

Pei, Harry

Working Paper

Reputation building with endogenous speed of learning

CSIO Working Paper, No. 0146

Provided in Cooperation with:

Department of Economics - Center for the Study of Industrial Organization (CSIO),
Northwestern University

Suggested Citation: Pei, Harry (2019) : Reputation building with endogenous speed of learning, CSIO Working Paper, No. 0146, Northwestern University, Center for the Study of Industrial Organization (CSIO), Evanston, IL

This Version is available at:

<https://hdl.handle.net/10419/274037>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Reputation Building with Endogenous Speed of Learning

Harry PEI*

Preliminary. Comments Welcome.

July 19, 2019

Abstract: I study reputation models in which information about the long-run player's past behavior is dispersed among short-run players. I identify two challenges to reputation building when such information is aggregated via the short-run players' actions. First, when the long-run player's action can only affect short-run players' future behaviors, the informativeness of short-run players' actions decreases when the long-run player becomes more patient. This leads to equilibria in which both players receive low payoffs. Second, 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 fails. 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 payoff. When the environment is not resistant to learning, the patient long-run player can secure his commitment payoff in all equilibria.

Keywords: dispersed information, reputation failure, endogenous signals, limited memory, informativeness, information aggregation

JEL Codes: C73, D82, D83

1 Introduction

Reputations are powerful tools to overcome lack-of-commitment problems. This intuition is formalized in Fudenberg and Levine (1989,1992), which shows 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 a number of settings. This is especially the case in many developing countries, where mistrust between business partners, lack of high-quality brands, and low government credibility are major obstacles for growth and development.

*Department of Economics, Northwestern University. harrydp@northwestern.edu

For example, the inability of central banks to convince citizens about their intentions (to lower inflation) reduces the effectiveness of monetary policies. Similarly, consumers' skepticism about product quality leads to low returns from a reputation for quality, which reduces firms' incentives to enhance their quality standards. The resulting lack of the supply of high quality products makes the consumers' beliefs self-fulfilling and hurts social welfare.¹

A common theme in these examples is the inability of the reputation building player to signal his future intentions via his current period behavior. These observations are inconsistent with the logic behind the canonical reputation results, which suggest that whenever consumers are skeptical about a seller's product quality, they will be *surprised* after observing the seller providing high quality and the probability with which they attach to the seller being committed goes up.

Motivated by these episodes of reputation failure, I study a reputation model where information about an informed player's past actions is *dispersed* among uninformed players. In my model, a patient long-run player (he, player 1, e.g., seller) interacts with an infinite sequence of myopic short-run players (she, player 2, 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 occurs with small but positive probability.

Different from the canonical reputation models (Fudenberg and Levine 1989, 1992), every short-run 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.² 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 fairly precise 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 more time and effort: a buyer needs to read the online reviews carefully or to ask their friends more detailed questions. It is usually the case that a buyer has limited capacity to acquire and process these detailed information.

Theorem 1 shows that under generic payoff functions, the patient long-run player's guaranteed

¹In the Chinese watermelon market, consumers are unwilling to pay high prices for high-quality melons, and future consumers are reluctant to reward the seller despite having observed high quality in the past (Bai 2018). In the market for malaria drugs, consumers purchase fake drugs despite effective ones are available at a lower price (Nygqvist, Svensson and Yanagizawa-Drott 2018).

²In many games of interest such as the product choice game, my result generalizes to stochastic sampling, namely, each short-run player observes a stochastic subset of her predecessors' actions.

equilibrium payoff is no more than his Nash equilibrium payoff in the stage-game.³ 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 commitment payoff in all equilibria.

Intuitively, such a distinction is driven by the differences between learning through *exogenous signals* (such as the signals in Fudenberg and Levine 1989, 1992) whose informativeness about the long-run player's action is fixed, 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 the stage-game payoffs and the discount factor.

I explain the ideas 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 two phases. Play starts from a *reputation building phase*, in which buyers do not trust and the seller mixes between high and low effort.⁴ 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. This is enforced via trigger strategies, namely, as soon as the seller cheats, the buyer in the next period chooses not to trust. Upon observing the non-trusting action following the trusting action, all future buyers know that the seller has cheated before and punish him by playing the non-trusting action. The transition probability is such that the strategic seller is indifferent between high and low effort in the reputation building phase.

In summary, although the buyers' actions in the reputation building phase can statistically identify the seller's action, their informativeness decreases when the seller becomes more patient. This result extends to more general sampling processes, such as each short-run player observes a stochastic subset of the long-run player's past actions. This extension is motivated by consumers' learning processes in practice. As a byproduct, it also clarifies the robustness of my result when

³My result applies to all games in which (1) players' stage-game payoff functions are generic in the sense that each player receives different payoffs under different action profiles, and (2) there exists a pure strategy Nash equilibrium in the stage-game.

