmax value. It is not sufficient

[5] Statement 1 of Theorem 2 only establishes a common property of all equilibria but does not establish the existence of equilibrium. Equilibrium existence does not follow from the result of Fudenberg and Levine (1983) when S is infinite.

[6] My result also differs from Mirrlees (1976) who shows that in principal-agent models, the principal can implement the first best outcome when there exists a signal realization that occurs with zero probability when the agent takes the first-best action and occurs with positive probability otherwise. This is because in my model, the rewards and punishments to player 1 are endogenously determined by player 2's future actions. Depending on the equilibrium being played, there are *multiple ways* in which the signal realizations are mapped to player 1's continuation payoffs.

since player 2 only cares about player 1's action and does not care about player 1's type per se. As a result, in pooling equilibria where the strategic type imitates the commitment type, player 2s cannot learn anything about player 1's type, yet player 1 can still receive his Stackelberg payoff.

The intuition behind Theorem 2 is that player 2_t observing an unboundedly informative private signal about a_t guarantees a uniform lower bound on the speed with which they learn from previous player 2s' actions, which does not depend on player 1's discount factor. Since each player 2 observes all of her predecessors' actions, player 1 receives at least his Stackelberg payoff in all equilibria. By contrast, when the likelihood ratio between the lowest action a' and the Stackelberg action a^* is uniformly bounded from below, the speed of observational learning can vanish to zero as $\delta \to 1$. Similar to the case without contemporaneous information, the prolonged learning process wipes out player 1's gains from reputations and leads to equilibria in which he receives his minmax payoff.

I sketch the proof for my reputation result in two steps, which can explain the role of MLRP and unbounded informativeness. To provide intuition, I focus on the case in which S is finite, i.e., $\mathbf{f}$ is unboundedly informative only when there exists $s^* \in S$ such that $f(s^*|a) > 0$ if and only if $a = a^*$.

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 $\mathbf{f}$ is unboundedly informative about a^*, player 2 has a strict incentive to play b^* when she observes s^*, which rules out the first reason.[7] If the second reason applies, then player 1's stage-game payoff is $u_1(a^*, b^*)$ when he imitates the commitment type.

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

$$\Pr(b_t = b^*|a_t = a^*) - \Pr(b_t = b^*|a_t = a') \geq c(1 - \Pr(b_t = b^*|a_t = a^*)), \tag{3.4}$$

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

When player 1 has three or more actions, player 2_t's incentive to play b^* can no longer be summa-

[7] When S is infinite and $\mathbf{f}$ is unboundedly informative about a^*, there exists a nonempty subset of signal realizations $S(\pi)$ 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 B.1 in Appendix B for details.

rized by a likelihood ratio. As a result, her action can be uninformative about player 1's type even when $\mathbf{f}$ is unboundedly informative about a^* and b_t is informative about s_t. I provide a counterexample in Section 4. Nevertheless, if $\mathbf{f}$ satisfies MLRP, then 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 B.2 of Appendix B that there exists $c > 0$ such that for every $\nu \in (0, 1)$, every α 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 distance between $\gamma(\alpha, \beta)$ and $\gamma(a^*, \beta)$ is at least $c\nu$. This implies that when player 1 imitates the commitment type, either b^* occurs with probability close to 1 under a^* and β, or the informativeness of b_t about player 1's type is uniformly bounded from below.

3.3 Unified Explanation for Theorems 1 and 2

I provide a unified explanation for my two theorems and the canonical reputation results in Fudenberg and Levine (1989, 1992). In the reputation models of Fudenberg and Levine (1989, 1992), Gossner (2011) shows that for any $\delta \in (0, 1)$ and any equilibrium (σ_1, σ_2),

$$\mathbb{E}^{(a^*, \sigma_2)}\Big[\sum_{t=0}^{\infty} d\Big(y_t(\cdot|a^*)\Big|\Big|y_t(\cdot)\Big)\Big] \leq -\log \pi_0, \tag{3.5}$$

where $y_t(\cdot|a^*)$ is the distribution over player 2's signals about a_t when player 1 plays a^*, $y_t(\cdot)$ is the distribution over player 2's signals about a_t in equilibrium, $d(\cdot||\cdot)$ is the KL-divergence, 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.

Inequality (3.5) applies to my model, both with and without contemporaneous information.

In the models of Fudenberg and Levine (1989, 1992) when player 2s' signals can statistically identify player 1's actions, $d\big(y_t(\cdot|a^*)\big|\big|y_t(\cdot)\big)$ is bounded away from 0 at every history where player 2_t does not have a strict incentive to play b^*. Inequality (3.5) implies that the expected number of such "bad periods" is uniformly bounded from above and this upper bound does not depend on δ. Therefore, player 1's expected payoff in every equilibrium is at least $u_1(a^*, b^*)$ when δ is close to 1.

In my constructive proof of Theorem 1, let $y_t(\cdot)$ be the distribution of b_{t+1} and let $y_t(\cdot|a^*)$ be the distribution of b_{t+1} conditional on $a_t = a^*$. It is still true that $d\big(y_t(\cdot|a^*)\big|\big|y_t(\cdot)\big) > 0$ when player 2 does not have a strict incentive to play b^*, but $d\big(y_t(\cdot|a^*)\big|\big|y_t(\cdot)\big)$ vanishes to 0 as δ goes to 1. Therefore, the expected number of periods where player 2 does not have a strict incentive to play b^* can grow without bound, which explains the existence of equilibria where the patient player receives his minmax payoff.

In the case with contemporaneous information, let $y_t(\cdot)$ be the distribution of b_t and let $y_t(\cdot|a^*)$

be the distribution of b_t conditional on $a_t = a^*$. When $\mathbf{f}$ is unboundedly informative, there exists a strictly increasing function $g : [0,1] \rightarrow \mathbb{R}_+$ with $g(0) = 0$ such that $d\big(y_t(\cdot|a^*)\big|\big|y_t(\cdot)\big) > g(\nu)$ at every history where player 2 plays b^* with probability less than $1-\nu$. Inequality (3.5) 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. This implies that a patient player receives at least a fraction $1-\nu$ of his Stackelberg payoff when he imitates the commitment type.

4 Discussions & Extensions

I discuss the robustness of my reputation failure result under alternative information structures and explain the role of MLRP in my reputation result. I also discuss an extension in which every buyer observes a bounded *stochastic* subset of the seller's past actions, in addition to all previous buyers' actions. Some of the claims made in this section are shown in a working paper version (Pei 2020).

