Trust and Betrayals
Reputation Building and Milking without Commitment

Harry Di Pei*

July 24, 2018

Abstract: I introduce a reputation model where all types of the reputation building agent are rational and are facing lack-of-commitment problems. I study a repeated *trust game* in which a patient player (e.g. seller) wishes to win the trust of some myopic opponents (e.g. buyers) but can strictly benefit from betraying them. Her benefit from betrayal is her persistent private information. I provide a tractable formula for the patient player's highest equilibrium payoff, which converges to her mixed Stackelberg payoff when the lowest benefit in the support of the prior belief vanishes. In equilibria that attain this highest payoff, reputations are built and milked gradually and the patient player's behavior must be non-stationary. This enables her to extract information rent in unbounded number of periods while minimizing her long-term reputation loss. Moreover, her reputation in equilibrium can be computed by counting the number of times she has betrayed as well as been trustworthy in the past. This captures some realistic features of online rating systems.

Keywords: lack-of-commitment problems, reputation, trust

JEL Codes: C73, D82, D83

1 Introduction

Trust is essential in many economic activities, yet it is also susceptible to opportunism and exploitation. To fix ideas, imagine a politician running for president pledging for massive tax cuts. Once elected, he might be tempted to breach his promise due to the growth in mandatory spending and rising budget deficits. Anticipating such possibilities of future betrayal, should the electorate vote for this candidate in the first place?[1] Alternatively, firms try to convince consumers about their high quality standards, but after receiving the upfront payments, they are tempted to undercut quality, especially on aspects that are hard to verify. Similar plights occur when incumbents deter entrants, central banks fight hyperinflation and entrepreneurs seek funding for their projects.

*Email: harrydp@northwestern.edu. Department of Economics, Northwestern University. I am indebted to Drew Fudenberg, Juuso Toikka and Daron Acemoglu for guidance and support. I thank Ricardo Alonso, Mehmet Ekmekci, Bob Gibbons, Debraj Ray, Jean Tirole, Chara Tzanetaki, Alex Wolitzky, Muhamet Yildiz and my seminar audience in Warwick for helpful comments. Errors are mine.

[1] A classic example is ex-president George H.W. Bush's 1988 acceptance speech at the New Orleans convention "*Read my lips, no new taxes.*" But after becoming president, he agreed to increase several existing taxes in order to reach a compromise with the Democrat-controlled Congress. Breaching this promise has hurt Bush politically during his 1992 campaign, as both Pat Buchanan and Bill Clinton cited his quotation and questioned his trustworthiness.

The common theme in this class of applications is a *lack-of-commitment problem* faced by the firms, politicians, central banks and entrepreneurs. As a response, these agents build reputations for being trustworthy, from which they can derive benefits in the future. In practice, a key challenge to reputation building is that all agents are facing temptations to renege, including those '*role-models*' that others wish to imitate. As a result, the heterogeneity across agents is more about how much temptation they are facing, rather than whether they are facing temptations or not.[2] This contrasts to the classic theories in Sobel (1985), Fudenberg and Levine (1989,1992), Benabou and Laroque (1992), etc. where some types of the agent are committed to playing pre-specified strategies, and others can establish reputations by imitating those *commitment types*.

This paper introduces a reputation model that incorporates these realistic concerns. To highlight the lack-of-commitment problems in the applications, I study the following *trust game* that is played repeatedly over the infinite time horizon between a patient long-run player (e.g. seller) and a sequence of myopic short-run players (e.g. buyers).[3] In every period, the long-run player wishes to win her opponent's trust by promising high effort, but has a strict incentive to renege and exert low effort once trust is granted. Her cost of high effort is her persistent private information, which I call her *type*. Every short-run player perfectly observes all the actions taken in the past and is willing to trust the long-run player if he expects effort to be high with probability above some cutoff.