⁴The probability of high effort is low enough so that buyers have incentives to play the non-trusting action.

the long-run player do not know which of his past actions the short-run players can observe.⁵

The economic mechanism identified above accounts for many episodes of reputation failures in practice, some of which have been documented empirically. In particular, rational buyers 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 imitates the commitment type reduces their willingness to trust upon observing the seller's commitment action. This reduces the informativeness of the buyer's actions, lowers the speed of learning of future buyers, which in turn lowers the seller's returns from reputation building.⁶

In the second part of the 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. In this environment, the long-run player can influence the short-run player's current period action. As a result, he may have an incentive to play the Stackelberg action even when his action has no direct impact on his future payoffs. This fits into situations in which a seller produces the good in advance, and potential buyers can inspect the product, observe a noisy signal about its quality, before making their purchasing decisions.

If the informativeness of this signal is bounded, then the long-run player's guaranteed equilibrium payoff is strictly bounded below his Stackelberg payoff. This follows from the same logic as in the baseline model: when the prior probability of commitment type is low enough and the short-run players believe that the strategic long-run player will play the commitment action with low probability, no signal realization can convince them to play their Stackelberg best reply. As a result, the signal has no impact on the 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 (which I call *binary action games*),

⁵When the short-run players have private information (such as which of the long-run player's actions she can observe), players cannot perfectly coordinate their actions based on the long-run player's action in the previous period. To overcome this challenge, I construct a belief-free equilibrium in the sense of Ely and Välimäki (2002) and Ely, Hörner and Olszewski (2005), while taken into account the constraints introduced by the short-run players' learning about the long-run player's type.

⁶In 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 completely uninformative, 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 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.

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. 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, similar to repeated games with private monitoring, the informativeness of a short-run player's action is different from her perspective and from her successors' perspective. For example, even if an action is very informative from the current period short-run player's perspective, it can be completely uninformative from some of her successors' perspectives. To overcome this difficulty caused by differences in beliefs, I use the observation that when the ex ante probability with which the short-run player plays the Stackelberg best reply is low, then the probability with which the long-run player plays the Stackelberg action in periods she observes cannot be too low. This is because otherwise, the short-run player will be convinced that the long-run player is committed. Given that the current period short-run player does not observe a rare event, the probability with which future short-run players believe that it occurs with very low probability is also bounded from below. This leads to a lower bound on the future short-run players' believed informativeness of the current period short-run player's action.

Given this connection between successful reputation formation and unboundedly informative signals in binary action games, one may wonder whether this relationship applies more generally. However, 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 exists 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 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 function of the short-run player (when she believes that this alternative action will be played), 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 to 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, then the environment is not resistant to learning against the commitment action when the long-run player's action choice is binary.

I show 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 for every payoff function of the long-run player, he can guarantee his commitment payoff in all equilibria when he is sufficiently patient.

Related Literature: My paper is related to several strands of literatures 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 the 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 social learning and addresses its welfare consequences.

The economic mechanism behind my reputation failure result is also different from the ones behind the herding results in Banerjee (1992) and Smith and Sørensen (2000). In those models, information aggregation fails because players' actions are not responsive to their private signals when the public signal becomes sufficiently precise. In my model, given that the long-run player plays the commitment action in every period, the short-run players *cannot* herd on an action that does not best reply to the commitment action. Therefore, the long-run player receiving low payoffs is not driven by short-run players joining a *bad herd*.

In fact, the public belief is never too precise in the adverse equilibrium of my model, and in the reputation building phase (i.e., the phase in which the long-run player's payoff is low), the short-run players' actions are responsive to their private observations. However, the responsiveness is low and decreases as the reputation-building player becomes more patient. This leads to low

returns from reputation building, reduces the long-run player's incentives to build his reputation, which makes the short-run players' beliefs self-fulfilling.

The reputation failure results in this paper contrast to the canonical reputation results of Fudenberg and Levine (1989,1992) and Gossner (2011). They show that the long-run player can secure his commitment payoff in all equilibria when the short-run players can observe a signal that statistically identifies his commitment action. In my model, short-run players learn primarily through their predecessors' actions, whose informativeness about the long-run player's actions are endogenous.⁷ When the long-run player becomes more patient, the informativeness of signal vanishes, which leads to low returns from reputation building.

In terms of studying the economic forces behind reputation failures, the mechanism suggested in this paper is different from the ones behind the *bad reputation* models of Ely and Välimäki (2003) and Ely, Fudenberg and Levine (2008). In their models, the short-run players can choose a *non-participating action* under which the public signal becomes completely uninformative about the long-run player's actions. In my model, the short-run players cannot unilaterally shut down learning. Instead, their successors will interpret their actions in an equilibrium context. In the reputation building phase of the low-payoff equilibrium, the short-run player's action in the current period can statistically identify the long-run player's action in the previous period. However, the speed of learning decreases as the long-run player becomes more patient, which leads to slow learning and low returns from reputation building.⁸

This paper is also related to reputation models with limited memories, such as the ones of Liu (2011) and Liu and Skrzypacz (2014). Different from my model, they study a model in which the short-run players cannot observe their predecessors' actions, for the purpose of shutting down channels for social learning. Their focus is on the long-run player's ability to unilaterally clean up histories and the Markov equilibria they characterize help to explain the reputation building-milking cycles. In contrast, each short-run player observes all the actions taken by her predecessors, so the long-run player cannot unilaterally clean up his past records. The focus of my model is instead on the effectiveness of reputation building through social learning.