Bounded Observation of Both Players' Actions: I consider an extension in which player 2_t observes $(a_{\max\{0,t-K\}}, ..., a_{t-1})$ and $(b_{\max\{0,t-M\}}, ..., b_{t-1})$ where $K \geq 1$ and $M \geq 1$ are finite integers.

The low-payoff equilibrium in the proof of Theorem 1 remains to be an equilibrium since (1) player 2's equilibrium strategy depends only on (a_{t-1}, b_{t-1}), which is feasible as long as $K \geq 1$ and $M \geq 1$, and (2) at every history where the strategic type player 1 needs to play a mixed action, in another word, when $(a_{t-1}, b_{t-1}) \neq (a^*, b^*)$, the probability with which player 2's posterior belief attaches to the commitment type is uniformly bounded from above. For some intuition, take the product choice game example and suppose $(a_{t-K}, ..., a_{t-1}) = (H, ..., H)$ and $(b_{t-M}, ..., b_{t-1}) = (N, ..., N)$. According to my construction, if

$$\Pr\Big((a_{t-K}, ..., a_{t-1}, b_0, ..., b_{t-1})\Big|(a_{t-K}, ..., a_{t-1}) = (H, ..., H) \text{ and } (b_{t-M}, ..., b_{t-1}) = (N, ..., N)\Big) > 0$$

and

$$\Pr\Big(\omega_c\Big|(a_{t-K}, ..., a_{t-1}, b_0, ..., b_{t-1})\Big) > 0,$$

then $(a_{t-K}, ..., a_{t-1}, b_0, ..., b_{t-1}) = (H, ..., H, N, ..., N)$. Since the posterior probability of commitment type is less than $q/2$ after observing $(a_{t-K}, ..., a_{t-1}, b_0, ..., b_{t-1}) = (H, ..., H, N, ..., N)$, we know that the posterior probability of commitment type is also less than $q/2$ after observing $(a_{t-K}, ..., a_{t-1}) = (H, ..., H)$ and $(b_{t-M}, ..., b_{t-1}) = (N, ..., N)$.

However, the conclusion that player 1's undiscounted average payoff from imitating the commitment type is at least $\frac{K}{K+1}u_1(a^*,b^*)+\frac{1}{K+1}u_1(a^*,b')$ in every equilibrium is no longer true. In fact, there exist equilibria in which player 1's undiscounted average payoff is close to his minmax payoff for every finite K and M, which means that reputation failure is not caused by slow learning, but rather, by bounded observation of players' past actions.

Claim 2. *In the product choice game with K and M being strictly positive and finite. For every $\varepsilon > 0$, there exists $\underline{\delta} \in (0,1)$ such that when $\delta > \underline{\delta}$, there exist equilibria in which*

$$\limsup_{t\to\infty} \frac{1}{t}\mathbb{E}^{(a^*,\sigma_2)}\Big[\sum_{s=0}^{t-1} u_1(a_s,b_s)\Big] \leq \varepsilon. \tag{4.1}$$

The proof is in Appendix D, and this conclusion generalizes to all monotone-supermodular games that satisfy a mild regularity condition. Intuitively, each of player 2's action is informative about player 1's past actions, although their informativeness is bounded. As a result, player 2 may never be convinced that player 1 will play H when she can only observe a finite number of past actions. If this is the case, then player 2_t may have an incentive to play N despite $(a_{t-1},b_{t-1})=(H,T)$. However, such an equilibrium will unravel when player 2 observes all of player 2's past actions, since player 2 will eventually be convinced that player 1 will play H in the future after observing a long string of T.

Detrimental Effects of Observing Player 2's Past Actions: I show by example that allowing buyers to observe previous buyers' actions can significantly reduce the seller's equilibrium payoff.

I focus on the product choice game in Section 1 with an additional parametric assumption that $c_N > c_T$. That is, players' actions are strategic complements.[8] I show in Claim 2 that when player 2_t can only observe a_{t-1}, a patient player 1 receives payoff close to 1 in *all* equilibria. This stands in contrast to the conclusion of Theorem 1, in which an arbitrarily patient player 1 receives his minmax payoff in some equilibria when player 2_t can also observe player 2's past actions in at least one period.

An implicit assumption is that player 2 cannot directly observe calendar time. She has a prior belief about t, observes player 1's action in the previous period and updates her belief about t according to Bayes Rule (e.g,. if she observes $a_{t-1}=\varnothing$, then she knows that $t=0$). In order to make player 2's prior belief about calendar time well-defined, I decompose player 1's discount factor δ into two parts.

1. Survival rate: In every period, player 1 survives in the next period with probability $\delta_1 \in (0,1)$, and with complementary probability, he dies or exits the market after which the game ends.

2. Time preference: One unit of utility in period t is worth $\delta_2 \in (0,1)$ unit in period $t-1$.

[8] Liu and Skrzypacz (2014) examine the product choice game under a different assumption that the seller's cost of effort is strictly greater when the buyer trusts the seller, i.e., players' payoffs are submodular.

By definition, $\delta = \delta_1\delta_2$. Under the above interpretation, player 2's prior belief attaches probability $(1-\delta_1)\delta_1^t$ to calendar time being t. I establish the following reputation result.

Claim 3. *Suppose $0 < c_T < c_N$ and each player 2 can only observe player 1's action in the period before, then for every $\pi_0 > 0$, there exists $\underline{\delta} \in (0,1)$ such that when $\delta > \underline{\delta}$,*

1. *There exists at least one equilibrium.*
2. *In every equilibrium, the strategic type player 1's payoff is at least $\delta - (1-\delta)c_N$.*

The proof is in Appendix E. For some intuition, when player 2_t cannot observe b_{t-1}, b_{t-1} affects player 1's incentive in period $t-1$ only through his stage-game payoff. The strategic complementarity between actions implies that player 1 has a stronger incentive to play H when player 2 plays T with higher probability. Therefore, whenever the strategic type has an incentive to impede learning by playing H when $a_{t-1} = L$, he must have a strict incentive to do so when $a_{t-1} = H$. If this is the case, then the buyer has a strict incentive to trust him whenever $a_{t-1} = H$ and a patient seller can guarantee his optimal commitment payoff by imitating the commitment type.

When player 2_t observes b_{t-1}, b_{t-1} can also affect player 1's continuation value. This weakens the implication of strategic complementarity in the stage game and leads to equilibria in which player 2_t's action is more responsive to player 2_{t-1}'s action rather than player 1's past actions. The strategic type player 1 always has an incentive to imitate the commitment type, although this incentive is weak and therefore, is insufficient to provide player 2 a strict incentive to play T.

