Reputation Building with Endogenous Speed of Learning

Harry PEI[*]

Preliminary and Incomplete. Comments are Welcomed.

August 17th, 2019

Abstract: I study reputation models where each short-run player observes a bounded subset of the long-run player's previous period actions, in addition to the entire history of her predecessors' actions. Reputation effects fail because the speed of learning decreases endogenously with the long-run player's patience. This leads to equilibria in which both players receive low payoffs. When each short-run player can also observe an informative signal about the long-run player's current period action, I propose a *resistent to learning* condition under which reputation effects fail. This is because the short-run player's action can be uninformative about the long-run player's type in periods where the latter receives a low stage-game payoff. When the environment is not resistent to learning, the patient long-run player can secure his commitment payoff in all equilibria. I compare my resistent to learning condition to the bounded/unbounded informativeness condition in observational learning models.

Keywords: reputation failure, endogenous signals, information aggregation, network monitoring
JEL Codes: C73, D82, D83

[*]Department of Economics, Northwestern University. Email: harrydp@northwestern.edu.

1 Introduction

Reputations are powerful tools to overcome lack-of-commitment problems. Such an intuition is formalized by Fudenberg and Levine (1989,1992), who show that patient players who have no commitment power can guarantee their optimal commitment payoffs (i.e., their Stackelberg payoffs) by building reputations. These theoretical results are well-supported empirically, as many success stories in the business world are attributed to reputations for supplying high quality products and providing good customer service (Bar-Isaac and Tadelis 2008).

Nevertheless, reputation mechanisms fail to work in many developing countries, where mistrust between business partners, lack of trustworthy brands, and low government credibility have become major obstacles for growth and development. For some concrete examples, the inability of central banks to convince citizens about their intentions (to lower inflation) reduces the effectiveness of monetary policies. Consumers' skepticism about product quality leads to low returns from a reputation for quality, which hurts firms' incentives to invest in quality. The resulting lack of the supply of high quality products makes the consumers' beliefs self-fulfilling.[1]

A common theme in these examples is the inability of the reputation building player to convey his future intentions via his current period behavior. These observations are inconsistent with the logic behind the reputation results of Fudenberg and Levine (1989, 1992), which suggests that whenever consumers are skeptical about a seller's product quality, they will be surprised after observing the seller supplying high quality and the probability with which they attach to the seller being committed goes up. After a bounded number of such surprises, consumers will be sufficiently convinced that the seller is committed to high quality and are willing to trust the seller in future periods.

Motivated by these episodes of reputation failures, I study a reputation model in which information about an informed player's past actions is *dispersed* among uninformed players and is *aggregated* via the uninformed players' actions. In my model, a long-lived player 1 (he, e.g., seller) interacts with an infinite sequence of short-lived player 2s (she, e.g., buyers), arriving one in each period and each plays the game only once. The long-run player is either an opportunistic type that maximizes his discounted average payoff, or a commitment type that mechanically plays his (pure) Stackelberg action in every period. I focus on situations in which the commitment type is arbitrarily rare.

Different from the canonical reputation models (Fudenberg and Levine 1989, 1992), every short-run

[1] Reputation failures have been widely documented in developing countries. In the market for watermelons, consumers are unwilling to pay high prices for high-quality melons, and future consumers are reluctant to reward sellers for supplying high quality (Bai 2018). In the market for malaria drugs, consumers purchase fake drugs despite effective ones are available on the market (Nyqvist, Svensson and Yanagizawa-Drott 2018).

player can only observe the long-run player's actions in the past K periods, in addition to observing the entire history of her predecessors' actions. This is motivated by the heterogeneous accessibility of different types of information. For example, by skimming through the summary statistics online, or by word-of-mouth communication with neighbors and friends, a potential buyer can know (or have a fair estimate about) the frequency with which the seller's product has been purchased as well as the timing of these purchases. However, figuring out the details of the seller's behavior requires time and effort, for example, a buyer needs to read the online reviews carefully, and it is usually the case that she has limited capacity to acquire and process such detailed information.

Theorem 1 shows that no matter how large K is, the patient long-run player's guaranteed equilibrium payoff is no more than his worst pure strategy Nash equilibrium payoff in the stage-game.[2] When the stage-game payoffs are monotone-supermodular, which is the case in the product choice game and the entry deterrence game, there exist equilibria in which both players' equilibrium payoffs are their respective minmax payoffs. These conclusions contrast to the reputation results of Fudenberg and Levine (1989, 1992), in which a patient player can secure his Stackelberg payoff in all equilibria.

Intuitively, such a distinction is driven by the differences between learning through *exogenous signals* (e.g., the ones in Fudenberg and Levine 1989, 1992) whose informativeness about the long-run player's action is exogenous, and learning through signals whose informativeness is endogenously determined in equilibrium (such as the short-run players' actions). In particular, the informativeness of these *endogenous signals* can vary with the other parameters of the game, such as players' stage-game payoffs and the long-run player's discount factor.

I explain the idea behind the proof using the well-known product choice game in Mailath and Samuelson (2001). A seller chooses between high and low effort, and each buyer chooses between a trusting action (e.g., buy the customized product) and a non-trusting action (e.g., buy the standardized product). The constructed equilibrium consists of three phases. Play starts from a *reputation building phase*, in which buyers do not trust and the seller mixes between high and low effort.[3] When a buyer observes low effort in the previous period, play remains in the reputation building phase. When a buyer observes high effort in the previous period, play transits to a *reputation maintenance phase* with positive probability, after which all buyers choose the trusting action and the seller exerts high effort on the equilibrium path. If the seller chooses low effort during the reputation maintenance phase, play transits to a *punishment phase*, in which all future buyers choose the non-trusting action and the seller chooses low effort. Such a punishment is implementable since all future buyers can observe that a

[2] My result applies to all games in which first, each player receives a different payoff under different action profiles, and second, there exists a pure strategy Nash equilibrium in the stage-game.

[3] The probability of high effort is low enough so that buyers have incentives to play the non-trusting action.

non-trusting action has been played after the trusting action has occurred, from which they learn that the seller is not committed. Therefore, myopic buyers have incentives to punish the seller by playing the non-trusting action. The transition probability between reputation building phase and reputation maintenance phase makes the strategic seller indifferent between high and low effort at every history of the reputation building phase. Importantly, the above transition probability *decreases* with the seller's discount factor. This reduces the speed of learning as the seller becomes more patient, which leads to low-payoff equilibria for both players.

In some applications, each buyer randomly samples among her predecessors and learns about their past experiences with the seller, or each buyer only observes the seller's actions against her neighbors in a stochastic network. Proposition 2 shows that the above insight remains robust as long as the number of neighbors each buyer can have is uniformly bounded from above, and the probability with which each buyer is observed by her immediate successor is uniformly bounded from below. Constructing equilibrium in this environment is challenging since the seller cannot observe whom each buyer samples, which leads to problems such as *private monitoring* and *private learning*. My proof overcomes this challenge by combining the belief-free approach (i.e., buyers' incentives do not depend on their beliefs about the seller's private history) with the belief-based approach (i.e., buyers are indifferent under their posterior beliefs). The latter is necessary due to the presence of commitment types, since a buyer who arrives sufficiently late has a strict incentive to play her Stackelberg best reply *if she knew* that the seller has exerted high effort in all previous periods.

The driving force behind my reputation failure result differs from that behind the herding results of Banerjee (1992) and Smith and Sørensen (2000). This is because in any equilibrium, given that the long-run player plays her Stackelberg action in every period, the short-run players *cannot* herd on an action that is not their Stackelberg best reply. Instead, the low-payoff equilibria are driven by the low speed of learning. This novel economic mechanism accounts for many episodes of reputation failures in practice. For example, rational consumers understand that opportunistic sellers have incentives to build reputations in order to milk them in the future. From this perspective, the more patient the seller is, the less informative his current period action is about his underlying type. The buyers' fear of facing an opportunistic seller who is imitating the commitment type reduces their willingness to trust upon observing the Stackelberg action. This undermines the informativeness of the buyers' actions about the seller's past actions, lowers the speed of learning, which in turn lowers the seller's returns from reputation building.[4]

[4]In the worst equilibrium, the buyer's action is informative about the seller's previous period action. To see this, suppose the buyer's action is uninformative about the seller's previous period action, then the strategic seller has a strict incentive to exert low effort. Therefore, after observing high effort in the current period, all buyers who can observe

In the second part of this paper, I study situations in which each short-run player can also observe an informative signal about the long-run player's current period action, in addition to what she can observe in the baseline model. This fits into situations in which a seller produces a product in advance, and potential buyers inspect the product, observe a noisy signal about its quality, before making their purchasing decisions. Since the long-run player's action choice affects the short-run player's action in the current period, he may have an incentive to play the Stackelberg action despite his action does not affect his continuation payoff.

If the informativeness of this signal is bounded, then player 1's guaranteed equilibrium payoff is strictly bounded below his Stackelberg payoff. This follows from the same logic as in the baseline model: when the commitment type is rare and player 2s believe that the strategic player 1 plays the commitment action with low but positive probability, no signal realization can convince them to play their Stackelberg best reply. As a result, the informative signal has no impact on the asymptotic speed of learning as well as the long-run player's equilibrium payoff in the worst case scenario.

Next, suppose there exists a signal realization that occurs with positive probability if and only if the long-run player plays the Stackelberg action (i.e., the informativeness of signal is *unbounded*), then in games where the long-run player's action choice is binary (i.e., *binary action games*), he can secure his Stackelberg payoff by playing his Stackelberg action in every period.

The proof of this result proceeds in two steps. First, regardless of the short-run players' belief about the long-run player's action as well as the other details of the signal structure, the probability with which the short-run players playing the Stackelberg best reply is *weakly higher* when the long-run player plays his Stackelberg action, compared to the case in which he plays the alternative action.[5] This leads to a lower bound on the informativeness of the short-run player's actions when the ex ante probability with which she plays the Stackelberg best reply is bounded away from one.

Second, due to the differences in private information, the informativeness of a short-run player's action can be different from her perspective and from her successors' perspectives. To circumvent this difficulty, I use the observation that when the ex ante probability with which player 2 plays the Stackelberg best reply in a given period is bounded away from one, then the probability with which the long-run player plays the Stackelberg action in the past K periods cannot be too low. This is because otherwise, the short-run player will be convinced that the long-run player is committed and the probability with which she plays the Stackelberg action cannot be bounded away from one. As a result, the probability with which her successors' higher order belief attaches to her true belief is

this will be fully convinced that the seller is committed. As the buyers' length of memory increases, the patient seller's guaranteed equilibrium payoff approaches his Stackelberg payoff no matter how likely the commitment type is.

[5] This relies on the unboundedly informative signal and the long-run player's action choice being binary.

uniformly bounded from below. This leads to a lower bound on the *believed informativeness* of the current period short-run player's action as a function of the probability with which she plays the Stackelberg best reply. The two parts together implies the commitment payoff theorem.

Given this connection between successful reputation formation and unboundedly informative signals in binary action games, one may wonder whether this relationship applies more generally. Unfortunately, this is not the case under generic payoffs. In particular, the short-run players' actions can be uninformative about the long-run player's type under some equilibrium beliefs about the latter's actions, even when there exist signal realizations that are unboundedly informative.

I introduce a joint condition on the short-run player's payoff function and the monitoring structure called *resistant to learning*, that characterizes this property. In particular, an environment is resistant to learning against a pure commitment action if there exists an alternative (mixed) action of the long-run player with the commitment action in its support, as well as a best reply of the short-run player against this alternative action, such that (1) the distribution of the short-run player's actions according this best reply is the same under the alternative action and under the commitment action; and (2) the resulting distribution is not the degenerate distribution on her best reply to the commitment action.