⁷As in Gossner (2011), the sum of entropy between the equilibrium distribution of player 2's actions and the distribution over player 2's actions under the Stackelberg action is uniformly bounded from above. However, in periods in which player 2 does not have a strict incentive to play her best reply to the Stackelberg action, the minimal entropy between player 2's action distribution and the Stackelberg action vanishes as the long-run player becomes patient. Moreover, the speed with which it vanishes is proportional to $1 - \delta$.

⁸The idea that higher temptations to deviate in the stage-game implies higher speed of learning also appears in the dynamic sender-receiver game of Sobel (1985). In the equilibrium of his model, truth-telling leads to larger increase in the sender's reputation when the sender's gain from lying is larger.

Logina, Lukyanov and Shamruk (2019) characterize Markov equilibria in a product choice game in which each buyer can observe an informative signal about the seller's current period action, but cannot observe the seller's actions in the previous periods. 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 high or low. The logic is similar to the one in social learning models: when reputation is very high or very low, the buyer's prior belief in that period (before observing his private signal) is sufficiently precise such that their behavior is irresponsive to their private signals. This eliminates the opportunistic seller's incentive to exert high effort.

2 Baseline Model

I set up the baseline model, in which each uninformed player observes the entire history of her predecessors' actions, as well as the informed player's actions in the past K periods. Section 4 extends the baseline model by allowing the uninformed player to observe a noisy signal about the informed player's current period action, in addition to what she can observe in the baseline model.

Primitives: Time is discrete, indexed by $t = 0, 1, 2, \dots$. 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. Before they choose their actions in period t , player 1 and the player 2 (who arrives in period t) can observe the realization of a public randomization device $\xi_t \in [0, 1]$.

Players' stage-game payoffs are $u_1(a_t, b_t)$ and $u_2(a_t, b_t)$. Let $BR_1 : \Delta(B) \rightrightarrows 2^A \setminus \{\emptyset\}$ and $BR_2 : \Delta(A) \rightrightarrows 2^B \setminus \{\emptyset\}$ 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. $u_i(a, b) \neq u_i(a', b')$ for all $(a, b) \neq (a', b')$ and $i \in \{1, 2\}$.

Under Assumption 1, player 2 has a unique best reply to each of player 1's pure actions, and player 1 has a unique (pure) Stackelberg action. Let a^* be his Stackelberg action, which by definition, is the unique element of the set:

$$\arg \max_{a \in A} \left\{ \min_{b \in BR_2(a)} u_1(a, b) \right\}. \quad (2.1)$$

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 (denoted by ω^c), who mechanically plays a_1^* in every period; or he is a strategic type (denoted by ω^s), who can flexibly choose his actions in order to maximize his discounted average payoff, given by:

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

Player 2's prior belief attaches probability π_0 to the commitment type, with $\pi_0 \in (0, 1)$. 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.⁹ Examining this simplified monitoring structure is to highlight the economic forces at work. In section 3.2, I generalize my main result to cases in which each player 2 observes a bounded stochastic subset of player 1's past actions.¹⁰ Formally, let h^t be a typical history of the player 2 who arrives in period t , with

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

Let $\pi(h^t)$ be her posterior belief at history h^t . To simplify notation, I write π_t instead of $\pi(h^t)$.

Player 1's private history consists of his type, the entire sequence of his actions and player 2's actions in the past, as well as the current and past realizations of the public randomization devices. Let h_1^t be a typical private history in period t , with

$$h_1^t \equiv \{\omega, a_0, \dots, a_{t-1}, b_0, \dots, b_{t-1}, \xi_0, \dots, \xi_t\}.$$

⁹My 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.

¹⁰I adopt a belief-free equilibrium approach (Ely, Hörner and Olszewski 2005) in the proof of this result under stochastic sampling. Therefore, my result allows the long-run player to have arbitrary information about which subset of his past actions can each player 2 observes, which includes, but not limited to cases in which he has no information about what the short-run players can observe.

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 $\sigma_1 : \mathcal{H}_1 \rightarrow \Delta(A)$, with $\sigma_1 \in \Sigma_1$. Player 2's strategy is $\sigma_2 : \mathcal{H} \rightarrow \Delta(B)$, 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 2. Let $NE(\delta, \pi_0, K) \subset \Sigma_1 \times \Sigma_2$ 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 the 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:

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

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:

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

My result also applies under other discount factors adopted by a social planner, as long as it is no greater than the long-run player's discount factor δ .

3 Reputation Failure under Endogenous Signals

For $i \in \{1, 2\}$, let \underline{v}_i be player i 's lowest pure strategy Nash equilibrium payoff in the stage-game. According to Assumption 2, \underline{v}_1 is strictly lower than his Stackelberg payoff $u_1(a^*, b^*)$. I show the following reputation failure result, which unveils a new challenge to reputation building when information about the informed player's past actions is dispersed, and the uninformed players learn primarily through signals whose informativeness (about the informed player's past actions) is endogenously determined in equilibrium.