The Role of MLRP: MLRP ensures that the informativeness of b_t about player 1's type is bounded away from 0 whenever the probability with which $b_t = b^*$ is bounded away from 1.

In order to demonstrate that MLRP is not redundant for my reputation result, I provide an example in which players' stage-game payoffs satisfy Assumptions 1 and 2, the signal distribution is unboundedly informative about a^*, but violates MLRP. I show that a patient player 1's payoff is bounded below his optimal commitment payoff in some equilibria. Players' 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 $f(s^*|a^*) = 2/3$, $f(\underline{s}|a^*) = 1/3$, $f(\overline{s}|\overline{a}) = 1$, $f(\overline{s}|\underline{a}) = 1/3$, and $f(\underline{s}|\underline{a}) = 2/3$.

One can verify that players' 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 $b^* \succ b'$. Player 1's Stackelberg action is a^*, his Stackelberg payoff is 2, $\mathbf{f}$ is unboundedly informative about a^*, but violates MLRP.

I construct an equilibrium in which player 1's payoff is 1, which is strictly bounded below his Stackelberg payoff. 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 $\alpha(\pi) \in \Delta(A)$, which is pinned down by:

$$(1-\pi) \circ \alpha(\pi) + \pi \circ a^* = 0.5 \circ a^* + 0.25 \circ \overline{a} + 0.25 \circ \underline{a}.$$

Player 2 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 action in the current period. Player 2 has a strict incentive to play b^* after observing $\overline{s}$ or s^*, and has an incentive to play b' after observing $\underline{s}$. Conditional on each type of player 1, the probability with which player 2 plays b^* 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.

Bounded Informativeness: I use the following example to explain why the condition in statement 2 cannot be replaced by a weaker bounded informativeness condition. Players' 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 $\overline{s} \succ s^* \succ \underline{s}$, $\mathbf{f}$ satisfies MLRP, and is not unboundedly informative about a^*.

Player 1 receives at least his Stackelberg payoff 2 in every equilibrium. This is because when player 1 plays a^*, player 2 observes either s^* or $\overline{s}$, and has a strict incentive to play b^*.

Games with Stochastic Network Monitoring: In the baseline model, each buyer observes the seller's actions in the last K periods.

In many applications of interest, each consumer observes a stochastic subset of the seller's past actions. For example, each buyer randomly samples previous buyers and asks the buyers in her sample about their personal experiences with the seller. Alternatively, every buyer may only talk to her friends, modeled as her neighbors in a stochastic network, about the seller's actions against them. In both scenarios, the seller does not know who does each buyer sample and who each buyer's friends are. This leads to private monitoring about the seller's past actions and private learning about the seller's type. Both of these features bring new challenges to conduct equilibrium analysis.

I study an extension in a working paper version (Pei 2020) where every buyer observes the entire history of previous buyers' actions as well as the seller's past actions according to a stochastic network among the buyers $\mathcal{N} \equiv \{\mathcal{N}_t\}_{t=1}^{\infty}$, with $\mathcal{N}_t \in \Delta\big(2^{\{0,1,...,t-1\}}\big)$. The realization of $\mathcal{N}_t$ is denoted by $N_t \subset \{0, 1, ..., t-1\}$, which is privately observed by player 2_t and is unbeknownst to player 1 and other short-run players. In the case without contemporaneous information, player 2_t observes N_t, $\{a_\tau\}_{\tau \in N_t}$, $\{b_0, ..., b_{t-1}\}$, and ξ_t. In the case with contemporaneous information, she observes N_t, $\{a_\tau\}_{\tau \in N_t}$, $\{b_0, ..., b_{t-1}\}$, ξ_t, and $s_t \in S$ drawn according to $f(\cdot|a_t) \in \Delta(S)$.

Statement 1 of Theorem 2 applies regardless of the stochastic network $\mathcal{N}$, since b_t is informative about a_t whenever $\Pr(b_t = b^*) < 1$. Theorem 1 and Statement 2 of Theorem 2 extend when $\mathcal{N}$ is such that (1) for every $s \neq t$, $\mathcal{N}_s$ and $\mathcal{N}_t$ are independent random variables, which is commonly assumed in the social learning literature including Acemoglu, Dahleh, Lobel and Ozdaglar (2011). (2) there exist $K \in \mathbb{N}$ and $\gamma \in (0, 1)$ such that $\Pr\big(|\mathcal{N}_t| \leq K\big) = 1$ and $\Pr\big(t-1 \in \mathcal{N}_t\big) \geq \gamma$ for every $t \geq 1$, i.e., every buyer observes a bounded stochastic subset of the seller's past actions and observes the seller's action against her immediate predecessor with probability bounded away from zero.

The proof of the reputation failure result shares similar ideas with the proof of Theorem 1. In particular, I construct an equilibrium that starts from a non-trusting phase and followed by a trusting phase and a punishment phase. The main challenge is to construct the non-trusting phase under private monitoring and private learning. The belief-free approach in the existing literature does not directly apply, since player 2s are myopic in my model. In equilibria where nontrivial learning takes place, player 2s' actions are sensitive to their posterior beliefs about player 1's type, making it hard to sustain belief-free incentives.

To illustrate how I construct the non-trusting phase, consider the product choice game earlier in this section. Let q^* be the minimal probability with which H needs to be played in order to provide player 2 an incentive to play T. Let $\{a_0, ..., a_{t-1}, b_0, ..., b_{t-1}\}$ be a *complete history* in period t.

First, consider providing belief-free incentives such that (1) conditional on each *complete history*, player 2 believes that H will be played with probability q^*, and (2) each player 2 mixes between T and N with probability that makes player 1 indifferent between H and L. Under this arrangement, both T and N are player 2's best replies, regardless of her belief about player 1's type and private history.

However, this belief-free arrangement is feasible in period t only if after observing *each* complete history in period t, player 2's posterior belief attaches probability less than q^* to the commitment type. Since player 2 plays N in the non-trusting phase, the probability with which player 1 plays H is at most q^*. Therefore, a hypothetical observer's posterior belief attaches probability arbitrarily close to 1 to the commitment type after observing a sufficiently long string of H. This implies the existence of a cutoff calendar time, such that the above belief-free arrangement is feasible only if calendar time is below this cutoff.