Conversely, an environment is not resistant to learning if for any alternative action and any best reply of the short-run player, the resulting distributions over the short-run players' actions are the same if and only if it is the degenerate distribution on her best reply against the commitment action. One can verify that (1) when the informativeness of all signal realizations are bounded, the environment is resistant to learning against all actions; (2) when there exists a signal realization whose informativeness is unbounded about a commitment action, the environment is not resistant to learning against the commitment action when the long-run player's action choice is binary.

I show in Theorem 3 that when the environment is resistant to learning against a pure commitment action, then there exists an open set of payoff functions of the long-run player, under which his worst equilibrium payoff is strictly bounded below his commitment payoff no matter how patient he is. Conversely, when the environment is not resistant to learning against a pure commitment action, then the patient long-run player can guarantee his commitment payoff regardless of his payoff function.

Related Literature: My paper is related to the literature on social learning and reputation formation. Compared to the existing results of social learning that focus on players' asymptotic beliefs (Banerjee 1992, Smith and Sørensen 2000), I study a patient player's returns from reputation building when information about his past behavior is dispersed and is aggregated via his opponents' action choices. Since the reputation-building player discounts future payoffs, my model takes into account the *speed* of observational learning and studies its welfare consequences. Reputation effects fail in my model despite short-run players never join any *bad herd*, which differentiates my results from the herding results of Banerjee (1992) and Smith and Sørensen (2000).

The reputation failure results in this paper contrast to the canonical reputation results of Fudenberg and Levine (1989,1992) and Gossner (2011), in which the long-run player secures his commitment payoff in all equilibria when his opponents can observe a signal that statistically identifies his commitment action. In my model, short-run players learn primarily through their predecessors' actions. When the long-run player becomes more patient, the informativeness of this endogenous signal vanishes, which leads to low returns from reputation building and low incentives to build reputations.[6]

The mechanism behind reputation failure suggested in this paper differs from the one in Ely and Välimäki (2003), Ely, Fudenberg and Levine (2008), and Deb, Mitchell and Pai (2019). In those models, the uninformed player(s) can choose a *non-participating action* under which the public signal is uninformative about the informed player's actions. In contrast, the uninformed players cannot unilaterally shut down learning in my model. In fact, whenever the informed player's continuation payoff is low enough, the uninformed player's action in the current period must be informative about the informed player's action in the previous period. Reputation effects fail because its informativeness vanishes as the informed player becomes arbitrarily patient.

This paper is also related to reputation models with limited memories, such as the ones of Liu (2011) and Liu and Skrzypacz (2014). They study models in which short-run players *cannot* observe their predecessors' actions, and focus on the long-run player's ability to unilaterally clean up histories. Instead, my analysis focuses on the effectiveness of reputation building through observational learning. In order to answer this question, I study a model in which each short-run player observes all the actions taken by her predecessors, so that the long-run player *cannot* unilaterally clean up his past records.

Logina, Lukyanov and Shamruk (2019) characterize Markov equilibria when each buyer can observe an informative signal about the seller's current period action, and the opportunistic seller's payoff from high effort is lower than his minmax payoff. They show that the opportunistic seller has an incentive to exert effort when his reputation is intermediate, but strictly prefers to shirk when his reputation is either very high or very low. Intuitively, when reputation is very high or very low, the buyer's prior belief is sufficiently precise such that her behavior is irresponsive to her private signals. This eliminates the opportunistic seller's incentive to exert high effort.

[6] Similar ideas also appear in Sobel (1985), in which truth-telling leads to a larger increase in the sender's reputation when his gain from lying is larger.

2 Baseline Model

Primitives: Time is discrete, indexed by $t = 0, 1, 2....$ A long-lived player 1 (he) with discount factor $\delta \in (0, 1)$ interacts with an infinite sequence of short-lived player 2s (she), arriving one in each period and each plays the game only once. In period t, players simultaneously choose their actions $(a_t, b_t) \in A \times B$, with A and B being finite sets. Players have access to a public randomization device, with the realization in period t denoted by $\xi_t \in [0, 1]$.

Players' stage-game payoffs are $u_1(a_t, b_t)$ and $u_2(a_t, b_t)$. Let $\text{BR}_1 : \Delta(B) \rightrightarrows 2^A \backslash \{\varnothing\}$ and $\text{BR}_2 : \Delta(A) \rightrightarrows 2^B \backslash \{\varnothing\}$ be player 1's and player 2's best reply correspondences in the stage-game. I make the following assumption, which is satisfied for generic payoff functions:

Assumption 1. *For every $(a, b) \neq (a', b')$, we have $u_1(a, b) \neq u_1(a', b')$ and $BR_2(a)$ is a singleton.*

Under Assumption 1, player 2 has a strict best reply against each of player 1's pure actions, and player 1 has a unique (pure) Stackelberg action, denoted by $a^* \in A$ and is the unique element of the set:

$$\arg\max_{a \in A} \Big\{ \min_{b \in \text{BR}_2(a)} u_1(a, b) \Big\}. \tag{2.1}$$

Let b^* be the unique element in $\text{BR}_2(a^*)$, namely, player 2's best reply to the Stackelberg action. The next assumption ensures that player 1 can benefit from committing to play pure actions, which rules out games such as matching pennies and rock-paper-scissors.

Assumption 2. *There exists a pure strategy Nash equilibrium in the stage-game.*

Information & Monitoring Structure: Player 1 is one of the two possible types $\omega \in \{\omega^s, \omega^c\}$, which is player 1's private information and is perfectly persistent. Either he is a commitment type ω^c, who mechanically plays a_1^* in every period; or he is a strategic type ω^s, who can flexibly choose his actions in order to maximize his discounted average payoff, given by:

$$\sum_{t=0}^{\infty} (1 - \delta)\delta^t u_1(a_t, b_t). \tag{2.2}$$

Player 2's prior belief attaches probability $\pi_0 \in (0, 1)$ to the commitment type. Player 2's private history consists of all the actions of her predecessors, the past realizations of the public randomization devices, as well as player 1's actions in the past K periods.[7] Formally, let h^t be a typical history of

[7] My result also applies when there is no public randomization device, or when the short-run player in period t can only observe the realization of the public randomization device in the current period, or can only observe a stochastic subset of the past realizations.

player 2 who arrives in period t, with

$$h^t \equiv \begin{cases} \{b_0, b_1, ..., b_{t-1}, a_{t-K}, a_{t-K+1}, ..., a_{t-1}, \xi_0, ..., \xi_t\} & \text{if } t \geq K \\ \{b_0, b_1, ..., b_{t-1}, a_0, a_1, ..., a_{t-1}, \xi_0, ..., \xi_t\} & \text{if } t < K. \end{cases} \tag{2.3}$$

Let $\pi(h^t)$ be the probability her posterior belief at h^t assigns to $\omega = \omega^c$. Sometimes, I write π_t instead of $\pi(h^t)$ for notation simplicity.

In the *agent normal form* of this repeated game, strategic player 1's private history consists of the entire sequence of action profiles and realizations of public randomizations. Let h_1^t be a typical private history in period t, with $h_1^t \equiv \{a_0, ..., a_{t-1}, b_0, ..., b_{t-1}, \xi_0, ..., \xi_t\}$. For every $t \in \mathbb{N}$, let $\mathcal{H}^t$ be the set of h^t and let $\mathcal{H} \equiv \cup_{t=0}^{\infty} \mathcal{H}^t$. Let $\mathcal{H}_1^t$ be the set of h_1^t and let $\mathcal{H}_1 \equiv \cup_{t=0}^{\infty} \mathcal{H}_1^t$. Strategic player 1's strategy is $\sigma_1 : \mathcal{H}_1 \to \Delta(A)$, with $\sigma_1 \in \Sigma_1$. Player 2's strategy is $\sigma_2 : \mathcal{H} \to \Delta(B)$, with $\sigma_2 \in \Sigma_2$.

Equilibrium: The solution concept is Bayesian Nash Equilibrium (or equilibrium for short), which consists of a strategy for the strategic type player 1, and a strategy for player 2s. Let $\text{NE}(\delta, \pi_0, K) \subset \Sigma_1 \times \Sigma_2$ be the set of equilibria under parameter configuration (δ, π_0, K). Let $\mathbb{E}^{(\sigma_1, \sigma_2, \pi_0)}[\cdot]$ be the expectation operator under the probability measure over histories induced by (σ_1, σ_2) when player 2s' prior is π_0. Let $\mathbb{E}_1^{(\sigma_1, \sigma_2)}[\cdot]$ be the expectation operator under the probability measure over histories induced by (σ_1, σ_2), conditional on player 1 being strategic. The strategic long-run player's equilibrium payoff is:

$$\mathbb{E}_1^{(\sigma_1, \sigma_2)} \Big[\sum_{t=0}^{\infty} (1-\delta)\delta^t u_1(a_t, b_t) \Big]. \tag{2.4}$$

I evaluate the short-run players' welfare according to their expected discounted average payoff with discount rate δ, namely, their welfare under strategy profile (σ_1, σ_2) is:

$$\mathbb{E}^{(\sigma_1, \sigma_2, \pi_0)} \Big[\sum_{t=0}^{\infty} (1-\delta)\delta^t u_2(a_t, b_t) \Big]. \tag{2.5}$$

My result also applies to other discount factors adopted by a social planner when he evaluates different generations of short-run players' payoffs, as long as it is no more than δ.

3 Reputation Failure under Endogenous Signals

Let $(a', b') \in A \times B$ be the worst pure strategy Nash equilibrium for player 1 in the stage-game. Let $\underline{v}_1 \equiv u_1(a', b')$, which by definition, is no more than player 1's Stackelberg payoff $u_1(a^*, b^*)$. In games where player 1 faces a strict lack-of-commitment problem (i.e., $a^* \notin \text{BR}_1(b^*)$), such as the product choice game and the entry deterrence game, $\underline{v}_1 < u_1(a^*, b^*)$. Let the cutoff discount factor $\underline{\delta}$ be:

$$\underline{\delta}_1 \equiv \begin{cases} \max\left\{\frac{\max_{a\in A} u_1(a,b^*)-u_1(a^*,b^*)}{\max_{a\in A} u_1(a,b^*)-\underline{v}_1}, \frac{\underline{v}_1-u_1(a^*,b')}{u_1(a^*,b^*)-u_1(a^*,b')}\right\} & \text{if } \underline{v}_1 < u_1(a^*,b^*) \\ 0 & \text{if } \underline{v}_1 = u_1(a^*,b^*). \end{cases}$$

The main result is stated as Theorem 1:

Theorem 1. *If the stage-game payoffs satisfy Assumptions 1 and 2, then for every $K \in \mathbb{N}$, there exists $\overline{\pi}_0 \in (0,1)$, such that for every $\pi_0 \in (0, \overline{\pi}_0)$ and $\delta \geq \underline{\delta}_1$, there exists $(\sigma_1^\delta, \sigma_2^\delta) \in NE(\delta, \pi_0, K)$, such that:*

$$\mathbb{E}_1^{(\sigma_1^\delta,\sigma_2^\delta)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t u_1(a_t,b_t)\Big] \leq \underline{v}_1. \tag{3.1}$$

According to Theorem 1, when information about player 1's past behavior is dispersed among short-run players and is aggregated via his opponents' actions, player 1's return from reputation building can be low, no matter how patient he is. This contrasts to the canonical reputation results in Fudenberg and Levine (1989, 1992) and Gossner (2011), which show that if player 2s have unbounded observations of player 1's actions, or noisy signals that can statistically identify player 1's actions, then a patient player 1 can *guarantee* his Stackelberg payoff in all equilibria of the reputation game.