I show that despite all types of the long-run player are tempted to renege, she can still overcome her lack-of-commitment problem and attain her commitment payoff from playing mixed actions. This includes her (mixed) *Stackelberg payoff* when the lowest cost in the support of the prior belief vanishes. The absence of commitment types also leads to interesting implications on the long-run player's behavior. In equilibria that attain her highest payoff, her strategy must be non-stationary and reputations are built and milked gradually. Moreover, her equilibrium reputation follows an intuitive rule-of-thumb and can be computed by counting the number of times she has exerted high and low effort in the past. This captures some realistic features of online rating systems such as eLance, Uber, Yelp, etc. (Dellarocas 2006, Dai et al.2018) in which a seller's score only depends on the number of times she has received each rating, instead of other more complicated metrics.[4]

My analysis starts with the complete information benchmark. If the long-run player's cost of effort is common knowledge, then according to Fudenberg, Kreps and Maskin (1990), her highest payoff in the *repeated complete information game* cannot exceed her payoff from trust and high effort, which is strictly below her Stackelberg payoff. Intuitively, this is because the long-run player needs to exert high effort with positive probability every

[2]For example, all firms can save cost by undercutting quality, but their costs can be different due to different production technologies.

[3]The assumption on myopia is motivated by the applications, such as in durable good markets where each buyer has unit demand, online platforms (Airbnb, Uber, Lyft) where buyers are unlikely to meet with the same seller twice. Relaxing this assumption will not affect the result on the attainability of Stackelberg payoff or the insight that the patient long-run player can overcome her lack-of-commitment problem. Nevertheless, it will affect the patient long-run player's equilibrium payoff set.

[4]Websites such as eBay and eLance only consider ratings obtained in the past six months when they compute sellers' scores. This is motivated by concerns such as the seller's type is changing over time, which is beyond the scope of this paper.

time she receives her opponent's trust. As a result, exerting high effort whenever she is trusted is her best reply, from which her stage-game payoff cannot exceed the above upper bound, so is her discounted average payoff.

My first result (Theorem 1) characterizes the set of equilibrium payoffs a patient long-run player can attain in the *repeated incomplete information game*. At the heart of this characterization is a simple formula for every type's highest equilibrium payoff, which equals to the product of her Stackelberg payoff and an *incomplete information multiplier*. The latter summarizes the effect of incomplete information, which is common for all types, strictly below one and only depends on the *lowest cost* in the support of the prior belief. My formula implies that every type except for the lowest cost one can strictly benefit from incomplete information, i.e. her highest equilibrium payoff increases compared to the complete information benchmark. Furthermore, when this lowest cost vanishes to zero, the multiplier converges to one. That is to say, every type of the long-run player can overcome her lack-of-commitment problem and attain her Stackelberg payoff.

This result is reminiscent of Fudenberg and Levine (1992), who show that if with positive probability, the long-run player is a *commitment type* who is mechanically playing a mixed strategy, then she can obtain her commitment payoff from that mixed strategy. Nevertheless, it remains unclear whether there are good ways to rationalize those mixed strategy commitment types. Theorem 1 provides a partial strategic foundation via rational types that have regular ordinal preferences over stage-game outcomes.[5] In particular, the mixed Stackelberg commitment type is rationalized by strategic types that have very low albeit positive cost to exert high effort. My approach maintains the sensible assumptions on the game's payoff structure, such as providing high quality is costly for a firm but it can benefit from consumers' purchases, which makes the conclusion more convincing.

Next, I explore how the absence of commitment types affects the long-run player's behavior. As a first step, I show in Theorem 2 that in every equilibrium that (approximately) attains the long-run player's highest payoff, including those Stackelberg equilibria, no type of the patient long-run player will play stationary strategies or completely mixed strategies. More interestingly, this conclusion also applies to types whose cost of exerting high effort is arbitrarily low or even zero. This stands in sharp contrast to the classic examples of stationary commitment types who are mechanically playing the same mixed action in every period.