In light of this observation, I use the following *belief-based construction* when calendar time is above the aforementioned cutoff. In particular, player 1's action depends on his private history, which is chosen such that each player 2 is indifferent under her posterior belief about player 1's private history. This is equivalent to establish the existence of solution to a system of linear equations, in which the number of player 1's private histories is the number of free variables, and the number of player 2's private histories is the number of linear constraints. In period t, the number of free variables is 2^t. Given that each player 2's sample size is at most K, the number of constraints is at most $2^K \sum_{j=0}^{K} \binom{t}{j}$. An important observation is that the linear system is under-determined *if and only if* t is large relative to K. This explains why the belief-free construction is used only when calendar time is below the cutoff, and the belief-based construction is used only when calendar time is above the cutoff.

5 Concluding Remarks

I examine a patient player's returns from investing in his reputation when his opponents have limited access to his past records and learn primarily from previous short-run players' actions.

My results relate the patient player's returns from good reputations to the speed with which myopic players learn from their predecessors' actions, which hinges on the private information each myopic player receives about the patient player's current-period action. When myopic players have no information about the patient player's current-period action or can only observe a boundedly informative signal, the speed of observational learning can vanish to zero as the patient player's discount factor approaches unity, leading to equilibria in which the patient player receives his minmax payoff. When every myopic player privately observes an unboundedly informative signal about the patient player's current-period action, the patient player receives at least his optimal commitment payoff in all equilibria since the speed of observational learning is uniformly bounded away from zero.

I construct a class of low-payoff equilibria in which every buyer's action is more responsive to the buyer's past actions compared to the seller's past actions. Those equilibria highlight the interaction between social learning and reputation building, and demonstrate the fragility of reputation effects when players have bounded memory or limited capacity to process detailed information. They also demonstrate the importance of high-quality record-keeping institutions and high-quality inspection technologies in encouraging sellers to build good reputations. In particular, observational learning from buyers' past actions is a powerful tool in motivating sellers to supply high quality once complemented with a minimal frequency of product certification.

I conclude by explaining the connections between this paper and the existing literature on reputation failure, reputation models with bounded memories, and social learning.

Reputation Failures: My Theorem 1 identifies a new mechanism for why reputation effects may fail despite the reputation-building player is patient and his opponents receive informative signals about his behavior. This stands in contrast to existing theories that are based on the lack-of-identification (e.g., Ely and Välimäki 2003) or the uninformed player being patient (e.g., Cripps and Thomas 1997).

Models with lack-of identification such as Ely and Välimäki (2003), Ely, Fudenberg and Levine (2008), and more recently Deb, Mitchell and Pai (2020) focus on *participation games* where the uninformed player(s) can take a non-participating action under which future uninformed players cannot receive informative signals about the informed player's current-period action.[9] The uninformed players' option to shut down learning leads to equilibria with low asymptotic payoffs. My model stands in contrast to theirs since no action of the uninformed player can stop her successors' learning. Despite the informed player is guaranteed to receive a high asymptotic payoff by building his reputation, his discounted average payoff is low since the speed of learning depends endogenously on his patience.

Schmidt (1993), Cripps and Thomas (1997), and Chan (2000) construct low-payoff equilibria when the uninformed player is patient and can observe the entire history of the informed player's actions. The takeaway message from their papers and other reputation models with complete records (e.g., Fudenberg and Levine 1989) is that the informed player's patience helps reputation building while the uninformed player's patience hurts reputation building. In particular, the cost of reputation building is lower when the informed player becomes more patient. The takeaway message from my analysis is entirely the opposite, that the informed player's patience lowers the rate of learning, prolongs the

[9] Levine (2019) examines a model where signals are less informative when the uninformed players do not participate.

process of reputation building, which can exactly cancel out the positive effects of patience.

Reputation Models with Limited Memory: Existing reputation models with bounded memories such as Liu (2011), Liu and Skrzypacz (2014), and Kaya and Roy (2020) study situations in which every short-run player observes a bounded sequence of the long-run player's past actions, but *cannot* observe other short-run players' actions.

This modeling difference between my paper and theirs leads to different insights. My model with unboundedly informative contemporaneous information implies that consumers' observational learning can provide powerful incentives for sellers to build reputations in *all* equilibria once complemented with a minimal frequency of product certification. In my model without contemporaneous information, player 1 can guarantee a high asymptotic payoff from building his reputation in *all* equilibria, which implies that slow learning is the only plausible driving force behind reputation failures.

By contrast, in the models of Liu (2011) and Liu and Skrzypacz (2014) when the prior probability of commitment type is below a cutoff, both the patient player's discounted average payoff and asymptotic payoff are bounded below his Stackelberg payoff in all Markov equilibria. This implies that my positive reputation results hinge on unbounded observation of previous consumers' actions. Furthermore, the reputation cycles in their models rely on the assumption that the seller can unilaterally clean up his records, which is not feasible in my model when consumers can observe previous consumers' choices.

Social Learning: My model can be viewed as a social learning game in which a sequence of myopic players observe their predecessors' actions and some private signals (e.g., the long-run player's actions in the last K periods) in order to forecast the current behavior of a strategic long-run player.

The object to learn differs from the social learning models of Banerjee (1992), Bikhchandani, et al. (1992), and Smith and Sørensen (2000) in which myopic players learn about an exogenous state.[10] As a result, the myopic players asymptotically learn the state (i.e., the patient player's type) is neither sufficient nor necessary for the patient player to receive his Stackelberg payoff.

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 (Banerjee 1992 and so on), their asymptotic rates of learning (Gale and Kariv 2003, Hann-Caruthers, Martynov and Tamuz 2018, Harel, Mossel, Strack and Tamuz 2020), and their asymptotic payoffs

[10] 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. 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.

(Rosenberg and Vieille 2019).[11] As highlighted by the comparison between Theorem 1 and Claim 1, the patient player's discounted average payoff can be low despite their asymptotic payoff is high.

In my model, the patient player receives at least his Stackelberg payoff in every equilibrium when each myopic player directly observes her predecessors' private signals. This is analogous to Banerjee (1992), Bikhchandani, et al. (1992), and Smith and Sørensen (2000) in which myopic players' actions are asymptotically efficient when they can directly observe their predecessors' private signals.

However, different forms of inefficiencies arise when the myopic players can only observe their predecessors' actions but not their private signals. In canonical social learning models, the asymptotic outcome is inefficient since players can herd on an inefficient action. In *every* equilibrium of my model *without* contemporaneous information, the myopic players never herd on any action other than b^*, and inefficiencies take the form of low discounted average payoff despite player 1 can guarantee a high asymptotic payoff by building his reputation.