The mechanism behind Theorem 1 differs from that in the social learning models of Banerjee (1992) and Smith and Sørensen (2000), in which information aggregation fails because the short-run players rationally ignore their private signals and herd on a *wrong* action. To see this, I argue that in my reputation model, player 2s cannot herd on any action b' ($\neq b^*$) when the strategic player 1 imitates the commitment type. This is because when herding occurs, the strategic type has no intertemporal incentives and in equilibrium, plays his myopic best reply against b'. I consider two cases separately. First, suppose $\mathrm{BR}_1(b') = \{a^*\}$, then player 2 believes that a^* is played with probability 1 after observing a^* in the past K periods, in which case her best reply is b^*. This leads to a contradiction. Next, suppose $\mathrm{BR}_1(b') \neq \{a^*\}$, then at histories where player 2s start to herd, their posterior attaches probability 1 to the commitment type after observing a^* in the past K periods. This is because strategic player 1 does not play a^* when player 2s start to herd. As a result, player 2s have strict incentives to play b^* after observing a^* in the past K periods, which leads to a contradiction.

To highlight the novel economic mechanism at work, I provide a constructive proof of Theorem 1. In the equilibria I construct, the informativeness of the uninformed player's actions vanishes as the informed player becomes more patient. As a result, despite information about the long-run player's type is aggregated asymptotically, the *speed of learning* vanishes to 0 as $\delta \to 1$. This eliminates player

1's returns from reputation building, making the short-run players' adverse beliefs self-fulfilling.

Proof of Theorem 1: Recall the definitions of (a', b') and (a^*, b^*). If $b' = b^*$, then according to Assumption 1, $a' = a^*$ and $\underline{v}_1$ can be attained by playing (a^*, b^*) in every period.

In what follows, I focus on the nontrivial case in which $b' \neq b^*$. Assumption 1 and the definitions of a', b', a^*, b^* imply that:

$$u_1(a^*, b^*) > u_1(a', b') > u_1(a^*, b'). \tag{3.2}$$

Let $q^* \in (0, 1)$ be small enough such that b' is player 2's best reply against player 1's mixed action $q^* \circ a^* + (1 - q^*) \circ a'$. The upper bound on player 2's prior $\overline{\pi}_0$ is given by:

$$\frac{\overline{\pi}_0}{1 - \overline{\pi}_0} = \Big(\frac{q^*}{2 - q^*}\Big)^{K+1}.$$

For every $\pi_0 \leq \overline{\pi}_0$, I construct the following *three-phase equilibrium* in which player 1's payoff is $\underline{v}_1$ regardless of δ. I start from describing players' strategies, and later verify players' incentive constraints taken into account player 2s' posterior beliefs.

Play starts from a *reputation building phase*, in which player 2 plays b', and strategic player 1 mixes between a^* and a' with probabilities $\frac{q^*}{2-q^*}$ and $\frac{2-2q^*}{2-q^*}$, respectively. In period $t \geq 1$, play remains in the reputation building phase if $a_{t-1} \neq a^*$. Play transits to a *reputation maintenance phase* with strictly positive probability if $a_{t-1} = a^*$, after which player 1 plays a^* and player 2 plays b^* on the equilibrium path. Whether play transits to the reputation maintenance phase or not depends on the realization of public randomization in the beginning of period t, before players choosing their actions. The transition probability r is pinned down by:

$$u_1(a', b') = (1 - \delta)u_1(a^*, b') + \delta\Big\{r u_1(a^*, b^*) + (1 - r)u_1(a', b')\Big\}, \tag{3.3}$$

which is between 0 and 1 when δ is close enough to 1. Future player 2s know the calendar time at which play transits to the reputation maintenance phase: it coincides with the first period in which player 2 plays b^*. If player 1 plays actions other than a^* after reaching the reputation maintenance phase, play transits to a *punishment phase*, in which (a', b') is played in all subsequent periods.

I verify players' incentives and the feasibility of player 1's behavioral strategy in the reputation building phase. First, when δ is large enough such that:

$$u_1(a^*, b^*) \geq (1 - \delta) \max_{a \in A} u_1(a, b^*) + \delta u_1(a', b'),$$

player 1 has an incentive to play a^* in the reputation maintenance phase. Second, player 1 is indifferent between a^* and a' in the reputation building phase according to (3.3). Moreover, he strictly prefers a' to actions other than a' and a^*. Third, I verify that player 2's incentive to play b' at the reputation building phase, by showing that at every history of this phase, player 2 believes that player 1 will play a^* with probability less than q^*. In particular, after observing a^* being played in the past K periods, player 2's posterior belief at h^t, denoted by π_t satisfies:

$$\frac{\pi_t}{1-\pi_t}\Big/\frac{\pi_0}{1-\pi_0} = \frac{\Pr^{(\sigma_1^\delta,\sigma_2^\delta)}(a^*,...,a^*|\omega^c)}{\Pr^{(\sigma_1^\delta,\sigma_2^\delta)}(a^*,...,a^*|\omega^s)} \cdot \frac{\Pr^{(\sigma_1^\delta,\sigma_2^\delta)}(b',...,b',\xi_0,...,\xi_t|a^*,...,a^*,\omega^c)}{\Pr^{(\sigma_1^\delta,\sigma_2^\delta)}(b',...,b',\xi_0,...,\xi_t|a^*,...,a^*,\omega^s)}. \tag{3.4}$$

in which $\Pr^{(\sigma_1^\delta,\sigma_2^\delta)}(\cdot)$ is the probability measure over $\mathcal{H}_1^t$ generated by strategy profile $(\sigma_1^\delta,\sigma_2^\delta)$. By construction,

$$\frac{\Pr^{(\sigma_1^\delta,\sigma_2^\delta)}(a^*,...,a^*|\omega^c)}{\Pr^{(\sigma_1^\delta,\sigma_2^\delta)}(a^*,...,a^*|\omega^s)} = \Big(\frac{q^*}{2-q^*}\Big)^{-K},$$

and

$$\frac{\Pr^{(\sigma_1^\delta,\sigma_2^\delta)}(b',...,b',\xi_0,...,\xi_t|a^*,...,a^*,\omega^c)}{\Pr^{(\sigma_1^\delta,\sigma_2^\delta)}(b',...,b',\xi_0,...,\xi_t|a^*,...,a^*,\omega^s)} \leq 1.$$

Since $\frac{\pi_0}{1-\pi_0} \leq \frac{\overline{\pi}_0}{1-\overline{\pi}_0} = \Big(\frac{q^*}{2-q^*}\Big)^{-K-1}$, we know that $\pi_t \leq \frac{q^*}{2}$ for every history h^t of the reputation building phase. Given strategic player 1's strategy, the probability with which player 2 believes that player 1 will play a^* at the reputation building phase is below q^*. This verifies player 2's incentive to play b'. $\square$

Theorem 1 highlights the payoff consequences of reputation failure from the long-run player's perspective. Next, I investigate the effect of slow-learning on the short-run players' welfare. Let $(a'',b'') \in A \times B$ be the worst pure strategy Nash equilibrium for player 2 in the stage game. If there are multiple such equilibria, pick the one that is worst for player 1. Let

$$\underline{\delta}_2 \equiv \begin{cases} \max\Big\{\frac{\max_{a\in A} u_1(a,b^*)-u_1(a^*,b^*)}{\max_{a\in A} u_1(a,b^*)-u_1(a'',b'')}, \frac{u_1(a'',b'')-u_1(a^*,b'')}{u_1(a^*,b^*)-u_1(a^*,b'')}\Big\} & \text{if } u_1(a'',b'') < u_1(a^*,b^*) \\ 0 & \text{if } u_1(a'',b'') = u_1(a^*,b^*) \end{cases}$$

be the cutoff discount factor. I show the following proposition:

Proposition 1. *Under Assumptions 1 and 2, for every $K \in \mathbb{N}$ and $\varepsilon > 0$, there exists $\overline{\pi}_0 \in (0,1)$ such that for every $\pi_0 \in (0,\overline{\pi}_0)$ and $\delta \geq \underline{\delta}_2$, there exists $(\sigma_1^\delta,\sigma_2^\delta) \in NE(\delta,\pi_0,K)$, such that:*

$$\mathbb{E}^{(\sigma_1^\delta,\sigma_2^\delta,\pi_0)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t u_2(a_t,b_t)\Big] \leq u_2(a'',b'') + \varepsilon. \tag{3.5}$$

Proof of Proposition 1: Consider the class of equilibria constructed in the proof of Theorem 1. Let V_2 be the short-run players' discounted average payoff in the reputation building phase, we have:

$$V_2 = (1-\delta)\Big\{q^* u_2(a^*, b'') + (1-q^*)u_2(a'', b'')\Big\} + \delta\Big\{(1-q^*)V_2 + q^*(1-r)V_2 + q^* r u_2(a^*, b^*)\Big\}, \quad (3.6)$$

in which $q^* \in (0,1)$ is small enough such that b'' is player 2's best reply against the mixed action of a^* with probability q^* and a'' with complementary probability, and r is the probability of transiting to the reputation maintenance phase after observing the long-run player played a^* in the previous period. Equation (3.3) implies that r is proportional to $1-\delta$. Let $\gamma \equiv r/(1-\delta)$, which is independent of δ. Plugging $r = (1-\delta)\gamma$ into (3.6) and rearranging terms, we have:

$$V_2 = \frac{(1-q^*)u_2(a'', b'') + q^* u_2(a^*, b'') + \delta q^* \gamma u_2(a^*, b^*)}{1 + \delta q^* \gamma} \quad (3.7)$$

For every $\varepsilon > 0$, there exists q^* small enough such that the RHS of (3.7) is strictly less than $u_2(a'', b'') + \varepsilon$. Let $\overline{\pi}_0 \equiv (q^*)^{K+1}$, the resulting strategy profile is an equilibrium in which the short-run players' (discounted average) welfare is no more than $u_2(a'', b'') + \varepsilon$. $\square$

I discuss the implications of Theorem 1 and Proposition 1 in the product choice game of Mailath and Samuelson (2001). Suppose the long-run player is a seller (row player) and the short-run players are a sequence of buyers. Their stage-game payoffs are given by:

–	T	N
H	$1, 1$	$-1, 0$
L	$2, -1$	$0, 0$

Suppose with probability π_0, the seller commits to play H in every period. My results imply that for every $\varepsilon > 0$, there exists $\overline{\pi}_0 > 0$, such that when π_0 is below $\overline{\pi}_0$, there exist equilibria in which the seller's discounted average payoff is 0 and the buyers' discounted average welfare is less than ε. These adverse equilibria exist regardless of how large δ is.

The equilibria constructed in these proofs shed light on some of the difficulties faced by reputation-building sellers in practice, which can account for some of the reputation failures documented in the empirical literature, such as the ones in Bai (2018) and Nyqvist, Svensson and Yanagizawa-Drott (2018). In particular, when the seller is patient, he is willing to sacrifice his current period payoff even though the probability of receiving a high continuation payoff is very low. When buyers understand the seller's strategic motives and when they believe that the seller cares more about his continuation payoffs, they attribute less to the seller having an intrinsic preference for supplying high quality after observing high quality. In equilibrium, their future actions become less responsive to the seller's current period action. This slows down the *speed of learning.* As shown in Theorem 1 and Proposition 1, the aforementioned channel can completely eliminate the returns from reputation building as well as players' surplus from their long-term relationship.