Theorem 1. *If the stage-game payoffs satisfy Assumptions 1 and 2, then there exists $\bar{\pi}_0 \in (0, 1)$,¹¹ such that for every $\pi_0 \in (0, \bar{\pi}_0)$ and δ large enough, there exists $(\sigma_1^\delta, \sigma_2^\delta) \in NE(\delta, \pi_0, K)$,*

¹¹The value of $\bar{\pi}_0$ depends on the buyers' memory length K and players' stage-game payoff functions.

such that:

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

According to Theorem 1, when information about player 1's past behavior is dispersed among his opponents and is aggregated via their actions, the long-run player's return from reputation building is low, no matter how patient he is. This contrasts to the conclusions in Fudenberg and Levine (1989, 1992) and Gossner (2011), which show that if player 2 can observe the entire history of player 1's actions, or noisy signals that can statistically identify player 1's actions, then a patient player 1 can guarantee his commitment payoff in all Nash 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 short-run players ignore their private signals and follow the public signal. In my reputation model, if player 1 deviates and plays a^* in every period, then the short-run players *cannot* herd on an action that is not b^* . To see why, suppose towards a contradiction that player 1 plays a^* in every period and player 2s herd on an action $b \in B$ that is not b^* . When herding occurs, the strategic type has no intertemporal incentives and plays his myopic best reply to b in equilibrium. Consider two cases. First, suppose $\text{BR}_1(b) = a^*$, then player 2 believes that a^* is played with probability 1 after observing a history that occurs with positive probability under the commitment type, in which case her best reply is b^* , which leads to a contradiction. Next, suppose $\text{BR}_1(b) \neq a^*$, then player 2 believes that player 1 is the commitment type after observing a history that occurs with positive probability under the commitment type, after which she has a strict incentive to play b^* . This contradicts the presumption that player 2s herd on an action that is not b^* .

In what follows, I provide a constructive proof of Theorem 1 which highlights the novel economic mechanism at work. In the class of equilibria I construct, the informativeness of the uninformed player's actions vanishes as the informed player becomes more patient. As a result, information about the long-run player's type is aggregated as $t \rightarrow \infty$, but the speed of learning vanishes as $\delta \rightarrow 1$, which eliminates all the returns from reputation building.

Proof of Theorem 1: Let $(a', b') \in A \times B$ be the worst pure strategy equilibrium for player 1 in the stage-game, which exists under Assumption 2.¹² If $b' = b^*$, then according to Assumption 1, $a' = a^*$ and \underline{v}_1 can be attained by playing (a^*, b^*) in every period.

¹²Let (a', b') be a generic pure strategy Nash equilibrium in the stage-game. Using this method, one can construct an equilibrium in which the long-run player's discounted average payoff equals $u_1(a', b')$ regardless of his discount factor. It generalizes to mixed strategy equilibria under certain conditions. See the end of section 3 for discussions.

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

$$u_1(a^*, b^*) > u_1(a', b') > u_1(a^*, b'). \quad (3.2)$$

Let $q^* \in (0, 1)$ be small enough such that b' is player 2's best reply against player 1's mixed action $q^* \circ a^* + (1 - q^*) \circ a'$. Let $\bar{\pi}_0 \equiv (q^*)^{K+1}$. When $\pi_0 < \bar{\pi}_0$, I construct the following two-phase equilibrium in which player 1's payoff is \underline{v}_1 regardless of δ . I start from describing players' strategies, and later verify the incentive constraints as well as their feasibility taken into account player 2s' posterior beliefs.

Play starts from a *reputation building phase*, in which player 2 plays b' , and the strategic player 1 mixes between a^* and a' such that according to player 2's belief, a^* is played with probability q^* . 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:

$$u_1(a', b') = (1 - \delta)u_1(a^*, b') + \delta \left\{ r u_1(a^*, b^*) + (1 - r)u_1(a', b') \right\}. \quad (3.3)$$

This equation admits a solution r that 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 has ever played actions other than a^* after reaching the reputation maintenance phase, player 2 plays b' at all subsequent histories and player 1 plays a' at all subsequent histories.

I verify players' incentives and the feasibility of player 1's mixed 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 posterior belief

attaches probability less than q^* to the commitment type at every history of the reputation building phase. When she has observed an action other than a^* , her posterior belief is 0. When the long-run player has played a^* in the past K periods, her posterior belief at h^t , denoted by $\pi(h^t)$ satisfies:

$$\frac{\pi_t(h^t)}{\pi_0} = \frac{\Pr^{(\sigma_1^\delta, \sigma_2^\delta)}(a^*, \dots, a^* | \omega^c)}{\Pr^{(\sigma_1^\delta, \sigma_2^\delta)}(a^*, \dots, a^*)} \cdot \frac{\Pr^{(\sigma_1^\delta, \sigma_2^\delta)}(b', \dots, b', \xi_0, \dots, \xi_t | a^*, \dots, a^*, \omega^c)}{\Pr^{(\sigma_1^\delta, \sigma_2^\delta)}(b', \dots, b', \xi_0, \dots, \xi_t | a^*, \dots, a^*)}. \quad (3.4)$$

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,

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

and

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

Since $\pi_0 \leq \bar{\pi}_0 = (q^*)^{K+1}$, we know that $\pi_t(h^t) \leq q^*$ for every $h^t \in \mathcal{H}$ of the reputation building phase. This verifies that player 1's mixed strategy in the reputation building phase is feasible. \square