My paper is also related to the literature on social learning with bounded memory. Drakopoulos et al (2012) examine a setting in which every myopic player learns about an exogenous state by observing some private signals and the actions of his last K immediate predecessors. They show that learning is possible when $K \geq 2$ but not when $K = 1$. By contrast, the myopic players learn about the behavior of a strategic long-run player in my model and the informativeness of their private signal (i.e., the patient player's past actions) is also endogenous. In contrast to the conclusion in Drakopoulos et al (2012), the length of memory does not play a key role in my result as long as it is strictly positive.

[11] A separate strand of works characterize mechanisms that maximize *a sequence of myopic players' discounted average payoff* in social learning games, which include Che and Hörner (2018) and Smith, Sørensen and Tian (2021).

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 by setting $u_1(a',b') = 0$ and $u_1(a^*,b^*) = 1$.

Let $\underline{q}$ be the largest $q \in [0,1]$ such that b' is not player 2's strict best reply to $qa^* + (1-q)a'$. Let $\overline{q}$ be the smallest $q \in [0,1]$ such that b^* is not player 2's strict best reply to $qa^* + (1-q)a'$. 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 $\{b^{**}, b'\} \subset \mathrm{BR}_2(\underline{q}a^* + (1-\underline{q})a')$ and $\{b^*, b''\} \subset \mathrm{BR}_2(\overline{q}a^* + (1-\overline{q})a')$. The MSM condition (i.e., Assumption 2) implies that $b^* \succ b''$, $b^{**} \succ b'$, and $b^* \succ b'$. This implies that there are three possibilities, which I consider separately in the proof.

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

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

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

I construct equilibrium in which (1) player 1's payoff is 0, (2) player 2_t's action depends only on (a_{t-1}, b_{t-1}) and 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, (3) player 1 plays a^* or a' on the equilibrium path, (4) 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 $u_1(a,b)$ strictly decreases in a and a' is player 1's lowest action, he strictly prefers a' to actions other than a^* and a' at any history.

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' such that player 2 believes that a^* is played with probability 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' such that player 2 believes that a^* is played with probability 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' such that player 2 believes that a^* is played with probability q.

4. When $(a_{t-1}, b_{t-1}) = (a', b^*)$, player 2 plays b^* and the strategic type 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 $(a_{t-1}, b_{t-1}) \neq (H, T)$ 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 belong to $(0, 1)$ and will be specified later on.

(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' such that a^* is played with probability $\underline{q}$ under player 2's posterior.

(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' such that a^* is played with probability $\underline{q}$ under player 2's posterior.

(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' such that a^* is played with probability $\underline{q}$ under player 2's posterior.

(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' such that a^* is played with probability $\overline{q}$ under player 2's posterior.

(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' such that a^* is played with probability $\overline{q}$ under player 2's posterior.

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

Next, I compute player 1's continuation values and verify his incentive constraints. From (1) and (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 $0 < V(a^*, b') < (1-\delta)u_1(a^*, b^{**}) + \delta V(a^*, b^{**})$.

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

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

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

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

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 strategy 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' such that a^* is played with probability $\underline{q}$ under player 2's posterior.

(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' such that a^* is played with probability $\underline{q}$ under player 2's posterior.

(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' such that a^* is played with probability $\underline{q}$ under player 2's posterior.

(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' such that a^* is played with probability $\overline{q}$ under player 2's posterior.

(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' such that a^* is played with probability $\overline{q}$ under player 2's posterior.

(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, when $u_1(a', , b'') \leq 1$, 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, when $u_1(a', b'') > 1$, by setting $\gamma = 1$, the LHS of (A.9) is strictly positive while the RHS 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 $u_1(a^*, b'')$ $\square$

Pick γ that satisfies the condition in Lemma A.1, 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 $0 < V(a^*, b') < (1-\delta)u_1(a^*, b'') + \delta V(a^*, b^{''})$.

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 $0 < V(a', b'') < (1-\delta)u_1(a', b'') + \delta V(a', b'')$.

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

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

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). $\square$

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,

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

The first term on the RHS 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 RHS 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.

B Proof of Theorem 2

Section B.1 establishes a payoff lower bound that applies to all equilibria when $\mathbf{f}$ is unboundedly informative. Section B.2 establishes the existence of equilibrium when $\mathbf{f}$ is unboundedly informative, S is countablely infinite, and δ is large enough. The existence of equilibrium when S is finite follows from the standard argument in Fudenberg and Levine (1983).

B.1 Proof of Statement 1

I start from a lemma showing that in every equilibrium, if player 1 plays a^* in every period, then there exists $\eta > 0$ that depends only on $\mathbf{f}$ 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 B.1. *Suppose* $\mathbf{f}$ *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)*,*

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

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 $\mathbf{f}$ is unboundedly informative, $S(\pi^*)$ is non-empty and $f(s|a^*) > 0$ for every $s \in S(\pi^*)$. Therefore, we have $\eta > 0$.

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 $\{\pi_t \geq \pi^*$ 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 $\{\Pr(b_t = b^*) \geq \eta$ 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 B.2. *Suppose* $\mathbf{f}$ *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. I omit the subscripts in the complete orders on S, A, and B and write $\succ$ instead. Since $u_2(a, b)$ has strictly increasing differences and $\mathbf{f}$ 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 $\mathbf{f}$ 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

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

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 $a\neq a^*$,

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

Since $\mathbf{f}$ satisfies MLRP,

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

These inequalities together with (B.2) imply that

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

and

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

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

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

First, I show that the LHS of (B.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 LHS of (B.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 LHS of (B.6) is $\eta(1 - \frac{1}{M(\pi^*)})$, which is greater than $\eta/2$ when $M(\pi^*)$ is large enough.

3. If the limit of S^* is unbounded from below, then similarly, the LHS of (B.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 $\{b^*\} = \mathrm{BR}_2(a^*)$, we know that for every $b \succ b^*$, there exists $\overline{r}^* \in \mathbb{R}_+$ such that $b \in \mathrm{BR}_2(\alpha)$ 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 $b \in \mathrm{BR}_2(\alpha)$ 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 (B.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 B.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 B.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 P^{σ,ω_c} be the probability measure induced by σ conditional on player 1 being the commitment type, and let P^{σ,ω_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

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

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

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

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

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

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

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 P^σ, the probability of $(a_{\max\{0,t-K\}}, ..., a_{t-1}) = (a^*, ..., a^*)$ conditional on $(b_0, ..., b_{t-1})$ is at least $\rho^*(\epsilon)$.