3.1 Minmax Payoff

I provide sufficient conditions under which the patient long-run player's guaranteed equilibrium payoff coincides with his minmax payoff. To account for the uninformed players' myopia, I adopt the notion of minmax payoff introduced by Fudenberg, Kreps and Maskin (1990). First, in *monotone-supermodular* games, player 1's lowest pure stage-game Nash equilibrium payoff coincides with his minmax payoff.

Assumption 3 (Monotone-Supermodularity). (u_1, u_2) *is monotone-supermodular if there exist a ranking* $\succ_a$ *on* A*, and a ranking* $\succ_b$ *on* B *under which:*[8]

1. u_1 *is strictly increasing in* b *and is strictly decreasing in* a*.*

2. u_2 *has strictly increasing differences in* (a, b)*.*

Assumption 3 is satisfied in the aforementioned product choice game: it is costly for a firm to supply high quality, but it can strictly benefit from consumers' trusting behaviors, and consumers have stronger incentives to play the trusting action when the firm supplies high quality. It is also satisfied in the entry deterrence game of Schmidt (1993), with stage-game payoffs given by:

–	O	E
F	$1, 0$	$-1, -1$
A	$2, 0$	$0, 1$

In this game, it is costly for the incumbent to lower prices (or *fight*), but it can strictly benefit from the entrants staying out. Furthermore, entrants have stronger incentives to stay out when incumbents are more likely to set low prices.

Let $\underline{a}$ be player 1's lowest action and let $\underline{b} \equiv \mathrm{BR}_2(\underline{a})$. According to the folk theorem result in Fudenberg, Kreps and Maskin (1990), player 1's minmax payoff taken into account player 2's myopia is $u_1(\underline{a}, \underline{b})$. This coincides with his lowest equilibrium payoff in the stage-game. The following result is an immediate corollary of Theorem 1 and Proposition 1.

[8] This monotone-supermodularity condition is similar to, albeit different from that in Pei (2018). In Pei (2018), the long-run player has persistent private information about a payoff relevant state, and monotone-supermodularity requires complementarity between the state and the action profile in players' payoff functions.

Corollary 1. *When the stage-game payoffs satisfy Assumptions 1, 2, and 3. Then for every* $K \in \mathbb{N}$ *and* $\varepsilon > 0$*, there exists* $\overline{\pi}_0 \in (0,1)$ *such that for every* $\pi_0 \in (0, \overline{\pi}_0)$ *and* δ *large enough, there exists an equilibrium in which player* 1*'s payoff equals his minmax payoff, and player* 2*'s payoff is below* ε *plus her minmax payoff.*

Next, I consider games in which player 2 needs to play a mixed action in order to minmax player 1. Let $\beta^* \in \Delta(B)$ be player 2's action that minmaxes player 1, and let $\alpha^* \in \Delta(A)$ be one of player 1's best replies to β^* such that every action in the support of β^* is player 2's pure best reply to α^*. A similar construction to the proof of Theorem 1 shows that player 1's guaranteed equilibrium payoff coincides with his minmax payoff in all of the following three cases:

1. $a^* \notin \text{supp}(\alpha^*)$ and $b^* \notin \text{supp}(\beta^*)$;
2. $a^* \in \text{supp}(\alpha^*)$ and $b^* \notin \text{supp}(\beta^*)$;
3. $a^* \in \text{supp}(\alpha^*)$ and $b^* \in \text{supp}(\beta^*)$.

The only case that is not covered is one in which $a^* \notin \text{supp}(\alpha^*)$ but $b^* \in \text{supp}(\beta^*)$, namely, the Stackelberg action is not player 1's stage-game best reply to player 2's minmax action, and in order to minmax player 1 while guaranteeing player 2's stage-game incentive constraint, player 2 needs to play the Stackelberg best reply b^* with positive probability.

3.2 Stochastic Sampling

In some applications of interest, consumers stochastically sample among their predecessors to learn about their experiences with the seller (Banerjee and Fudenberg 2004), or each consumer only talks to a subset of his predecessors, interpreted as her friends, before making her purchasing decision (Acemoglu, Dahleh, Lobel and Ozdaglar 2011). Importantly, the seller does not know who do each buyer samples nor does he know each buyer's social network.

Motivated by these applications, I focus on the *product choice game* and generalize the insights of Theorem 1 to settings with stochastic sampling or stochastic social networks. For every $t \geq 1$, let

$$\mathcal{N}_t \in \Delta\Big(2^{\{0,1,\dots,t-1\}}\Big)$$

be the distribution over the tth short-run player's neighborhood, and let N_t be the realization of $\mathcal{N}_t$. The public history consists of

$$h^t \equiv \Big\{b_0, b_1, \dots, b_{t-1}, \xi_0, \dots, \xi_t\Big\}.$$

Player 2's *private history* in period t is:

$$h_2^t \equiv \Big\{N_t, b_0, b_1, ..., b_{t-1}, \Big(a_s\Big)_{s\in\mathcal{N}_t}, \xi_0, ..., \xi_t\Big\}. \tag{3.8}$$

Let $\mathcal{H}_2^t$ be the set of h_2^t, and let $\mathcal{H}_2 \equiv \cup_{t\in\mathbb{N}}\mathcal{H}_2^t$ be the set of player 2's private histories. Importantly, player 1 cannot observe the current and past realizations of $\mathcal{N}_t$, and therefore, he may not know player 2's posterior beliefs about his type and about his private history. I make the following assumption on the stochastic network $\{\mathcal{N}_t\}_{t\in\mathbb{N}}$:

Assumption 4. *For every $t \neq s$, $\mathcal{N}_t$ and $\mathcal{N}_s$ are independent random variables. Moreover, there exist $K \in \mathbb{N}$ and $\gamma \in (0,1)$ such that for every $t \geq 1$,*

$$\Pr\Big(|\mathcal{N}_t| \leq K\Big) = 1 \text{ and } \Pr\Big(t-1 \in \mathcal{N}_t\Big) \geq \gamma.$$[9]

The first part of Assumption 4 requires that players' neighborhoods to be independently distributed, which is satisfied in Acemoglu, Dahleh, Lobel and Ozdaglar (2011). The second part implies that it is *common knowledge* that each buyer only samples a bounded subset of his predecessors' experiences. This bound is interpreted as a constraint on the buyers' ability to acquire or process detailed information. The third part requires that each buyer samples her immediate predecessor with probability bounded from below. This assumption rules out uniform sampling (i.e., the agent samples K out of t predecessors, and each predecessor is sampled with equal probability) since the probability with which the immediate predecessor's action being observed vanishes as the sample size becomes large. Without this part of Assumption 4, the buyers' actions will not be adequate to motivate the seller to play H as time goes to infinity.

I show the following result in context of the product choice game, which applies more generally to monotone-supermodular games in which player 2's action choice is binary.

Proposition 2. *In the product choice game, if the sampling process satisfies Assumption 4, then there exists $\overline{\pi}_0 \in (0,1)$, such that for every $\pi_0 \in (0, \overline{\pi}_0)$ and δ large enough, there exists $(\sigma_1^\delta, \sigma_2^\delta) \in$ $NE(\delta, \pi_0, K)$, such that:*

$$\mathbb{E}_1^{(\sigma_1^\delta,\sigma_2^\delta)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t u_1(a_t,b_t)\Big] \leq \underline{v}_1. \tag{3.9}$$

According to Proposition 2, Theorem 1 is robust against the *private monitoring* of the long-run player's actions and the short-run players' *private learning.* The result only requires the following

[9] Abusing notation, I use $t-1 \in \mathcal{N}_t$ to denote the event that $t-1 \in N_t$ given that N_t is distributed according to $\mathcal{N}_t$.

aspects of the stochastic network to be common knowledge: first, player 2's observations about player 1's past actions is uniformly bounded from above; and second, player 1's action against the current period player 2 is observed in the next period with probability uniformly bounded from below.

The proof is in Appendix A, which constructs three-phase equilibria similar to those in the proof of Theorem 1. To overcome the challenges brought by private monitoring and private learning, I use a combination of *belief-free equilibria* and *belief-based approach*.

For some intuition, let q^* be the cutoff probability above which player 2 has an incentive to play T. When the calendar time t is low enough such that the probability of commitment type is below q^* conditional on any *complete history* (i.e., one that consists of action profiles in all previous periods), the strategic long-run player mixes between H and L with probabilities such that conditional on each complete history, player 2 believes that H is played with probability q^*. As a result, player 2s' incentives are belief free with respect to player 1's private history.

When the calendar time t is larger than some cutoff $M \in \mathbb{N}$, the probability of commitment type is above q^* conditional on player 1 playing H in all previous periods, then player 2 has a strict incentive to play T if *she knew* that the complete history is $\{(H,N),(H,N),...,(H,N)\}$. This explains why the equilibrium cannot be belief-free with respect to player 1's private history when calendar time is large. To address this issue, I use a belief-based approach that relies on two observations. First, in period t, the number of player 2's private histories with length no more than K is no more than $2^K \sum_{j=0}^{K} \binom{t}{j}$. Second, strategic player 1 can condition the probability with which he plays H on his private history, and there are 2^t realizations of his private histories. For all M relatively large compared to K (which is the case when $\overline{\pi}_0$ is small enough), we have $2^K \sum_{j=0}^{K} \binom{t}{j} < 2^t$ for all $t \geq M$. As a result, under any stochastic network that satisfies Assumption 4, there exists a mapping from player 1's private history to his mixed actions such that conditional on each h_2^t with $t \geq M$, player 2 believes that H will be played with probability q^*.[10]

3.3 Comparison to Existing Reputation Models

I compare Theorem 1 to the canonical reputation results in Fudenberg and Levine (1989, 1992) and Gossner (2011). The key is to distinguish between the *noisy endogenous signals* in my model (e.g., short-run players' actions), and the *noisy exogenous signals* in theirs. In the current model, there are two obstacles to learn about player 1's type. First, player 1's action can be uninformative about his

[10] This belief-based construction only works for large enough calendar time. This is because when t is not large enough compared to K, $2^K \sum_{j=0}^{K} \binom{t}{j} > 2^t$, which means that under generic stochastic networks, there exists no strategy of player 1 under which player 2 is indifferent between T and N at all of her private histories. As a result, the belief-free construction when calendar time is below M is indispensable.

type. This is the case when the strategic type plays a^* with high probability. Second, player 2's action can be uninformative about player 1's past actions.

The first obstacle has negligible payoff consequences for the patient long-run player. This is because when player 2 expects a^* to be played with high enough probability in the current period, she has a strict incentive to play b^* and player 1 can secure his Stackelberg payoff in that period by playing a^*. The second obstacle is novel and has significant payoff consequences. Focusing on the product choice game with $K = 1$, I argue that in the worst equilibrium for player 1, b_{t+1} is informative about a_t, but its informativeness vanishes as $\delta \to 1$.

To start with, consider a candidate equilibrium in which b_{t+1} is *uninformative* about a_t. According to Assumption 2, player 1 has a strict incentive to play L. As a result, after observing H in period t, player 2 who arrives in period $t+1$ will be convinced that player 1 is the commitment type. Hence, she has a strict incentive to play T in period $t+1$. If this cycle persists, then a patient player 1's average payoff across the two periods is approximately $\frac{1}{2}(\underline{v}_1 + u_1(a^*, b^*))$ by playing H in every period.[11]

The lesson learnt from the above reasoning process is: in order to motivate the strategic type to play a^*, b_{t+1} needs to vary with a_t. However, the *minimal amount of variation* needed to motivate player 1 decreases when he becomes more patient. This is because player 1 puts more weight on his continuation payoffs relative to his current-period payoff. In particular, the required level of informativeness vanishes to 0 as $\delta \to 1$.