Short-run Players' Welfare: One can use the same idea to show the following proposition:

Proposition 1. *Under Assumptions 1 and 2, for every $\varepsilon > 0$, there exists $\bar{\pi}_0 \in (0, 1)$ such that for every $\pi_0 \in (0, \bar{\pi}_0)$ and δ large enough, there exists $(\sigma_1^\delta, \sigma_2^\delta) \in NE(\delta, \pi_0, K)$, such that:*

$$\mathbb{E}^{(\sigma_1^\delta, \sigma_2^\delta, \pi_0)} \left[\sum_{t=0}^{\infty} (1 - \delta) \delta^t u_2(a_t, b_t) \right] \leq v_2 + \varepsilon. \quad (3.5)$$

Proof of Proposition 1: Consider the class of equilibria constructed in the proof of Theorem 1. Let (a', b') be player 2's worst stage-game Nash equilibrium. Let V_2 be the short-run players' welfare in this equilibrium, we have:

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

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 $\bar{\pi}_0 \equiv (q^*)^{K+1}$, the resulting strategy profile is an equilibrium in which the short-run players' welfare is no more than $u_2(a', b') + \varepsilon$. \square

Implications on Product Choice Game: I discuss the implications of Theorem 1 and Proposition 1 in the product choice game of Mailath and Samuelson (2001), which is a classic model that captures the lack-of-commitment problem in business transactions. 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:¹³

–	B	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 $\bar{\pi}_0 > 0$, such that when π_0 is below $\bar{\pi}_0$, 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 equilibrium constructed in the proof sheds 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). 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 for choosing H , they tend to attribute more to the seller's strategic concerns instead of to his intrinsic preference after observing the seller supplying high quality in the previous periods. In equilibrium, the buyers' actions are less responsive to the seller's previous actions when the seller is more patient or when his temptation to deviate is low. 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 the surplus from the long-term relationship.

¹³I interpret N as a non-trusting action, instead of “not buying”. Importantly, the seller's action in period t is observable to some future buyers despite the buyer in period t has chosen N .

3.1 Guaranteed Equilibrium Payoff & 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, I show that 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:¹⁴

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 $\underline{b} \equiv \text{BR}_2(\underline{a})$. 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.

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 $\bar{\pi}_0 \in (0, 1)$ such that for every $\pi_0 \in (0, \bar{\pi}_0)$ and δ large enough, there exists an equilibrium in which player 1's payoff equals his minmax, and player 2's payoff is ε -close to her minmax payoff.*

¹⁴This 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.

When stage-game payoffs are not monotone-supermodular, player 2's action that minmaxes player 1's action is nontrivially mixed. 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^* . The same construction in the proof of Theorem 1 can be used to show that the patient long-run player'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 (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), but the seller does not know players' social network.

Motivated by these applications, I generalize the insights of Theorem 1 in environments with stochastic sampling. For every $t \geq 1$, let

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

be the distribution over the t th short-run player's neighborhood. For every $s \neq t$, I assume that \mathcal{N}_s and \mathcal{N}_t are independent random variables.¹⁵ The short-run player's history in period t is:

$$h^t \equiv \left\{ b_0, b_1, \dots, b_{t-1}, \left(a_s \right)_{s \in \mathcal{N}_t}, \xi_0, \dots, \xi_t \right\}, \quad (3.8)$$

and the long-run player cannot directly observe the current and past realizations \mathcal{N}_t . I make the following assumption on the network structure $\{\mathcal{N}_t\}_{t=1}^\infty$:

¹⁵This rules out networks with correlated neighborhoods, such as the case studied by Lobel and Sadler (2015).

Assumption 4. *There exists $K \geq 1$ and $\gamma > 0$ such that for every $t \geq 1$, $\Pr(|\mathcal{N}_t| \leq K) = 1$ and $\Pr(t - 1 \in \mathcal{N}_t) \geq \gamma$.¹⁶*

The first part of this assumption implies that each short-run player only samples a bounded subset of his predecessors' experiences. The second part says that the probability with which each player observing the long-run player's action in the previous period is uniformly 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. I show the following result in the product choice game, which applies more generally to monotone-supermodular game 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 $\bar{\pi}_0 \in (0, 1)$, such that for every $\pi_0 \in (0, \bar{\pi}_0)$ and δ large enough, there exists $(\sigma_1^\delta, \sigma_2^\delta) \in NE(\delta, \pi_0, K)$, such that:*

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

Different from the baseline model, players cannot perfectly coordinate rewards and punishments based on player 1's previous period action, since it is no longer common knowledge between the players. To overcome this challenge, I construct a belief-free equilibrium while taking into account the constraints imposed by the short-run players' posterior beliefs.

3.3 Comparison to Existing Reputation Models

I compare Theorem 1 and Proposition 2 to the canonical reputation results in Fudenberg and Levine (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 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.