Suppose towards a 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 (B.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 $P(\sigma, t, s) \in \Delta(\Delta(A^K))$ 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 $P(\sigma, t)$ 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:

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

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:

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

Applying (B.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:

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

Let $\epsilon\to 0$ and $\delta \to 1$, (B.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.

B.2 Existence of Equilibrium when f is Unboundedly Informative

I establish the existence of Perfect Bayesian equilibrium when $\mathbf{f}$ 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

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

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

and plays

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

if inequality (B.13) is violated.

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

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

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

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, (B.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 LHS of (B.13) is player 1's continuation value from playing a^*, and the RHS is his continuation value from playing $\widetilde{a} \neq a^*$. This verifies his incentive constraint. When there exists $\tau < t$ such that $b_\tau \neq b^*$, player 2 plays $b(s)$ upon observing s, and it is optimal for player 1 to play $\arg\max_{a \in A} v(a)$.

C Proof of Claim 1

I establish the no herding result in Section C.1, which only uses Assumption 1 but not Assumption 2. I establish the lower bound on player 1's undiscounted average payoff in Section C.2. and construct equilibria in which player 1's asymptotic payoff equals the right-hand-side of (3.1) in Section C.3.

C.1 Proof of No Herding Result

Suppose toward a contradiction that player 2s herd on $b \neq b^*$ at h^t, then the strategic type has no intertemporal incentive at h^t and at every h^t_* that differs from h^t only in $\{a_0, ..., a_{t-K}\}$. In equilibrium, strategic-type player 1 plays his myopic best reply to b at those histories. Consider two cases. First, suppose $\mathrm{BR}_1(b) = \{a^*\}$, then in equilibrium, both types of player 1 play a^* at h^t and at every h^t_* that differs from h^t only in $\{a_0, ..., a_{t-K}\}$. As a result, player 2_t has a strict incentive to play b^* instead of b at h^t. This contradicts the presumption that $b \neq b^*$. Second, suppose $\mathrm{BR}_1(b) \neq \{a^*\}$, then in equilibrium, the strategic type has no incentive to play a^* at h^t and at every h^t_* that differs from h^t only in $\{a_0, ..., a_{t-K}\}$. Since $\pi(h^t) > 0$, player 2_{t+1}'s belief 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 $b \neq b^*$.

C.2 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 $a \prec a^*$, MSM 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

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

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

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

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 (C.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:

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

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

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

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

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

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

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

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 (C.3) and (C.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 $\tau \neq \tau'$,

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

The former is straightforward. For the latter, suppose toward a 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 (C.3) implies that player 1's expected average payoff at histories in $\mathcal{H}_0^\tau$ is at least the RHS of (3.1). Since $\mathcal{H}_0^\tau \cap \mathcal{H}_0^{\tau'} = \{\varnothing\}$ for every $\tau \neq \tau'$, it implies that player 1's expected average payoff at histories in $\bigcup_{\tau=0}^{K} \mathcal{H}_0^\tau$ is at least the RHS of (3.1). For every $\tau \in \{0, 1, ..., K\}$, (C.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

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

C.3 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 time-average payoff from playing a^* in every period equals the RHS of (3.1). I verify players' incentive constraints. Since b^* best replies to a^* and b' best replies to a', player 2's incentive constraints are satisfied. In what follows, I verify player 1's incentives. At every on-path history h^t,

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

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

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.

D Asymptotic Payoff under Bounded Memory

I state and show a general result on player 1's asymptotic payoff when he imitates the commitment type in games where player 2_t observes player 1's actions in the last K periods and player 2's actions in the last M periods, where K and M are finite integers. This result implies Claim 2.

D.1 Statement of General Result

Since players' payoffs are monotone-supermodular, $u_1(a^*,b^*) > u_1(a',b')$. Without loss of generality, I normalize player 1's payoff so that $u_1(a',b')=0$ and $u_1(a^*,b^*)=1$. Let $\underline{q}$ be the largest $q \in [0,1]$ such that b' is not player 2's strict best reply to $qa^*+(1-q)a'$. Let $\overline{q}$ be the smallest $q \in [0,1]$ such that b^* is not player 2's strict best reply to $qa^*+(1-q)a'$. Since b' is a strict best reply to a' and b^* is a strict best reply to a^*, there exist $b^{**} \neq b'$ and $b'' \neq b^*$ such that $\{b^{**}, b'\} \subset \mathrm{BR}_2(\underline{q}a^*+(1-\underline{q})a')$

and $\{b^*, b''\} \subset \mathrm{BR}_2(\overline{q}a^* + (1-\overline{q})a')$. My monotone-supermodularity assumption implies that either $b^* = b^{**}$ and $b' = b''$, or $b^* \succ b'' \succ b^{**} \succ b'$, or $b^* \succ b'' = b^{**} \succ b'$.

Definition 1. *Players' payoffs are irregular if* $b'' = b^{**}$ *and* $u_1(a^*, b^{**}) < -1$*. Otherwise, players' payoffs are regular.*

Players' payoffs are regular in the product choice game in which $b^* = b^{**} \succ b'' = b'$ and in more general monotone-supermodular games in which player 1's cost of playing a^* is not too large when player 2 plays actions in between b' and b^*. An example of an irregular game is shown as follows:

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

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

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

The rest of this section shows this proposition. I consider three cases separately, depending on the value of $u_1(a^*, b^{**})$ and the comparison between b^*, b^{**}, b', and b''.

D.2 The Case in which $u_1(a^*, b^{**}) > 0$

Consider the following strategy profile in which player 2_t's action depends only on (a_{t-1}, b_{t-1}), and the strategic type player 1 mixes between a^* and a' with probabilities such that player 2_t is indifferent between b^{**} and b'. I verify later than there exist such mixing probabilities by bounding the posterior probability of commitment type from above.