To better understand the connections, I apply the lower bound of Gossner (2011) to the baseline model and explain why it provides an uninformative answer when the informativeness of signals is endogenous. According to Gossner (2011), the sum of KL-divergence (between the probability measure over histories generated by the commitment type, and the equilibrium probability measure) is bounded from above by:

$$-\log \pi_0. \tag{3.10}$$

When a^* is played with probability q^*, the divergence between the probability measure generated by the commitment type and that generated by the equilibrium probability measure is approximately

$$\log\Big(1 + (1-q^*)(1-\delta)\Big). \tag{3.11}$$

Using the Taylor's expansion, the above expression is of the magnitude $(1-\delta)$. As a result, when the

[11] For any $K \in \mathbb{N}$, if b_{t+1} is uninformative about a_t, then player 1 can guarantee an average payoff close to $\frac{K}{K+1}u_1(a^*, b^*) + \frac{1}{1+K}\underline{v}_1$ from period t to $t+K$ by playing a^* in every period. As K converges to infinity, the above guaranteed average payoff converges to his Stackelberg payoff.

strategic player 1 imitates the commitment type, the expected number of periods with which player 2's belief about player 1's action being far away from a^* explodes as $\delta \to 1$. This contrasts to the case with exogenous signals in which the number of such periods is uniformly bounded from above.

4 Informative Signal about Current Period Action

I investigate situations in which each uninformed player can observe an informative signal about the informed player's current period action before making her own action choice. I call this *reputation game with informative signals*, as compared to the baseline model.

Consider the following *sequential-move* stage-game. In period t, player 1 chooses $a_t \in A$ after observing his private history h_1^t. In addition to observing h^t defined in (2.3), player 2 in period t also observes a noisy signal $s_t \in S$, drawn according to distribution $f(\cdot|a_t)$, before choosing $b_t \in B$. Let $\mathbf{f}$ be the stochastic matrix $\{f(\cdot|a)\}_{a\in A}$, which summarizes the signal structure. I introduce the definitions of bounded informativeness and unbounded informativeness, which is introduced by Smith and Sørensen (2000) in social learning models.

Definition 1. *For any given $a^* \in A$,*

1. *$\mathbf{f}$ is unboundedly informative about $a^* \in A$ if there exists $s \in S$ such that $f(s|a) > 0$ iff $a = a^*$.*
2. *$\mathbf{f}$ is boundedly informative about $a^* \in A$ if it is not unboundedly informative about a^*.*

Let $\mathrm{NE}(\delta, \pi_0, K, \mathbf{f})$ be the set of Nash equilibria in the reputation game with public signals. Recall that

$$\mathbb{E}_1^{(\sigma_1,\sigma_2)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t u_1(a_t,b_t)\Big]$$

is the strategic long-run player's equilibrium payoff under strategy profile (σ_1, σ_2). Let

$$\underline{V}_1(\pi_0, K, \mathbf{f}) \equiv \liminf_{\delta\to 1} \inf_{(\sigma_1,\sigma_2)\in\mathrm{NE}(\delta,\pi_0,K,\mathbf{f})} \mathbb{E}_1^{(\sigma_1,\sigma_2)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t u_1(a_t,b_t)\Big].$$

be a patient long-run player's *guaranteed equilibrium payoff*.

4.1 Signals with Bounded Informativeness

I show that if $f(\cdot|a)$ has full support for every $a \in A$, then the reputation failure result in Theorem 1 extends regardless of the statistical precision of $\mathbf{f}$. More generally, if $\mathbf{f}$ is boundedly informative about a^*, then player 1's guaranteed equilibrium payoff is strictly bounded below his Stackelberg payoff.

Corollary 2. *If the stage-game payoffs satisfy Assumptions 1 and 2, and* $\mathbf{f}$ *has full support, then in the reputation game with signals, there exists* $\overline{\pi}_0 \in (0,1)$ *such that for every* $\pi_0 \in (0, \overline{\pi}_0)$ *and* δ *large enough, there exists* $(\sigma_1^\delta, \sigma_2^\delta) \in NE(\delta, \pi_0, K, \mathbf{f})$*, such that:*

$$\mathbb{E}_1^{(\sigma_1^\delta,\sigma_2^\delta)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t u_1(a_t,b_t)\Big] \leq \underline{v}_1.$$

Proof of Corollary 2: Let a' be player 1's action in his worst stage-game Nash equilibrium. Let

$$l^*(\mathbf{f}) \equiv \max_{s \in S} \frac{f(s|a^*)}{f(s|a')}. \tag{4.1}$$

Consider the construction in the proof of Theorem 1 with one modification: the overall probability with which player 1 plays a^* is:

$$\widehat{q} \equiv \frac{q^*}{q^* + (1-q^*)l^*(\mathbf{f})}, \tag{4.2}$$

and the probability with which he plays a' is $1-\widehat{q}$. Let $\overline{\pi}_0 = \widehat{q}^K$, player 2 has an incentive to play b in the reputation building phase, as opposed to b^*, regardless of her observation of player 1's action in the past K periods, and regardless of the signal she receives about player 1's action in the current period. The rest of the proof follows from that of Theorem 1. □

4.2 Signals with Unbounded Informativeness: Binary Action Games

Next, I consider the case in which $\mathbf{f}$ is unboundedly informative about player 1's Stackelberg action a^*. I establish a positive reputation result when player 1's action choice is binary:

Theorem 2. *If the stage-game payoffs satisfy Assumption 1,* $|A| = 2$ *and* $\mathbf{f}$ *is unboundedly informative about the Stackelberg action* a^**, then for every* $K \in \mathbb{N}$ *and* $\pi_0 > 0$*:*

$$\underline{V}_1(\pi_0, K, \mathbf{f}) \geq u_1(a^*, b^*).$$

The proof is in Appendix B. The binary action game studied in Theorem 2 includes the two leading examples that demonstrate reputation effects, namely, the product choice game and the entry deterrence game. It provides a sufficient condition for player 1 to guarantee his commitment payoff when uninformed players have limited memories about the informed player's actions, and they are learning about the informed player's type via their predecessors' actions.

The requirement of unboundedly informative signals is reminiscent of the well-known conclusion in Smith and Sørensen (2000), that players' actions are asymptotically correct *if and only if* their signals are unboundedly informative about the payoff-relevant state. However, establishing a reputation for playing the Stackelberg action is more challenging than aggregating information about an exogenous state. This is because in reputation models, this signal is related to the informed player's type through the informed player's actions, and the latter is endogenously determined in equilibrium. As will be clear in the next subsection, under some adverse belief about the strategic type's behavior (which is very different from the commitment behavior), b_t can be uninformative about a_t although $\mathbf{f}$ is unboundedly informative about a^*.

Theorem 2 implies that in binary-action games, player 1 can overcome the aforementioned challenge and secure his Stackelberg payoff in all equilibria. Compared to games with boundedly informative signals, player 2 has a strict incentive to play b^* after observing the signal realization that only occurs under the Stackelberg action, regardless of her belief about strategic player 1's strategy. In addition, when player 1's action choice is binary, as long as the unconditional probability with which $b_t = b^*$ occurs is bounded away from 1, then the following likelihood ratio:

$$\frac{\Pr(b_t = b^*|a_t = a^*)}{\Pr(b_t = b^*|a_t \neq a^*)},$$

is bounded from below by a number that is strictly above 1. This inequality bounds the informativeness of b_t about a_t from below, which uniformly applies (1) across all discount factors, and (2) across all histories at which player 2 believes (before observing the current period realization of s) that the probability with which she plays b^* is bounded away from 1.

Another challenge arises from the differences in the short-run players' beliefs across different periods, which also occurs in other repeated game models with private monitoring. In particular, short-run players who arrive in different periods have access to different information about player 1's past play. Therefore, it could be the case that b_t is very informative about ω according to the belief of player 2 in period t, but is uninformative from the perspective of player 2 in period $s(> t)$.

I use the following argument to bound the payoff consequences of such differences in beliefs. If player 2 in period t observes that a^* has been played in the past K periods, and believes (before observing s_t) that $b_t = b^*$ with probability at most $1 - \epsilon$, then the probability with which $(a_{t-K}, ..., a_{t-1}) = (a^*, ..., a^*)$ under the equilibrium strategy profile must be bounded from below. This is because otherwise, player 2 in period t believes that the commitment type occurs with probability close to 1, and the probability with which she plays b^* in period t cannot be bounded away from 1. Given that $(a_{t-K}, ..., a_{t-1}) = (a^*, ..., a^*)$ occurs with probability bounded from below, the probability with which player 2 in period s believes that it occurs with very low probability is bounded from above. Therefore, for any given lower bound on b_t's informativeness about ω from the perspective of player 2 in period t, one can derive another lower bound on b_t's informativeness about ω from the perspective of player 2 in period s. The latter lower bound applies with probability close to 1.

4.3 Signals with Unbounded Informativeness: Beyond Binary Actions

Before generalizing Theorem 2 to games in which player 1 has three or more actions, I present two counterexamples highlighting the issues that arise. In particular, s_t can be uninformative about ω despite the probability with which $b_t = b^*$ is bounded away from 1.

Example 1: Consider the following stage game in which player 1 has three actions and player 2 has two actions.

-	b^*	b'
a^{**}	$8, 2$	$2, 0$
a^*	$10, 1$	$6, 0$
a'	$12, -1$	$8, 0$

Let $S \equiv \{s^*, s^{**}, s'\}$. The signal distribution $\mathbf{f}$ is given by $f(s^{**}|a^{**}) = 1$, $f(s'|a') = 1$, $f(s^{**}|a^*) = f(s^*|a^*) = 1/4$ and $f(s'|a^*) = 1/2$. One can check that player 1's Stackelberg action is a^*, the game satisfies Assumptions 1 and 2, and moreover, $\mathbf{f}$ is unboundedly informative about a^*.

Consider the following strategy profile: strategic player 1 mixes between a^{**}, a^*, and a' with equal probabilities in every period. Player 2 plays b^* if $s_t \in \{s^*, s^{**}\}$ and plays b' if $s_t = s'$. This strategy profile is an equilibrium when $\pi_0 < 3^{-K-1}$. Player 1's equilibrium payoff is 8, which is strictly bounded below his Stackelberg payoff 10.

In this example, b_t is uninformative about player 1's type because there are multiple actions of player 1 that can induce player 2 to play b^*. In the example, the two actions are a^* and a^{**}, in which a^{**} leads to an inferior payoff for the long-run player. When the commitment type plays a^{**} with positive probability, the conditional probability of b^* is the same regardless of player 1's type.

Example 2: Consider the following stage game:

-	b^*	b'
a^*	$1, 1$	$-1, 0$
a'	$0, -0.1$	$1, 0$
a''	$2, -10$	$0, 0$

Let $S \equiv \{s^*, s', s''\}$. The signal distribution $\mathbf{f}$ is given by $f(s^*|a^*) = 0.1$, $f(s'|a^*) = 0.4$, $f(s''|a^*) = 0.5$, $f(s'|a') = 1$ and $f(s''|a'') = 1$. Player 1's Stackelberg action is a^*, the game satisfies Assumptions 1 and 2, and $\mathbf{f}$ is unboundedly informative about a^*.