To understand why, suppose towards a contradiction that one of the types is playing a completely mixed strategy, then both exerting high effort at every history and exerting low effort at every history are her best replies. According to a result on one-shot signalling games developed by Liu and Pei (2017),[6] every type with strictly higher cost will exert low effort at every on-path history and every type with strictly lower cost will exert

[5] By *partial*, I mean that the long-run player can attain her mixed commitment payoff in the repeated game without commitment. Nevertheless, her equilibrium payoff set can be different once comparing the game with and without commitment types.

[6] Liu and Pei (2017) show by counterexample that supermodularity of players' payoffs are not sufficient to guarantee the monotonicity of the sender's action with respect to her type in one-shot signalling games. They also show that every Nash equilibrium is monotone when players' stage game payoffs are monotone-supermodular, which is satisfied in the stage game studied in this paper.

high effort at every on-path history. In what follows, I will argue that none of these pure stationary strategies can arise when the long-run player approximately attains her highest equilibrium payoff. To be more precise, they are inconsistent with all types except one benefiting from incomplete information.

1. Suppose a type always exerts low effort, then according to the learning argument in Fudenberg and Levine (1989,1992), the short-run players will eventually believe that low effort will occur with sufficiently high probability in every future period, after which they will stop trusting the long-run player.

2. Suppose a type always exerts high effort, then she cannot extract information rent. Moreover, the type with cost immediately above her cannot extract information rent either. This is because after shirking for one period, she will be separated from all types with strictly lower cost and therefore, she will be the lowest cost type according to the short-run players' posterior belief. Therefore, her continuation value after the first time she exerts low effort cannot exceed her highest payoff in the repeated complete information game.

Theorem 2 implies that in every equilibrium that attains the long-run player's highest payoff, the discounted average frequency of low effort along every action path of the lowest cost type cannot exceed a certain cutoff. This cutoff converges to the probability of low effort in the mixed Stackelberg action when the lowest cost vanishes. This requires the long-run player to *cherry-pick* her actions based on the game's history.

To gain a better understanding of behavior, my proof of Theorem 1 constructs equilibria that attain the long-run player's highest equilibrium payoff. The key challenge arises from the observation that extracting information rent (i.e. shirk while winning her opponent's trust) inevitably reveals information about her type, which undermines her informational advantage as well as her ability to extract information rent in the future. This tension grows when the long-run player is patient, as she needs to extract information rent in unbounded number of periods to obtain a discounted average payoff significantly above her complete information payoff.

I overcome this challenge by constructing equilibria that exhibit *reputation building-milking cycles* and *slow learning*. In periods where active learning takes place, the short-run players play trust and the long-run player's reputation (i.e. probability that she is the lowest-cost type) improves after high effort and deteriorates after low effort.[7] Every high-cost type will play a non-trivially mixed action unless her reputation is sufficiently close to one, at which point she will shirk for one period and extract information rent. Nevertheless, she can always *rebuild her reputation* after milking it, which allows such cycles to persist in the long-run and therefore, learning and rent extraction can occur in unbounded number of periods.

Next, I explain why slow learning can increase the patient long-run player's payoff. First, the short-run players' incentives to trust imply that high effort needs to occur with probability above some cutoff in every

[7]To ensure different types of long-run player's incentives to mix, learning will stop and play will transit to one of the two absorbing phases either when the long-run player has built a perfect reputation, or after she has shirked for sufficiently many periods.

period. This leads to an upper bound on the ratio between the magnitude of reputation improvement after high effort and that of reputation deterioration after low effort (or the *relative rate* of learning). Second, fixing the long-run frequencies of high and low effort, the amount of reputation loss per period increases with the *absolute rate* of learning. Therefore, keeping the relative rate fixed while decreasing the absolute rate allows the long-run player to increase her long-run frequency of low effort without compromising on her long-term reputation as well as the short-run players' willingness to trust.[8] Such adjustment increases her equilibrium payoff.