¹⁶My result also applies once we replace the second part of Assumption 3 with:

$$\exists k \in \mathbb{N} \text{ and } \gamma > 0 \text{ such that } \Pr(\{t - k, \dots, t - 1\} \cap \mathcal{N}_t \neq \{\emptyset\}) \geq \gamma \text{ for every } t \in \mathbb{N},$$

which means that the probability with which a short-run player observes the long-run player's recent actions is uniformly bounded from below.

The first obstacle does not hinder player 1 from receiving high payoffs: if player 2 expects a^* to be played with high enough probability in the current period, then although learning is slow, player 2 will have 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 case in which $K = 1$, I argue that in the worst equilibrium for player 1: (1) b_{t+1} is informative about a_t , but (2) its informativeness vanishes as $\delta \rightarrow 1$.

To start with, consider candidate equilibria in which b_{t+1} is completely uninformative about a_t . According to Assumption 2, player 1 has a strict incentive to deviate to low effort. As a result, after observing high effort 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 b^* in period $t + 1$. If this cycle persists, then a patient player 1's average payoff across the two periods is approximately $\frac{1}{2}(v_1 + u_1(a^*, b^*))$ by playing a^* in every period.¹⁷

Next, 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 patient players place more weight on continuation payoffs relative to current period payoffs. In particular, the required level of informativeness vanishes to 0 as $\delta \rightarrow 1$.

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 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

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

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

¹⁷For any $K \in \mathbb{N}$, if b_{t+1} is uninformative about a_t , then player 1 can guarantee an average payoff close to $\frac{K}{K+1}u_1(a^*, b^*) + \frac{1}{1+K}v_1$ 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.

to the case with exogenous signals in which the number of such periods is uniformly bounded.

4 Informative Signal about Current Period Action

In this section, I expand the baseline model by allowing each uninformed player to 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 a *sequential-move* stage-game in which 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 $f(s|a) > 0$ if and only if $a = a^*$.
2. Otherwise, \mathbf{f} is *boundedly informative* about $a^* \in A$.

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

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

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

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

be a patient long-run player's *guaranteed equilibrium payoff*, namely, his lowest equilibrium payoff when he is sufficiently patient.

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 the signal. More generally, if \mathbf{f} is boundedly

informative about a^* , then the patient long-run player'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 $\bar{\pi}_0 \in (0, 1)$ such that for every $\pi_0 \in (0, \bar{\pi}_0)$ and δ large enough, there exists $(\sigma_1^\delta, \sigma_2^\delta) \in NE(\delta, \pi_0, K, \mathbf{f})$, such that:*

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

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

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

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

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

and the probability with which he plays a is $1 - \hat{q}$. Let $\bar{\pi}_0 = \hat{q}^K$, 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. \square

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. 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 A. 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 (1) uninformed players have limited memories about the informed player's actions, and (2) 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 result in Smith and Sørensen (2000), that players choose the correct action in the long run *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 successfully overcome the aforementioned challenge and secure himself his optimal commitment 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 low.

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 from the perspective of player 2 in period $s (> t)$, it is very uninformative.

I use the following argument to bound the payoff consequences of such differences in beliefs. If player 2 in period t (1) observes that a^* has been played in the past K periods, and (2) believes

(before observing s_t) that $b_t = b^*$ with probability at most $1 - \epsilon$, then the probability with which $(a_{t-K}, \dots, a_{t-1}) = (a^*, \dots, a^*)$ under the equilibrium strategy profile must be bounded from below. This is because otherwise, player 2 in period t will believe 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 $(a_{t-K}, \dots, a_{t-1}) = (a^*, \dots, a^*)$ 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 to an event that occurs 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 challenges. 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

The set of signal realizations is $S \equiv \{s^*, s^{**}, s'\}$. 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, f is unboundedly informative about a^* .

Consider the following strategy profile: strategic player 1 mixes between a^{**} and a' with equal probability. Player 2 plays b^* if $s_t \in \{s^*, s^{**}\}$ and plays b' if $s_t = s'$. This strategy profile is an equilibrium when $\pi_0 < 10^{-K-1}$. 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 in which player 1 has three actions and player 2 has two actions.

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

The set of signal realizations is $S \equiv \{s^*, s', s''\}$. 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 $f(s''|a'') = 1$. One can check that player 1's Stackelberg action is a^* , the game satisfies Assumptions 1 and 2, and moreover, there exists an unboundedly informative signal that perfectly identifies a^* .

Consider the following strategy profile: strategic player 1 mixes between the three of his actions with probabilities such that according to player 2's belief, the probability distribution of player 1's action is $0.5 \circ a^* + 0.25 \circ a' + 0.25 \circ a''$. 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. 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 propensity to play b' against different actions of player 1's. In particular, player 2's propensity to play b' is much stronger under a'' than 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 about a^* , and has an incentive to play b' following a signal realization that leads to a high posterior about a^* . This situation is implicitly ruled out when $|A| = 2$ since there is only one *bad action*, but could occur under some stage-game payoff and some signal structure 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. For some extra notation, for every $\alpha \in \Delta(A)$, matrix \mathbf{f} that

summarizes the signal distribution, and $\beta : S \rightarrow \Delta(B)$, 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 distributed according to \mathbf{f} , and (3) player 2 behaves according to β after observing the realization of 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 $a^* \in \text{supp}(\alpha)$, and $\beta : S \rightarrow \Delta(B)$ which is a best reply against α under u_2 , such that:

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

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

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

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, action a^* induces player 2 to play her complete information best reply with probability 1.