Consider the following strategy profile. Strategic player 1's mixed action only depends on player 2's posterior belief about his type. If player 2's posterior assigns probability π_t to the commitment type, then strategic player 1 plays $\alpha(\pi) \in \Delta(A)$ such that:

$$(1-\pi) \circ \alpha + \pi \circ a^* = 0.5 \circ a^* + 0.25 \circ a' + 0.25 \circ a''.$$

Player 2 plays b^* if $s_t \in \{s^*, s'\}$ and plays b' if $s_t = s''$. Notice that conditional on each type, the probability with which $b_t = b^*$ is $1/2$. This strategy profile is an equilibrium when π_0 is small enough, such that player 2's posterior belief at any history is bounded from above by $1/2$. Player 1's equilibrium payoff is 0, which is strictly bounded below his Stackelberg payoff 1.

In this example, b_t is uninformative about player 1's type because there is *heterogeneity* in player 2's incentive to play b' against different actions of player 1's. In particular, player 2 has stronger incentive to play b' under a'' compared to that under a'. As a result, there exists $\mathbf{f}$ such that player 2 has an incentive to play b^* following a signal realization that leads to a low posterior probability about a^*, and has an incentive to play b' following a signal realization that leads to a high posterior probability about a^*. This situation is implicitly ruled out when $|A| = 2$ since there is one action in A that is not the Stackelberg action, but will occur generically when $|A| \geq 3$.

Resistant to Learning: Motivated by these examples, I introduce the definition of *resistance to learning*, which is a joint condition on $(\mathbf{f}, u_2)$, that characterizes situations in which observing informative signals about the long-run player's current period action (in addition to observing the previous short-run players' actions) is *sufficient* or *insufficient* for the patient long-run player to guarantee his commitment payoff.

Formally, for every $\alpha \in \Delta(A)$, signal distribution $\mathbf{f}$, and $\beta : S \to \Delta(B)$, let $\alpha \cdot \mathbf{f} \cdot \beta \in \Delta(B)$ be the distribution of b when (1) player 1 plays α, (2) the signals are generated according to $\mathbf{f}$, and (3) player 2 behaves according to β after observing s. Abusing notation, I use $a \in A$ and $b \in B$ to denote the Dirac measures on a and b, respectively.

Definition 2. *For any given $a^* \in A$,*

1. *$(\mathbf{f}, u_2)$ is **resistant to learning** against a^* if there exist $\alpha \in \Delta(A)$ with $a^* \in supp(\alpha)$, and*

$\beta : S \to \Delta(B)$ which is a best reply against α under u_2, such that:

$$\alpha \cdot \mathbf{f} \cdot \beta = a^* \cdot \mathbf{f} \cdot \beta \neq BR_2(a^*). \tag{4.3}$$

2. *$(\mathbf{f}, u_2)$ is* ***not resistant to learning*** *against a^* if for every $\alpha \in \Delta(A)$ with $a^* \in supp(\alpha)$, and $\beta : S \to \Delta(B)$ which is a best reply against α under u_2,*

$$\alpha \cdot \mathbf{f} \cdot \beta = a^* \cdot \mathbf{f} \cdot \beta \quad \textit{implies} \quad \alpha \cdot \mathbf{f} \cdot \beta = a^* \cdot \mathbf{f} \cdot \beta = BR_2(a^*). \tag{4.4}$$

By definition, for every $u_2, \mathbf{f}$ and a^*, either $(\mathbf{f}, u_2)$ is resistant to learning against a^*, or $(\mathbf{f}, u_2)$ is not resistant to learning against a^*. Intuitively, resistant to learning implies that player 2 is not playing the complete information best reply against a^*, and moreover, her action choices are uninformative about the long-run player's type under some belief about the long-run player's actions α, and some of her reply β against α. On the other hand, not resistant to learning implies that as long as player 2's action distribution cannot distinguish between a^* and some other action distribution of player 1's, player 1 can induce player 2 to play b^* with probability 1 by playing a^*.

Applying the resistant to learning definition to some of my previous results, if $\mathbf{f}$ is boundedly informative about a^*, and player 2's best reply depends on player 1's action, then $(\mathbf{f}, u_2)$ is resistant to learning against a^*. If $\mathbf{f}$ is unboundedly informative about a^* and player 1's action choice is binary, then $(\mathbf{f}, u_2)$ is not resistant to learning against a^*. In the two counterexamples of this subsection, although $\mathbf{f}$ is unboundedly informative about a^*, $(\mathbf{f}, u_2)$ is resistant to learning against a^*, which leads to failures to build reputations. My next theorem generalizes these insights by connecting resistant to learning with the success or failure of reputation building:

Theorem 3. *If $(\mathbf{f}, u_2)$ is not resistant to learning against a^*, then for every u_1 that satisfies Assumption 1, $K \in \mathbb{N}$ and $\pi_0 > 0$:*

$$\underline{V}_1(\pi_0, K, \mathbf{f}) \geq u_1(a^*, BR_2(a^*)).$$

If $(\mathbf{f}, u_2)$ is resistant to learning against a^*, then there exist $\overline{\pi}_0 > 0$ as well as an open set of u_1, such that for every u_1 within this open set, a^* is player 1's Stackelberg action, but for every $\pi_0 < \overline{\pi}_0$ and δ large enough, there exists $(\sigma_1^\delta, \sigma_2^\delta) \in NE(\delta, \pi_0, K, \mathbf{f})$ such that:

$$\mathbb{E}_1^{(\sigma_1^\delta, \sigma_2^\delta)} \Big[\sum_{t=0}^{\infty} (1-\delta)\delta^t u_1(a_t, b_t) \Big] \leq \underline{v}_1.$$

The proof of Theorem 3 is in Appendix C. The requirement that $K \geq 1$ is needed for the second statement to hold under an open set of u_1. Intuitively, this is because when $K = 0$, player 1's action in the current period cannot directly affect player 2's actions in the future. In order to motivate the strategic type to play α that makes b_t uninformative about ω, player 1 needs to be indifferent in the stage game, which can happen only under knife-edge payoff functions.

To better understand how to apply Theorem 3, I provide sufficient conditions on the primitives for *resistent to learning* and *not resistent to learning.* I start from introducing a regularity condition on u_2 that captures the hetergeneity in player 2's propensity to play b^*.

Definition 3 (Admissibility)**.** *u_2 is admissible if*

1. *$u_2(a, b) \neq u_2(a, b')$ for every $a \in A$ and $b \neq b'$.*
2. *there exists $a, a' \in A$ such that $BR_2(a) \neq BR_2(a')$.*
3. *for every $a' \neq a''$ and $b' \neq b''$, $u_2(a', b') - u_2(a', b'') \neq u_2(a'', b') - u_2(a'', b'')$.*

The first two requirements are already implied by Assumptions 1 and 2. The third requirement is novel, which says that player 2's gain from playing b' instead of b'' depends on player 1's action choice. This third condition is generic, and is satisfied, for example, when A and B are ordered sets and u_2 has strictly increasing differences in a and b. This leads to the following result:

Lemma 4.1. *When $|A| \geq 3$, for every $a^* \in A$ and every admissible u_2, there exists $\mathbf{f}$ that is (1) unboundedly informative about a^*, but (2) $(\mathbf{f}, u_2)$ is resistent to learning against a^*.*

Theorem 3 and Lemma 4.1 together imply the following corollary:

Corollary 3. *When $|A| \geq 3$, for every admissible u_2, there exist u_1 that satisfies Assumptions 1 and 2, $\mathbf{f}$ that is unboundedly informative about a^*, and $\overline{\pi}_0 \in (0, 1)$, such that for every $\pi_0 < \overline{\pi}_0$ and δ large enough, there exists $(\sigma_1^\delta, \sigma_2^\delta) \in NE(\delta, \pi_0, K)$, such that:*

$$\mathbb{E}^{(\sigma_1^\delta, \sigma_2^\delta)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t u_1(a_t, b_t)\Big] \leq \underline{v}_1.$$

Proofs of Lemma 4.1 and Corollary 3: For every $a^* \in A$ and admissible u_2, let b^* be the unique element in $\mathrm{BR}_2(a^*)$. Set $u_1(a^*, b^*) = 1$, and $u_1(a^*, b) = 0$ for all $b \neq b^*$. Since u_2 is admissible, there exist $\alpha \in \Delta(A)$ and $b' \neq b^*$ such that:

1. α has full support on A,

2. $\mathrm{BR}_2(\alpha) = \{b^*, b'\}$.

From the second and third requirement on admissibility and the assumption that $|A| \geq 3$, there exist $a', a'' \in A \backslash \{a^*\}$ such that:

$$u_2(a', b') - u_2(a', b^*) < u_2(a'', b') - u_2(a'', b^*), \tag{4.5}$$

and $u_2(a'', b') - u_2(a'', b^*) > 0$.[12] For every $g \in (0, 1)$, consider the following signal structure f with three signal realizations $S \equiv \{s^*, s', s''\}$:

1. $f(s^*|a^*) = \epsilon_1$, $f(s'|a^*) = g - \epsilon_1$ and $f(s''|a^*) = 1 - g$.

2. $f(s'|a') = g + \epsilon_2\alpha(a'')$ and $f(s''|a') = 1 - g - \epsilon_2\alpha(a'')$.

3. $f(s'|a'') = g - \epsilon_2\alpha(a')$ and $f(s''|a'') = 1 - g + \epsilon_2\alpha(a')$.

4. $f(s'|a) = g$ and $f(s''|a) = 1 - g$ for all $a \notin \{a^*, a', a''\}$.

When both ϵ_1 and ϵ_2 are small enough, player 2's best reply following any signal realization is either b^* or b'. When ϵ_2 is relatively large compared to ϵ_1, player 2 has an incentive to play b^* after observing s^* or s', and has an incentive to play b' after observing s''. Under this information structure, if player 1 plays the mixed action α, player 2 plays b^* with probability g and b' with probability $1 - g$; if player 1 plays a^*, player 2 plays b^* with probability g and b' with probability $1 - g$. Find such ϵ_1 and ϵ_2, one can then complete the construction of u_1.

1. $u_1(a', b^*)$ and $u_1(a', b')$ are such that

$$(g + \epsilon_2\alpha(a''))u_1(a', b^*) + (1 - g - \epsilon_2\alpha(a''))u_1(a', b') = g.$$

2. $u_1(a'', b^*)$ and $u_1(a'', b')$ are such that first, $u_1(a'', b^*) > 1$; and second,

$$(g - \epsilon_2\alpha(a'))u_1(a'', b^*) + (1 - g + \epsilon_2\alpha(a'))u_1(a'', b') = g.$$

3. For every $a \notin \{a^*, a', a''\}$, $u_1(a, b^*)$ and $u_1(a, b')$ are such that

$$gu_1(a, b^*) + (1 - g)u_1(a, b') = g.$$

[12] This is because $u_2(a^*, b') - u_2(a^*, b^*) < 0$, and player 2's ordinal preference between b' and b^* depends on a according to the second requirement.

4. When $b \notin \{b^*, b'\}$, set $u_1(a,b)$ to be negative for every $a \in A$.

As a result, when π_0 is small enough, the following strategy profile is an equilibrium for every δ: player 1 plays α in every period, and player 2 chooses $b_t = b^*$ after observing $s_t \in \{s^*, s'\}$, and chooses $b_t = b'$ after observing $s_t = s''$. Player 1's equilibrium payoff is g, which is strictly below his Stackelberg payoff 1. □

Next, I focus on stage-games that have monotone-supermodular payoffs (Assumption 3). Recall that in monotone-supermodular games, players' actions can be ranked according to $(A, \succ_a)$ and $(B, \succ_b$). I show that player 1 can guarantee his commitment payoff from playing his highest action whenever **f** that is unbounded informative about his highest action and possesses the standard *monotone likelihood ratio property* (or MLRP for short).