Related Literature: This paper contributes to the literatures on credibility, reputations and repeated incomplete games from several different angles. From a modeling perspective, I introduce a new framework with realistic informational assumptions to study trust building and reputations. Compared to Sobel (1985), Fudenberg and Levine (1989,1992), Benabou and Laroque (1992), etc. all types of the reputation builder are rational and share the same ordinal preferences over stage-game outcomes. Aside from being motivated by various trust-building problems in reality, my approach also addresses the concerns raised by Weinstein and Yildiz (2007) that when all forms of incomplete information are allowed, one can rationalize almost every outcome by introducing types that have qualitatively different payoff functions and beliefs. Their finding calls for a more careful selection of the types included in incomplete information game models. In particular, every type should have reasonable incentives including those that occur with very low probability, which is in the spirit of my model.

My paper also contributes to the literature that studies agents' behaviors when facing reputation concerns, such as Benabou and Laroque (1992), Tirole (1996), Phelan (2006), Ekmekci (2011), Liu (2011), Jehiel and Samuelson (2012), Liu and Skrzypacz (2014), etc. Compared to those papers, I characterize players' behaviors in environments *with multiple strategic types* and *without commitment types*. This expands the theory's applicability and moreover, leads to interesting implications on the reputation dynamics. I will elaborate more on the details in subsection 4.3. Sobel (1985), Schmidt (1993), Ghosh and Ray (1996) also study models without commitment types. However, in those models, one of the strategic type's behavior is either trivial or is exogenously assumed, so therefore, can be treated as a commitment type in the analysis. In contrast, no strategic type's behavior is trivial in my model as all of them are patient and have strict preferences over stage-game outcomes.

My characterization of the patient long-run player's equilibrium payoff set is related to the study of repeated incomplete information games, such as Hart (1985), Aumann and Maschler (1995), Cripps and Thomas (2003), Hörner and Lovo (2009), Pęski (2014), etc. Instead of studying games where all players are patient, I focus on games where one player is patient but her opponents are myopic. This asymmetry in discount factors introduces

[8] Decreasing the absolute rate of learning can hurt the long-run player when her discount factor is low, as it increases the number of periods required to build a reputation (i.e. reaching a point at which she can shirk with probability 1 for one period). Nevertheless, the payoff consequences for this bounded number of periods becomes negligible as the long-run player becomes patient.

novel constraints on the set of equilibrium payoffs. Moreover, my characterization can be viewed as a benchmark to study private value reputation models with commitment types. The techniques I developed to construct equilibria have analogs in interdependent value environments, which is addressed in Pei (2017).

In terms of rationalizing commitment types in the canonical reputation models, Weinstein and Yildiz (2016) study a complementary problem that rationalizes *non-stationary pure commitment types* in finitely repeated games (such as tit-for-tat) using strategic types that have repeated game payoffs. In contrast, my paper rationalizes *mixed strategy commitment types* using strategic types that not only have repeated game payoffs, but also share the same ordinal preferences over stage-game outcomes as the normal type.

2 The Baseline Model

In this section, I introduce a repeated trust game that captures the lack-of-commitment problem in many socioeconomic interactions. Different from the canonical reputation models with commitment types, all types of the reputation building player are rational and moreover, have qualitatively similar payoff functions. This is motivated by the concern that in reality, none of the agents are immune to reneging temptations, including those reputational types that others wish to imitate. Despite my model is framed in context of business transactions, the underlying economic mechanism applies to alternative settings such as public policies, entry deterrence, etc.

2.1 The Stage Game

Consider the following game between a seller (player 1, she) and a buyer (player 2, he). The buyer moves first, deciding whether to purchase a product from the seller (i.e. *trusting the seller*, taking action T) or not (i.e. *not trusting the seller*, taking action N). If he takes action N, then both players' stage game payoffs are 0. If he takes action T, then the seller chooses between high effort (action H) and low effort (action L). If the seller chooses L, then her stage game payoff is normalized to 1 and the buyer's payoff is $-c$. If the seller chooses H, then her stage game payoff is $1-\theta$ and the buyer's payoff is b, where:

- $\triangleright$ $b > 0$ is the buyer's return from the seller's high effort;
- $\triangleright$ $c > 0$ is the buyer's loss from the seller's low effort (or *betrayal*);
- $\triangleright$ $\theta \in \Theta \equiv \{\theta_1, ...\theta_m\} \subset (0,1)$ is the seller's cost of exerting high effort, or more generally, player 1's temptation to betray her opponents' trust. Without loss, I assume $0 < \theta_1 < \theta_2 < ... < \theta_m < 1$.