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

Let

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

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

$$V(a,b^{**})=\frac{X-(1-\delta)u_1(a,b^{**})}{\delta} \text{ for every } a\in\{a',a^*\}. \tag{D.3}$$

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

$$r(a,b)=\frac{V(a,b)}{X}. \tag{D.4}$$

I verify that $V(a,b)\leq X$ so that $r(a,b)$ is a well-defined probability. First, $V(a^*,b')\leq X$ by definition. This is because $u_1(a',b^{**})>u_1(a^*,b^{**})$ and (D.3). Second, I show that $V(a',b^{**})<V(a^*,b^{**})\leq X$, which is equivalent to

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

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

According to the construction of these continuation values, we have

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

and

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

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

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

$$L\equiv\min\left\{\frac{r(a^*,b')}{r(a^*,b^{**})},\frac{r(a^*,b^{**})}{r(a^*,b')}\right\}. \tag{D.5}$$

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

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

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

$$\frac{\overline{\pi}_0}{1-\overline{\pi}_0}=\frac{q^*/2}{1-q^*/2}\Big(\frac{q^*}{2}\Big)^{K+M+1}L^M \tag{D.6}$$

I show by induction that $\pi_t \leq q^*/2$ for every $t \in \mathbb{N}$ if $\pi_0 < \overline{\pi}_0$. Without loss of generality, we only need to consider histories where $(a_{t-K},...,a_{t-1})=(a^*,...,a^*)$ and $(b_{t-M},...,b_{t-1})=(b^{**},...,b^{**})$. First, condition (D.6) implies that $\pi_0 \leq q^*/2$. Second, suppose $\pi_s \leq q^*/2$ for every $s \leq t-1$, then the strategic type plays a^* with probability at least $q^*/2$ at every history from period 0 to period $t-1$. Recall from Section 3 that P^{ω_c} is the probability measure induced by the commitment type and P^{ω_s} is the probability measure induced by the strategic type. According to Bayes rule, we have

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

According to the induction hypothesis,

$$\frac{P^{\omega_c}(a_{t-i}=a^*|a_{t-i+1},...,a_{t-1})}{P^{\omega_s}(a_{t-i}=a^*|a_{t-i+1},...,a_{t-1})}\leq (q^*/2)^{-1}. \tag{D.7}$$

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

$$\frac{P^{\omega_c}(b_{t-i}=b^{**}|a_{t-i+1},...,a_{t-1},b_{t-i+1},...,b_{t-1})}{P^{\omega_s}(b_{t-i}=b^{**}|a_{t-i+1},...,a_{t-1},b_{t-i+1},...,b_{t-1})}\leq (q^*/2)^{-1}L^{-1} \tag{D.8}$$

Plugging inequalities (D.7) and (D.8) into the expression for π_t, we have $\pi_t \leq q^*/2$.

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

$$\frac{r(a^*,b')}{1-r(a^*,b^{**})+r(a^*,b')}u_1(a^*,b^{**})+\frac{1-r(a^*,b^{**})}{1-r(a^*,b^{**})+r(a^*,b')}u_1(a^*,b'). \tag{D.9}$$

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

D.3 The Case in which $b^* \succ b'' \succ b^{**} \succ b'$

Consider the following strategy profile:

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

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

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

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

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

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

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

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

$$0 = (1-\delta)u_1(a^*, b') + \delta V(a^*, b') = (1-\delta)u_1(a', b') + \delta V(a', b'). \tag{D.10}$$

$$X(b^{**}) \equiv (1-\delta)u_1(a^*, b^{**}) + \delta V(a^*, b^{**}) = (1-\delta)u_1(a', b^{**}) + \delta V(a', b^{**}). \tag{D.11}$$

$$Y(b'') \equiv (1-\delta)u_1(a^*, b'') + \delta V(a^*, b'') = (1-\delta)u_1(a', b^{''}) + \delta V(a', b''). \tag{D.12}$$

$$Y(b^*) \equiv (1-\delta)u_1(a^*, b^*) + \delta V(a^*, b^*) = (1-\delta)u_1(a', b^*) + \delta V(a', b^*), \tag{D.13}$$

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

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

and

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

In order to deliver these continuation values, we need

$$Y(b'') \leq V(a^*, b^{**}) = \frac{X(b^{**}) - (1-\delta)u_1(a^*, b^{**})}{\delta} \leq Y(b^*) \tag{D.14}$$

$$0 \leq V(a^*, b'') = \frac{Y(b'') - (1-\delta)u_1(a^*, b'')}{\delta} \leq X(b^{**}) \tag{D.15}$$

$$Y(b'') \leq V(a^*, b^*) = \frac{Y(b^*) - (1-\delta)u_1(a^*, b^*)}{\delta} \leq Y(b^*), \tag{D.16}$$

$$Y(b'') \leq V(a', b^*) = \frac{Y(b^*) - (1-\delta)u_1(a', b^*)}{\delta} \leq Y(b^*), \tag{D.17}$$

$$0 \leq V(a', b^{**}) = \frac{X(b^{**}) - (1-\delta)u_1(a', b^{**})}{\delta} \leq X(b^{**}), \tag{D.18}$$

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

Let

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

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

$$\frac{\overline{\pi}_0}{1-\overline{\pi}_0} = \frac{\underline{q}/2}{1-\underline{q}/2}\Big(\frac{\underline{q}}{2}\Big)^{K+M+1} L^M. \tag{D.19}$$

The same argument as that in Section D.1 implies that player 2's posterior belief attaches probability less than $\underline{q}/2$ to the commitment type if her prior belief satisfies $\pi_0 \leq \overline{\pi}_0$.

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

D.4 The Case in which $b^* \succ b'' = b^{**} \succ b'$ and $u_1(a^*, b^{**}) \in (-1, 0]$

Consider the following strategy profile:

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

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

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

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

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

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

Let

$$X \equiv \max\Big\{(1-\delta)u_1(a', b^{**}), -\frac{1-\delta}{\delta}u_1(a^*, b')\Big\}, \tag{D.20}$$

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

$$\frac{X - (1-\delta)u_1(a^*, b^{**})}{\delta} < Y < (1-\delta)u_1(a^*, b^*) + \delta X. \tag{D.21}$$

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

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

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

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

$$Y = (1-\delta)u_1(a', b^*) + \delta V(a', b^*) = (1-\delta)u_1(a^*, b^*) + \delta V(a^*, b^*). \tag{D.23}$$

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

E Proof of Claim 3

Payoff Lower Bound: Player 2's strategy can be summarized by a triple $(r_\varnothing, r_H, r_L)$, where r_x is the probability of playing 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 has no incentive to play H, we have $r_L = 0$. 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 RHS of (E.2) is strictly greater than the RHS of (E.1), which implies

- 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 L when $a_{t-1} = L$, and moreover, $r_H < 1$. I show that there is no such equilibria when δ is close enough to 1. Let p_t be the ex ante probability of the event that:

- player 1 is the strategic type and plays H in period t.