Applying Definition 2 to some of our 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 $\bar{\pi}_0 > 0$ 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 $\pi_0 < \bar{\pi}_0$ and δ large enough, there exists $(\sigma_1^\delta, \sigma_2^\delta) \in NE(\delta, \pi_0, K, \mathbf{f})$ such that:

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

The proof of Theorem 3 is in Appendix B. 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 *resistant to learning* and *not resistant to learning*. I start from introducing a regularity condition on u_2 that captures the heterogeneity 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 resistant 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 $\bar{\pi}_0 \in (0, 1)$, such that for every $\pi_0 < \bar{\pi}_0$ and δ large enough, there exists $(\sigma_1^\delta, \sigma_2^\delta) \in NE(\delta, \pi_0, K)$, such that:*

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

Proofs of Lemma 4.1 and Corollary 3: For every $a^* \in A$ and admissible u_2 , let b^* be the unique element in $BR_2(a^*)$. 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. $BR_2(\alpha) = \{b^*, b'\}$.

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

$$u_2(a', b') - u_2(a', b^*) < u_2(a'', b') - u_2(a'', b^*), \quad (4.5)$$

and $u_2(a'', b') - u_2(a'', b^*) > 0$.¹⁸ 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 .

¹⁸This 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.

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

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

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. \square

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 \mathbf{f} that is unbounded informative about his highest action and possesses the standard *monotone likelihood ratio property* (or MLRP for short).

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

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

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

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

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

Proof of Lemma 4.2: Since \mathbf{f} is unboundedly informative about \bar{a} , there exists $s^* \in S$ such that $f(s^*|a) > 0$ if and only if $a = \bar{a}$. Since \bar{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 $b^* \equiv \text{BR}_2(\bar{a})$ 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. \square

A 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

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

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

$$\frac{1 - g(h^t, \omega^c)}{1 - g(h^t)}. \quad (\text{A.1})$$

Let $A \equiv \{a^*, a'\}$ and $S \equiv \{s^*, s_1, s_2, \dots, s_m\}$. 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 .¹⁹ I suppress the dependence on h^t in order to simplify notation. Since $\{b^*\} = \text{BR}_2(a^*)$ 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 $1 - r \geq C(1 - g)$.²⁰

According to the first implication, it is without loss of generality to label the signal realizations such that $p(s_1) \geq p(s_2) \geq \dots \geq p(s_m)$, and moreover, there exists $k \in \{1, 2, \dots, m\}$ such that player 2 plays b^* for sure after observing s_1, \dots, s_{k-1} , and does not play b^* otherwise.²¹ Therefore,

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

¹⁹Notice that r, τ, p depend on player 1's action choice at h^t , which is endogenously determined in equilibrium.

²⁰This 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.

²¹Ignoring 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 fact that $p(s_1) \geq p(s_2) \geq \dots \geq p(s_m)$, we know that:

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

As a result,

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

and

$$\sum_{i=k}^m \tau(s_i) (1 - p(s_i)) \geq \frac{1 - r}{1 - r f(s^*|a^*)} (1 - g). \quad (\text{A.4})$$

Therefore,

$$\frac{1 - g(\omega^c)}{1 - g} \leq \frac{1 - f(s^*|a^*)}{1 - r f(s^*|a^*)}, \quad (\text{A.5})$$

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

$$\frac{1 - g(\omega^c)}{1 - g} \leq \frac{1 - f(s^*|a^*)}{1 - f(s^*|a^*) + C f(s^*|a^*) (1 - g)}. \quad (\text{A.6})$$

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

$$\frac{g(\omega^c)}{g} \geq 1 + \frac{f(s^*|a^*) (1 - g(h^t))}{g - r f(s^*|a^*)} \geq 1 + \frac{f(s^*|a^*) (1 - g)}{g - f(s^*|a^*) (1 - C(1 - g))} \quad (\text{A.7})$$

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 (A.6) and (A.7) imply the following lower bound on the KL divergence between $\beta(h^t)$ and $\beta(h^t, \omega^c)$:

$$d(\beta(h^t) \parallel \beta(h^t, \omega^c)) \leq \mathcal{L}(1 - g(h^t)), \quad (\text{A.8})$$

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

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 $\{a_{t-K}, \dots, a_{t-1}\}$, 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 (A.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}^σ 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 write the posterior likelihood ratio as the product of the likelihood ratio of the signals observed in each period:

$$\begin{aligned} & \frac{\pi(h^t)}{1 - \pi(h^t)} \bigg/ \frac{\pi_0}{1 - \pi_0} \\ = & \frac{\mathcal{P}^{\sigma, \omega^c}(b_0)}{\mathcal{P}^{\sigma, \omega^s}(b_0)} \cdot \frac{\mathcal{P}^{\sigma, \omega^c}(b_1|b_0)}{\mathcal{P}^{\sigma, \omega^s}(b_1|b_0)} \cdot \dots \cdot \frac{\mathcal{P}^{\sigma, \omega^c}(b_{t-1}|b_{t-2}, \dots, b_0)}{\mathcal{P}^{\sigma, \omega^s}(b_{t-1}|b_{t-2}, \dots, b_0)} \cdot \frac{\mathcal{P}^{\sigma, \omega^c}(a_{t-K}, \dots, a_{t-1}|b_t, b_{t-1}, \dots, b_0)}{\mathcal{P}^{\sigma, \omega^s}(a_{t-K}, \dots, a_{t-1}|b_t, b_{t-1}, \dots, b_0)} \end{aligned} \quad (\text{A.9})$$