The benefit and cost parameters, b and c, are common knowledge. The cost of high effort is the seller's private information, or her *type*. This assumption is reasonable when θ depends on the seller's production technology.

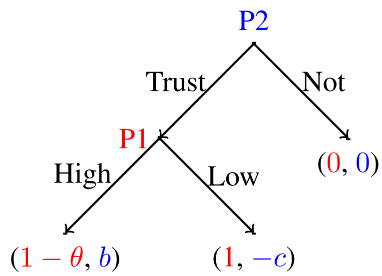

Figure 1: The stage game, where $\theta \in (0,1)$, $b>0$, $c>0$

Stage Game Equilibrium & Commitment Benchmark: The unique Nash equilibrium outcome in the stage game is N and the resulting payoff for the seller is 0. This is because the seller has a strict incentive to choose L after the buyer plays T, which motivates the latter to choose N.

Next, consider a benchmark scenario in which the seller could *pre-commit* to an (possibly mixed) action $\alpha_1 \in \Delta(A_1)$ before the buyer chooses between T and N. Every type will optimally commit to randomize between H and L, with the probability of playing H being $\gamma^* \equiv \frac{c}{b+c}$. For every $j \in \{1,2,...,m\}$, type θ_j's payoff under her optimal commitment is:

$$v_j^{**} \equiv 1 - \gamma^* \theta_j, \tag{2.1}$$

where v_j^{**} is her *Stackelberg payoff* and $\gamma^* H + (1-\gamma^*)L$ is her *Stackelberg action*.

The comparison between the seller's Nash equilibrium payoff and her Stackelberg payoff highlights a *lack-of-commitment* problem, which is of first order importance not only in business transactions (Mailath and Samuelson 2001, Ely and Välimäki 2003, Ekmekci 2011), but also in fiscal and monetary policies (Barro 1986, Phelan 2006), sovereign debt market (Cole, Dow and English 1995), corruption (Tirole 1996) and corporate finance (Tirole 2006). The rest of this article explores the extent to which the seller's persistent private information can mitigate her lack-of-commitment problem and improve her payoff when this stage game is played repeatedly.

2.2 The Repeated Game

Time is discrete, indexed by $t = 0,1,2,...$. The seller is interacting with an infinite sequence of buyers, arriving one in each period and plays the game only once. The stage game proceeds according to Figure 1. Players have access to a public randomization device in the beginning of each period, with $\xi_t \in [0,1]$ a typical realization.

The seller's cost of high effort, θ, is perfectly persistent over time and is her private information. The buyer's prior belief about θ is $\pi_0 \in \Delta(\Theta)$, which is assumed to have full support. Both players' action choices in the past can be perfectly observed. Let $a_t \in \{N,H,L\}$ be outcome in period t. Let $h^t = \{a_s, \xi_s\}_{s=0}^{t-1} \in \mathcal{H}^t$ be the public history in period t with $\mathcal{H} \equiv \bigcup_{t=0}^{+\infty} \mathcal{H}^t$ the set of public histories. Let $A_1 \equiv \{H, L\}$ and $A_2 \equiv \{T, N\}$. Let $\sigma_2 : \mathcal{H} \to \Delta(A_2)$ be the buyer's strategy. Let $\sigma_\theta : \mathcal{H} \to \Delta(A_1)$ be type θ's strategy, which specifies her action choices after receiving the buyers' trust. Strategy σ_θ is *stationary* if it takes the same value for all elements in $\mathcal{H}$. Let $\sigma_1 \equiv (\sigma_\theta)_{\theta \in \Theta}$ be the seller's strategy.