Definition 4. **f** *has MLRP if there exists a ranking on S, denoted by $\succ_s$, such that for every $a \succ a'$ and $s \succ s'$,*

$$\frac{f(s|a)}{f(s'|a)} \geq \frac{f(s|a')}{f(s'|a')}. \tag{4.6}$$

Intuitively, under ranking $\succ_s$ of the signal realizations, higher signals are more likely to occur under higher actions of the informed player. Let $\overline{a} \equiv \max A$.

Lemma 4.2. *If the stage-game payoffs satisfy Assumptions 1 and 3, and* **f** *is unboundedly informative about $\overline{a}$ and satisfies MLRP, then for every $K \in \mathbb{N}$ and $\pi_0 > 0$:*

$$\underline{V}_1(\pi_0, K, \mathbf{f}) \geq u_1\Big(\overline{a}, BR_2(\overline{a})\Big).$$

Proof of Lemma 4.2: Since **f** is unboundedly informative about $\overline{a}$, there exists $s^* \in S$ such that $f(s^*|a) > 0$ if and only if $a = \overline{a}$. Since $\overline{a}$ is player 1's highest action, the MLRP implies that s^* is the highest signal realization. For every distribution over player 1's actions $\alpha \in \Delta(A)$, there exists $s' \in S$ such that player 2 has an incentive to play $b^* \equiv \mathrm{BR}_2(\overline{a})$ if and only if $s \succeq s'$, and has a strict incentive not to play b^* otherwise. The probability with which $s \succeq s'$ is higher under a^* than any other action $\alpha \in \Delta(A)$. As a result, the probability of b^* is strictly higher under a^* than under α, as long as this probability is not 1. The lower bound on a patient player 1's equilibrium payoff follows from Theorem 3. □

5 Conclusion

This paper highlights the challenges to reputation building when uninformed players learn primarily through *endogenous signals*. An example of which is their predecessors' actions whose informativeness is endogenously determined and decreases with the informed player's discount factor. From the perspective of applications, my analysis captures buyers' suspicion after observing sellers' consummate behaviors. It provides an explanation for why such suspicion can persist over time and can have payoff consequences for patient players. I also provide sufficient conditions for effective reputation building through endogenous signals. My resistant to learning condition relates to, albeit different from, the unbounded informativeness condition in models of observational learning. In particular, it takes into account players' responses when the object to be learnt is endogenous to the equilibrium.

A Appendix: Proof of Proposition 2

Recall that the product choice game has the following stage-game payoff:

$-$	T	N
H	$1, 1$	$-1, 0$
L	$2, -1$	$0, 0$

In this game, the cutoff belief above which player 2 plays T is $q^* \equiv 1/2$. I construct an equilibrium in which patient strategic long-run player's payoff is 0 for all large enough δ. The equilibrium consists of three phases, a *reputation building phase*, a *reputation maintenance phase* and a *punishment phase.* I describe players' behaviors in the three phases one by one.

Punishment Phase: If player 2 in period t observes that $(b_0, ..., b_{t-1})$ is such that there exists N that occurs after T, then since $(b_0, ..., b_{t-1})$ is common knowledge among players, they coordinate on the stage-game Nash equilibrium (L, N) in all subsequent periods.

Reputation Maintenance Phase: If player 2 in period t observes that $(b_0, ..., b_{t-1})$ is such that T has occurred before, and N has not occurred after T, then let t^* be the first period with which T occurs. Strategic player 1 plays H with probability 1 if $(b_0, ..., b_{t-1})$ satisfies the above conditions and L has not been played from period $t^* + 1$ to $t - 1$; and plays H with probability q^* if L has been played from period $t^* + 1$ to $t - 1$. Player 2 plays T with probability 1 if at least one of the following three conditions is satisfied: first, $t - 1 = t^*$; second, $t - 1 \notin N_t$; third, $t - 1 \in \mathcal{N}_t$ and $a_{t-1} = H$. If $t - 1 > t^*$, $t - 1 \in N_t$, and $a_{t-1} = L$, then she mixes between T and N. The probability with which she plays N, denoted by p_t, satisfies:

$$\frac{1-\delta}{\delta} = p_t \Pr(t - 1 \in \mathcal{N}_t). \tag{A.1}$$

Since $\Pr(t - 1 \in \mathcal{N}_t)$ is bounded away from 0, p_t is strictly between 0 and 1 when δ is large enough.

Reputation Building Phase: Let $M \in \mathbb{N}$ be a large enough integer such that for every $n \geq M$, we have:

$$2^K \sum_{j=0}^{K} \binom{n}{j} < 2^n - 1. \tag{A.2}$$

Such an M exists due to two observations, namely, for any integer $n \in \mathbb{N}$, $\binom{n}{j}$ is increasing in j when $j < n/2$ and is decreasing in j when $j > n/2$; and moreover, $\sum_{j=0}^{n} \binom{n}{j} = 2^n$. Pick $\overline{\pi}_0 \in (0, 1)$ small enough such that:

$$\frac{\overline{\pi}_0}{1-\overline{\pi}_0}\Big(\frac{1}{q^*/2}\Big)^M < \frac{q^*/2}{1-q^*/2}. \tag{A.3}$$

I consider two subphases separately, depending on the comparison between calendar time t and M.

When $t \leq M$, strategic player 1 plays H with probability q^* at private histories such that L has been played before. At histories such that L has not been played before, the probability with which he plays H, denoted by β_t, is defined recursively via:

$$\beta_t \Pr(\omega^c|H,H,...,H) + \Big(1-\Pr(\omega^c|H,H,...,H)\Big) = q^*. \tag{A.4}$$

Such $\beta_t \in (0,1)$ exists and is greater than $q^*/2$ according to the upper bound on $\overline{\pi}_0$ in (A.3). Player 2 plays T with probability $\frac{1-\delta}{\delta(2-\delta)p_t}$ in period t if $t \geq 1$, $t-1 \in N_t$, and $a_{t-1} = H$, where $p_t \equiv \Pr(t-1 \in \mathcal{N}_t)$. Player 2 plays N with probability 1 in period t otherwise.

When $t > M$, let $\beta(h_1^t)$ be the probability with which strategic type player 1 plays H at private history h_1^t. I fix $\beta(H,H,...,H)$ to be 0. For every private history of player 2's h_2^t, let $\boldsymbol{\kappa(h_2^t)} \in \Delta(\mathcal{H}_1^t)$ be her belief about player 1's private history, and let $\pi(h_2^t) \in [0,1]$ be the probability she attaches to the commitment type. Let $\pi(h_1^t) \in [0,1]$ be the probability that an outside observer attaches to the commitment type if he shares the same prior belief as the short-run players and observes player 1's private history h_1^t. Let $\boldsymbol{\beta^t}$ be an $|\mathcal{H}_1^t|$-dimensional vector defined as:

$$\boldsymbol{\beta^t} \equiv \begin{cases} \beta(h_1^t) & \text{if } h_1^t \neq (H,H,...,H) \\ \pi(h_1^t) & \text{if } h_1^t = (H,H,...,H). \end{cases} \tag{A.5}$$

In what follows, I compute and define $\boldsymbol{\beta^t}$, $\boldsymbol{\kappa(h_2^t)}$, $\pi(h_2^t)$, $\pi(h_1^t)$, and players' behaviors in the reputation building phase after period M recursively. For every $t \geq M+1$, given players' behaviors from period 0 to $t-1$, as well as the distribution over player 2's neighborhood $\mathcal{N}_t$, one can compute $\boldsymbol{\kappa(h_2^t)}$, $\pi(h_2^t)$, and $\pi(h_1^t)$ according to Bayes Rule.

Given (A.3) and (A.4), and the assumption that $\Pr(|\mathcal{N}_t| \leq K) = 1$, we know that $\pi(h_2^t)$ is bounded from above by $q^*/2$ for every h_2^t that occurs with positive probability. Moreover, the probability with which $\boldsymbol{\kappa(h_2^t)}$ attaches to $(H,H,...,H)$ is bounded from above $q^*/2$. This is because conditional on $a_s = H$ for all $s \in \mathcal{N}_t$, $h_1^t = (H,H,...,H)$ if and only if $\omega = \omega^c$. As a result, the probability player 2 attaches to player 1's private history being $(H,H,...,H)$ is bounded from above by the probability she attaches to commitment type at private history h_2^t. I choose the other entries of $\boldsymbol{\beta}$, aside from the one for private history $(H,H,...,H)$ that is fixed to be $\pi(h_1^t)$, such that each of these aforementioned entries is between $q^*/2$ and 1, and moreover:

$$\boldsymbol{\kappa}(\boldsymbol{h_2^t}) \cdot \boldsymbol{\beta^t} = q^* \text{ for every } h_2^t \in \mathcal{H}_2^t. \tag{A.6}$$

The above linear system admits at least one solution for the following reasons:

1. The probability with which $\boldsymbol{\kappa}(\boldsymbol{h_2^t})$ attaches to $(H, H, ..., H)$ is bounded from above $q^*/2$.

2. Since each player 2 can observe at most K of her predecessors' interactions with player 1, the cardinality of $\mathcal{H}_2^t$ is at most $2^K \sum_{j=0}^{K} \binom{t}{j}$, which corresponds to the number of linear constraints; The cardinality of $\mathcal{H}_1^t$ is 2^t, namely, one can choose $2^t - 1$ free variables.

   According to the construction of M in (A.2), the number of free variables is strictly larger than the number of linear constraints when $t \geq M$.

For every $t > M$ and given that play remains in the reputation-building phase, player 2 plays T with probability $\frac{1-\delta}{\delta(2-\delta)p_t}$ in period t if $t-1 \in N_t$, and $a_{t-1} = H$, where $p_t \equiv \Pr(t-1 \in \mathcal{N}_t)$. Player 2 plays N with probability 1 in period t otherwise.

One can verify that the strategic player 1 is indifferent between H and L at every history in the reputation building phase since his continuation payoff is $\frac{1-\delta}{\delta}$ at every private history such that $t \geq 1$, $t-1 \in N_t$, and $a_{t-1} = H$. His continuation payoff is 0 at other private histories. Player 2 is indifferent between T and N at every history of the reputation building phase given (A.6).

B Appendix: Proof of Theorem 2

For every public history h^t, let $g(h^t)$ be the probability with which player 2 plays b^* at h^t. Let $g(h^t, \omega^c)$ be the probability with which player 2 plays b^* at h^t conditional on player 1 is the commitment type. For any public history h^t such that

$$\{a_{\max\{0,t-K\}}, ..., a_{t-1}\} = \{a^*, ..., a^*\},$$

namely, player 2's belief at h^t (before observing s_t) attaches positive probability to the commitment type, I derive a lower bound on:

$$\frac{g(h^t, \omega^c)}{g(h^t)},$$

as a function of $g(h^t)$, or equivalently, an upper bound on

$$\frac{1-g(h^t,\omega^c)}{1-g(h^t)}. \tag{B.1}$$

Let $A \equiv \{a^*, a'\}$ and $S \equiv \{s^*, s_1, s_2, ..., s_m\}$. Let $r(h^t)$ be the probability that a^* is played at h^t, let $\tau(s_i)(h^t)$ be the probability that signal s_i occurs at h^t, and let $p(s_i)(h^t)$ be the posterior probability of a^* conditional on observing s_i at h^t.[13] I suppress the dependence on h^t in order to simplify notation. Since $\{b^*\} = \mathrm{BR}_2(a^*)$ and $|A| = 2$, we have the following two implications:

1. there exists a cutoff belief $p^* \in (0,1)$ such that player 2 has a strict incentive to play b^* after observing s_i if and only if $p(s_i) > p^*$.