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

$$\mathcal{P}^{\sigma, \omega^c} \left(\pi^\sigma(b_0, b_1, \dots, b_{t-1}) < \epsilon \pi_0 \right) \leq \epsilon \frac{1 - \pi_0}{1 - \pi_0 \epsilon}, \quad (\text{A.10})$$

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

$$\rho^*(\epsilon) \equiv \frac{\epsilon \pi_0}{1 - \epsilon \pi_0}. \quad (\text{A.11})$$

Next, I show that if:

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

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

To see this, suppose towards a contradiction that the probability with which $(a_{t-K}, \dots, a_{t-1}) = (a^*, \dots, a^*)$ is strictly less than $\rho^*(\epsilon)$ conditional on (b_0, \dots, b_{t-1}) . According to (A.11), after observing

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

$$\frac{\eta\rho^*(\epsilon)(1 - \rho^*(\epsilon))}{(1 - \eta\rho^*(\epsilon))\rho^*(\epsilon)} = \eta \frac{1 - \rho^*(\epsilon)}{1 - \rho^*(\epsilon)\eta}. \quad (\text{A.12})$$

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 (A.6) and (A.7) also apply in this setting, namely,

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

and

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

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

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

Applying (A.15) to (A.13) and (A.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:

$$\left(1 - \epsilon - \epsilon \frac{1 - \pi_0}{1 - \pi_0\epsilon}\right) u_1(a^*, b^*) + \left(\epsilon + \epsilon \frac{1 - \pi_0}{1 - \pi_0\epsilon}\right) \min_{a,b} u_1(a, b) - \epsilon. \quad (\text{A.16})$$

Taking $\epsilon \rightarrow 0$ and $\delta \rightarrow 1$, (A.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, Harvard Kennedy School.
- [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] Ely, Jeffrey, Johannes Hörner and Wojciech Olszewski (2005) “Belief-Free Equilibria in Repeated Games,” *Econometrica*, 73(2), 377-415.
- [7] Ely, Jeffrey and Juuso Välimäki (2002) “A Robust Folk Theorem for the Prisoner’s Dilemma,” *Journal of Economic Theory*, 102(1), 84-105.
- [8] Ely, Jeffrey and Juuso Välimäki (2003) “Bad Reputation,” *Quarterly Journal of Economics*, 118(3), 785-814.
- [9] Ely, Jeffrey, Drew Fudenberg and David Levine (2008) “When is Reputation Bad?” *Games and Economic Behavior*, 63(2), 498-526.
- [10] 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.
- [11] Fudenberg, Drew and David Levine (1989) “Reputation and Equilibrium Selection in Games with a Patient Player,” *Econometrica*, 57(4), 759-778.
- [12] Fudenberg, Drew and David Levine (1992) “Maintaining a Reputation when Strategies are Imperfectly Observed,” *Review of Economic Studies*, 59(3), 561-579.
- [13] Gossner, Olivier (2011) “Simple Bounds on the Value of a Reputation,” *Econometrica*, 79(5), 1627-1641.
- [14] Liu, Qingmin (2011) “Information Acquisition and Reputation Dynamics,” *Review of Economic Studies*, 78(4), 1400-1425.
- [15] Liu, Qingmin and Andrzej Skrzypacz (2014) “Limited Records and Reputation Bubbles,” *Journal of Economic Theory* 151, 2-29.
- [16] Lobel, Ilan and Evan Sadler (2015) “Information Diffusion in Social Networks,” *Theoretical Economics*, 10(3), 807-851.

- [17] Logina, Ekaterina, Georgy Lukyanov and Konstantin Shamruk (2019) “Reputation and Social Learning,” Working Paper.
- [18] Mailath, George and Larry Samuelson (2001) “Who Wants a Good Reputation?” *Review of Economic Studies*, 68(2), 415-441.
- [19] 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.
- [20] Pei, Harry (2018) “Reputation Effects under Interdependent Values,” Working Paper, Northwestern University.
- [21] Pei, Harry (2019) “Trust and Betrayals: Reputational Payoffs and Behaviors without Commitment,” Working Paper, Northwestern University.
- [22] Schmidt, Klaus (1993) “Reputation and Equilibrium Characterization in Repeated Games with Conflicting Interests,” *Econometrica*, 61(2), 325-351.
- [23] Smith, Lones and Peter Norman Sørensen (2000) “Pathological Outcomes of Observational Learning,” *Econometrica*, 68(2), 371-398.
- [24] Sobel, Joel (1985) “A Theory of Credibility,” *Review of Economic Studies*, 52(4), 557-573.