The seller discounts future payoffs by factor $\delta \in (0,1)$. Let $u_1(\theta, a_t)$ be the seller's stage game payoff when her cost is θ and the outcome is a_t. Type θ seller maximizes her expected discounted average payoff, given by:

$$\mathbb{E}^{(\sigma_\theta, \sigma_2)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t u_1(\theta, a_t)\Big], \tag{2.2}$$

with $\mathbb{E}^{(\sigma_\theta, \sigma_2)}[\cdot]$ the expectation over $\mathcal{H}$ under the probability measure induced by $(\sigma_\theta, \sigma_2)$. The seller's *payoff* in this repeated incomplete information game is summarized by a vector $v = (v_1, v_2, ..., v_m) \in \mathbb{R}^m$, with v_j the discounted average payoff of type θ_j.

Equilibrium Payoff Set: I introduce two versions of a patient seller's equilibrium payoff set: a lower version that adopts a stringent solution concept and takes the lower limit (in the set inclusion sense), as well as an upper version that adopts a permissive solution concept and takes the upper limit. Other versions of her payoff set are bounded between the two. I will show in Theorem 1 that the two versions coincide, which implies that the resulting characterization is robust against the choice of solution concepts and the ways of taking the limits.

Formally, let $\underline{V}(\pi_0, \delta) \subset \mathbb{R}^m$ be the set of payoffs the seller can attain in *sequential equilibrium* under parameter configuration $(\pi_0, \delta) \in \Delta(\Theta) \times (0,1)$. Let $\text{clo}(\cdot)$ be the closure of a set. The lower version of the patient seller's equilibrium payoff set is given by:

$$\underline{V}(\pi_0) \equiv \text{clo}\Big(\liminf_{\delta \to 1} \underline{V}(\pi_0, \delta)\Big).^{9} \tag{2.3}$$

Similarly, let $\overline{V}(\pi_0, \delta) \subset \mathbb{R}^m$ be the set of payoffs the seller can attain in *Nash equilibrium*. The upper version of the patient seller's equilibrium payoff set is given by:

$$\overline{V}(\pi_0) \equiv \text{clo}\Big(\limsup_{\delta \to 1} \overline{V}(\pi_0, \delta)\Big). \tag{2.4}$$

Complete Information Benchmark: When θ is common knowledge (or $m = 1$), the seller's equilibrium payoff cannot exceed $1 - \theta$ no matter how patient she is, which is strictly less than her Stackelberg payoff $1 - \gamma^*\theta$. Intuitively, this is because in every period where the buyer plays T, the seller needs to play H with

[9] For a family of sets $\{E_\delta\}_{\delta \in (0,1)}$, let $\liminf_{\delta \to 1} E_\delta \equiv \bigcup_{\delta' \in (0,1)} \bigcap_{\delta \geq \delta'} E_\delta$ and $\limsup_{\delta \to 1} E_\delta \equiv \bigcap_{\delta' \in (0,1)} \bigcup_{\delta \geq \delta'} E_\delta$.

strictly positive probability. Therefore, playing H in every period where she receives the buyer's trust is one of the seller's best replies to the buyer's strategy, from which her payoff in every period is at most $1-\theta$ and this leads to the payoff upper bound.[10] This conclusion implies that the seller's patience alone is not sufficient to overcome her lack-of-commitment problem when her opponents are not sufficiently forward-looking.

3 Main Results

In this section, I state the main results of the paper which examine the patient seller's payoff and behavior in the repeated incomplete information game (or $m \geq 2$). Theorem 1 characterizes a patient seller's equilibrium payoff set. I provide a tractable formula for every type's highest equilibrium payoff, which converges to her Stackelberg payoff when the lowest cost in the support of the prior belief vanishes. Theorem 2 examines a patient seller's behavior and shows that no type will use stationary strategies or completely mixed strategies in any Nash equilibrium that approximately attains her highest equilibrium payoff. This is in contrast to the classic examples of stationary commitment types who are mechanically playing the same mixed action in every period.