2. there exists a constant $C \in \mathbb{R}_+$ such that $1 - r \geq C(1-g)$.[14]

According to the first implication, it is without loss of generality to label the signal realizations such that $p(s_1) \geq p(s_2) \geq ... \geq p(s_m)$, and moreover, there exists $k \in \{1, 2, ..., m\}$ such that player 2 plays b^* for sure after observing $s_1, ..., s_{k-1}$, and does not play b^* otherwise.[15] Therefore,

$$r(1-f(s^*|a^*)) = \sum_{i=1}^{m} \tau(s_i)p(s_i), \quad 1-r = \sum_{i=1}^{m} \tau(s_i)(1-p(s_i)), \quad \text{and } \sum_{i=k}^{m} \tau(s_i) = 1-g.$$

Using the fact that $p(s_1) \geq p(s_2) \geq ... \geq p(s_m)$, we know that:

$$\frac{\sum_{i=1}^{k-1} \tau(s_i)p(s_i)}{\sum_{i=1}^{k-1} \tau(s_i)(1-p(s_i))} \geq \frac{r(1-f(s^*|a^*))}{1-r} \geq \frac{\sum_{i=k}^{m} \tau(s_i)p(s_i)}{\sum_{i=k}^{m} \tau(s_i)(1-p(s_i))}. \tag{B.2}$$

As a result,

$$\sum_{i=k}^{m} \tau(s_i)p(s_i) \leq \frac{r(1-f(s^*|a^*))}{1-rf(s^*|a^*)}(1-g), \tag{B.3}$$

and

$$\sum_{i=k}^{m} \tau(s_i)(1-p(s_i)) \geq \frac{1-r}{1-rf(s^*|a^*)}(1-g). \tag{B.4}$$

Therefore,

$$\frac{1-g(\omega^c)}{1-g} \leq \frac{1-f(s^*|a^*)}{1-rf(s^*|a^*)}, \tag{B.5}$$

[13] Notice that r, τ, p depend on player 1's action choice at h^t, which is endogenously determined in equilibrium.

[14] This is implied by the results on Bayesian persuasion once player 1's action at h^t is viewed as the state. The probability with which b^* not being played leads to an upper bound on the probability with which state a^* occurs.

[15] Ignoring the possibility that player 2 plays a mixed action following certain signal realizations is without loss of generality in proving the current theorem. This is because when player 2 mixes between n actions after one signal realization, we can split this signal realization into n signal realizations with the same posterior belief, such that player 2 plays a pure action following each of these signal realizations.

Using the second implication, namely, $r \leq 1 - C(1-g)$, we have:

$$\frac{1-g(\omega^c)}{1-g} \leq \frac{1-f(s^*|a^*)}{1-f(s^*|a^*)+Cf(s^*|a^*)(1-g)}. \tag{B.6}$$

Similarly, the lower bound on the likelihood ratio with which b^* occurs is given by:

$$\frac{g(\omega^c)}{g} \geq 1 + \frac{f(s^*|a^*)(1-g(h^t))}{g - rf(s^*|a^*)} \geq 1 + \frac{f(s^*|a^*)(1-g)}{g - f(s^*|a^*)(1-C(1-g))} \tag{B.7}$$

Let $\beta(h^t) \in \Delta(B)$ be the distribution over player 2's action at h^t, and let $\beta(h^t, \omega^c) \in \Delta(B)$ be the distribution over player 2's action at h^t conditional on player 1 being the commitment type. Inequalities (B.6) and (B.7) imply the following lower bound on the KL divergence between $\beta(h^t)$ and $\beta(h^t, \omega^c)$:

$$d\Big(\beta(h^t)\Big|\beta(h^t,\omega^c)\Big) \leq \mathcal{L}\big(1-g(h^t)\big), \tag{B.8}$$

with $\mathcal{L}(\cdot)$ vanishing to 0 as $1 - g(h^t) \to 0$.

This lower bound on the KL divergence bounds the speed of learning at h^t from below, as a function of the probability with which player 2 at h^t does not play b^*. This implies a lower bound on the speed of learning when player 2 in the future observes b^* in period t, *given that he knew* that the probability with which player 2 plays b^* at h^t is no more than $g(h^t)$. However, unlike models with unbounded memory, future player 2's information does not nest that of player 2's in period t. This is because future player 2s may not observe $\{a_{t-K}, ..., a_{t-1}\}$, and hence, cannot interpret the meaning of b_t in the same way as player 2 in period 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 who arrives in period s, as a function of the informativeness of b_t (about player 1's type) from the perspective of player 2 who arrives in period t. This together with (B.8) establishes a lower bound on the informativeness of b_t from the perspective of future player 2s as a function of the probability with which b^* is not being played. Applying the result in Gossner (2011), one obtains the commitment payoff theorem.

Let $\pi(h^t)$ be player 2's belief about ω at h^t before observing the period t signal s_t. By definition, $\pi(h^0) = \pi_0$. For every strategy profile σ, let $\mathcal{P}^\sigma$ be the probability measure over $\mathcal{H}$ induced by σ, let $\mathcal{P}^{\sigma,\omega^c}$ be the probability measure induced by σ conditional on player 1 being the commitment type, and let $\mathcal{P}^{\sigma,\omega^s}$ be the probability measure induced by σ conditional on player 1 being the strategic type. One can the write the posterior likelihood ratio as the product of the likelihood ratio of the signals observed in each period:

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

$$= \frac{\mathcal{P}^{\sigma,\omega^c}(b_0)}{\mathcal{P}^{\sigma,\omega^s}(b_0)} \cdot \frac{\mathcal{P}^{\sigma,\omega^c}(b_1|b_0)}{\mathcal{P}^{\sigma,\omega^s}(b_1|b_0)} \cdot ... \cdot \frac{\mathcal{P}^{\sigma,\omega^c}(b_{t-1}|b_{t-2},...,b_0)}{\mathcal{P}^{\sigma,\omega^s}(b_{t-1}|b_{t-2},...,b_0)} \cdot \frac{\mathcal{P}^{\sigma,\omega^c}(a_{t-K},...,a_{t-1}|b_t,b_{t-1},...,b_0)}{\mathcal{P}^{\sigma,\omega^s}(a_{t-K},...,a_{t-1}|b_t,b_{t-1},...,b_0)} \tag{B.9}$$

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

$$\mathcal{P}^{\sigma,\omega^c}\Big(\pi^\sigma(b_0,b_1,...b_{t-1}) < \epsilon\pi_0\Big) \leq \epsilon\frac{1-\pi_0}{1-\pi_0\epsilon}, \tag{B.10}$$

in which $\pi^\sigma(b_0,b_1,...b_{t-1}) \in \Delta(\Omega)$ 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 $\rho^*(\epsilon)$ be defined as:

$$\rho^*(\epsilon) \equiv \frac{\epsilon\pi_0}{1-C\epsilon}. \tag{B.11}$$

Next, I show that if:

1. $\pi^\sigma(b_0,b_1,...b_{t-1}) \geq \epsilon\pi_0$,
2. player 2 in period t believes that $b_t = b^*$ occurs with probability less than $1-\epsilon$ after observing $(a_{t-K},...,a_{t-1}) = (a^*,...,a^*)$,

then under probability measure $\mathcal{P}^\sigma$, the probability of $\{a_{t-K},...,a_{t-1}\} = \{a^*,...,a^*\}$ conditional on $(b_0,...,b_{t-1})$ is at least $\rho^*(\epsilon)$.

To see this, 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.11), 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 in period s in the event that $\pi^\sigma(b_0,b_1,...b_{t-1}) \geq \epsilon\pi_0$. Let $\mathcal{P}(\sigma,t,s) \in \Delta(\Delta(A^K))$ be player 2's signal structure in period $s(\geq t)$ about $\{a_{t-K},...,a_{t-1}\}$ under equilibrium σ. For every small enough $\eta > 0$, given that $\mathcal{P}(\sigma,t)$ attaches probability at least $\rho^*(\epsilon)$ to $\{a_{t-K},...,a_{t-1}\} = \{a^*,...,a^*\}$, the probability with which $\mathcal{P}(\sigma,t,s)$ attaches to the event that $\{a_{t-K},...,a_{t-1}\} = \{a^*,a^*,...,a^*\}$ occurs with probability less than $\eta\rho^*(\epsilon)$

conditional on $\{a_{t-K}, ..., a_{t-1}\} = \{a^*, a^*, ..., a^*\}$ is bounded from above by:

$$\frac{\eta\rho^*(\epsilon)(1-\rho^*(\epsilon))}{(1-\eta\rho^*(\epsilon))\rho^*(\epsilon)} = \eta\frac{1-\rho^*(\epsilon)}{1-\rho^*(\epsilon)\eta}. \tag{B.12}$$

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. The conclusions in (B.6) and (B.7) also apply in this setting, namely,

$$\frac{1-g(t,\omega^c|h^s)}{1-g(t|h^s)} \leq \frac{1-f(s^*|a^*)}{1-f(s^*|a^*)+Cf(s^*|a^*)(1-g(t|h^s))} \tag{B.13}$$

and

$$\frac{g(t,\omega^c|h^s)}{g(t|h^s)} \geq 1+\frac{f(s^*|a^*)(1-g(t|h^s))}{g(t|h^s)-f(s^*|a^*)(1-C(1-g(t|h^s)))} \tag{B.14}$$

Whenever player 2 in period s believes that $\{a_{t-K}, ..., a_{t-1}\} = \{a^*, a^*, ..., a^*\}$ occurs with probability more than $\eta \cdot \rho^*(\epsilon)$, we have:

$$g(t|h^s) \leq 1 - \epsilon\eta\rho^*. \tag{B.15}$$

Applying (B.15) to (B.13) and (B.14), 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 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 $\underline{\delta}$ such that when $\delta > \underline{\delta}$, the strategic player 1's payoff by playing a^* in every period is at least:

$$\Big(1-\epsilon-\epsilon\frac{1-\pi_0}{1-\pi_0\epsilon}\Big)u_1(a^*,b^*)+\Big(\epsilon+\epsilon\frac{1-\pi_0}{1-\pi_0\epsilon}\Big)\min_{a,b}u_1(a,b)-\epsilon. \tag{B.16}$$

Taking $\epsilon \to 0$ and $\delta \to 1$, (B.16) implies the commitment payoff theorem.

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] Banerjee, Abhijit and Drew Fudenberg (2004) "Word-of-mouth Learning," *Games and Economic Behavior*, 46, 1-22.

[5] Bar-Isaac, Heski and Steven Tadelis (2008) "Seller Reputation," *Foundations and Trends in Microeconomics.*

[6] Deb, Rahul, Matthew Mitchell, and Mallesh Pai (2019) "Our Distrust is Very Expensive," Working Paper.

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

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

[9] 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.

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

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

[12] Gossner, Olivier (2011) "Simple Bounds on the Value of a Reputation," *Econometrica*, 79(5), 1627-1641.

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

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

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

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

[17] 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.

[18] Pei, Harry (2018) "Reputation Effects under Interdependent Values," Working Paper.

[19] Pei, Harry (2019) "Trust and Betrayals: Reputational Payoffs and Behaviors without Commitment," Working Paper.

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

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

[22] Sobel, Joel (1985) "A Theory of Credibility," *Review of Economic Studies*, 52(4), 557-573.