Since the strategic type player 1 plays L for sure 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 is indifferent between T and N after observing $a_{t-1} = H$ if and only if

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

Notice that $\pi_0 + 2p_t - p_{t-1} \leq \frac{\pi_0}{2}$ if and only if $p_{t-1} - p_t \geq p_t + \frac{\pi_0}{2} \geq \frac{\pi_0}{2}$. This suggests that there can be at most

$$T \equiv \left\lceil \frac{2(1-\pi_0)}{\pi_0} \right\rceil$$

such periods. Since $\pi_0 + 2p_t - p_{t-1} \geq -(1-2\pi_0)$,

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

The RHS 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.

Existence of Equilibrium: I establish the existence of equilibrium in the game where $K = 1$ and $M = 0$ by constructing an equilibrium in which player 1's discounted average payoff is 1. Both players' strategies depend only on a_{t-1}:

1. When $a_{t-1} = \varnothing$ or H, the strategic type player 1 plays H and player 2 plays T.

2. When $a_{t-1} = L$, the strategic type player 1 plays H with probability 1/2 and L with complementary probability. Player 2 plays T with probability

$$\frac{\delta - (1-\delta)c_N}{(1+c_T) - (1-\delta)(1+c_N)},$$

and N with complementary probability.

One can verify that the strategic type player 1's continuation value is 1 when $a_{t-1} = \varnothing$ or H and his continuation value is $\frac{\delta-(1-\delta)c_N}{(1+c_T)-(1-\delta)(1+c_N)}(1+c_T)$ when $a_{t-1} = L$. The strategic type player 1 is indifferent between L and H when $a_{t-1} = L$ and strictly prefers H when $a_{t-1} = H$ or $\varnothing$.

References

[1] Acemoglu, Daron, Munther Dahleh, Ilan Lobel and Asu Ozdaglar (2011) "Bayesian Learning in Social Networks," *Review of Economic Studies*, 78(6), 1201-1236.

[2] Bai, Jie (2018) "Melons as Lemons: Asymmetric Information, Consumer Learning and Quality Provision," Working Paper.

[3] Banerjee, Abhijit (1992) "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, 107(3), 797-817.

[4] Barro, Robert (1986) "Reputation in a Model of Monetary Policy with Incomplete Information," *Journal of Monetary Economics*, 17(1), 3-20.

[5] Benabou, Roland and Jean Tirole (2006) "Incentives and Prosocial Behavior," *American Economic Review*, 96(5), 1652-1678.

[6] 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.

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

[8] Cai, Hongbin, Yuyu Chen and Hanming Fang (2009) "Observational Learning: Evidence from a Randomized Natural Field Experiment," *American Economic Review*, 99(3), 864-882.

[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(2), 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] Conley, Timothy and Christopher Udry (2010) "Learning about a New Technology: Pineapple in Ghana," *American Economic Review*, 100(1), 35-69.

[13] Deb, Rahul, Matthew Mitchell, and Mallesh Pai (2020) "(Bad) Reputation in Relational Contracting," Working Paper.

[14] Drakopoulos, Kimon, Asuman Ozdaglar and John Tsitsiklis (2012) "On Learning with Finite Memory," *IEEE Journal of Selected Topics in Signal Processing*, 6(7), 780-794.

[15] Ely, Jeffrey and Juuso Välimäki (2003) "Bad Reputation," *Quarterly Journal of Economics*, 118(3), 785-814.

[16] Ely, Jeffrey, Drew Fudenberg and David Levine (2008) "When is Reputation Bad?" *Games and Economic Behavior*, 63(2), 498-526.

[17] Ekmekci, Mehmet (2011) "Sustainable Reputations with Rating Systems," *Journal of Economic Theory*, 146(2), 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(4), 555-573.

[19] Fudenberg, Drew and David Levine (1983) "Subgame Perfect Equilibria of Finite and Infinite Horizon Games," *Journal of Economic Theory*, 31(2), 251-268.

[20] Fudenberg, Drew and David Levine (1989) "Reputation and Equilibrium Selection in Games with a Patient Player," *Econometrica*, 57(4), 759-778.

[21] Fudenberg, Drew and David Levine (1992) "Maintaining a Reputation when Strategies are Imperfectly Observed," *Review of Economic Studies*, 59(3), 561-579.

[22] Gale, Douglas and Shachar Kariv (2003) "Bayesian Learning in Social Networks," *Games and Economic Behavior*, 45(2), 329-346.

[23] Gossner, Olivier (2011) "Simple Bounds on the Value of a Reputation," *Econometrica*, 79(5), 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 (2020) "Rational Groupthink," *Quarterly Journal of Economics*, forthcoming.

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

[27] Levine, David (2019) "The Reputation Trap," Working Paper.

[28] Liu, Qingmin (2011) "Information Acquisition and Reputation Dynamics," *Review of Economic Studies*, 78(4), 1400-1425.

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

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

[31] Mailath, George and Larry Samuelson (2001) "Who Wants a Good Reputation?" *Review of Economic Studies*, 68(2), 415-441.

[32] Mirrlees, James (1976) "The Optimal Structure of Incentives and Authority within an Organization," *Bell Journal of Economics*, 7(1), 105-131.

[33] Nyqvist, Martina Björkman, Jakob Svensson and David Yanagizawa-Drott (2018) "Can Competition Reduce Lemons? A Randomized Intervention in the Antimalarial Medicine Market in Uganda," Working Paper.

[34] Pei, Harry (2020) "Reputation Building under Observational Learning," Working Paper, arXiv:2006.08068.

[35] Phelan, Christopher (2006) "Public Trust and Government Betrayal," *Journal of Economic Theory*, 130(1), 27-43.

[36] Rahmat, Suharni, Chew Boon Cheong and Mohd Syaiful (2016) "Challenges of Developing Countries in Complying Quality and Enhancing Standards in Food Industries," *Procedia-Social and Behavioral Sciences*, 224, 445-451.

[37] Rosenberg, Dinah, and Nicolas Vieille (2019) "On the Efficiency of Social Learning," *Econometrica*, 87(6), 2141-2168.

[38] Schmidt, Klaus (1993) "Reputation and Equilibrium Characterization in Repeated Games with Conflicting Interests," *Econometrica*, 61(2), 325-351.

[39] Smith, Lones and Peter Norman Sørensen (2000) "Pathological Outcomes of Observational Learning," *Econometrica*, 68(2), 371-398.

[40] Smith, Lones, Peter Norman Sørensen and Jianrong Tian (2021) "Informational Herding, Optimal Experimentation, and Contrarianism," *Review of Economic Studies*, forthcoming.