3.1 Equilibrium Payoffs

I define a payoff for every type of the seller and will relate this to her highest equilibrium payoff when she is patient. For every $\theta_j \in \Theta$, let

$$v_j^* \equiv \underbrace{(1-\gamma^*\theta_j)}_{\text{Type } \theta_j\text{'s Stackelberg payoff}} \underbrace{\frac{1-\theta_1}{1-\gamma^*\theta_1}}_{\text{incomplete information multiplier}}, \tag{3.1}$$

which is the product of type θ_j's Stackelberg payoff and an *incomplete information multiplier*. The latter summarizes the effect of incomplete information on the patient seller's payoff, which is common for all types, strictly below one and only depends on the lowest cost in the support of the prior belief. In another word, the effect of incomplete information is independent of the other possible costs and the prior probability of each type.

Let $v^* \equiv (v_1^*, ..., v_m^*)$. Let V^* be the convex hull of $\Big\{v^*, (0,0,...,0), (1-\theta_1, ..., 1-\theta_m)\Big\}$, with an example depicted in Figure 2. Theorem 1 claims that V^* is the set of payoffs a patient seller can attain in equilibrium.[11]

Theorem 1. $\overline{V}(\pi_0) = \underline{V}(\pi_0) = V^*$.

[10] Since outcome (T, H) is enforceable via grim-trigger strategies when the seller is patient, her equilibrium payoff set is $[0, 1-\theta]$ when δ is above some cutoff. A general folk theorem for this class of games is established by Fudenberg, Kreps and Maskin (1990).

[11] Despite Theorem 1 is stated in the context of the sequential move stage game with perfect monitoring, it can be generalized to stage games in which players move simultaneously, or the informed long-run player is choosing from a continuum of effort levels and her effort choice is observed with noise. Both extensions will be addressed in section 5.

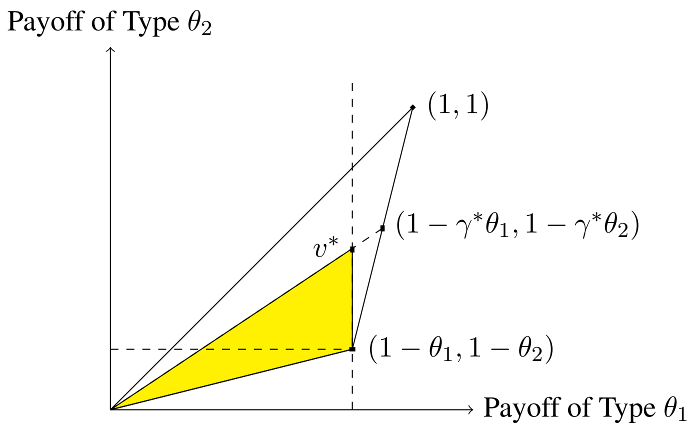

Figure 2: The limiting equilibrium payoff set V^* (in yellow) when $m = 2$.

The proof is in Appendices A and B with the intuitions behind explained in section 4. To better understand this result, note that set V^* is characterized by two linear constraints. First, the equilibrium payoff of the lowest cost type (i.e. type θ_1) cannot exceed her highest payoff in the repeated complete information game (i.e. $1 - \theta_1$). Intuitively, this is because she has no good candidate to imitate in the repeated incomplete information game.

Second, if one writes every feasible payoff vector as a convex combination of the payoffs from the three stage-game outcomes, namely, N, (T, H) and (T, L), then the ratio between the convex weight of (T, H) and the convex weight of (T, L) is no less than $\gamma^*/(1 - \gamma^*)$. This is because aside from a bounded number of periods, buyers will be able to predict the seller's future actions with arbitrarily high precision at every history they play trust. As a result, they will play trust at a history only when H will be played with probability above γ^*. I will later relate this to an upper bound on the relative speed between reputation improvement and reputation deterioration, which is also driven by the buyers' myopic incentives.

Implications: Next, I outline the economic implications of Theorem 1. To draw connections with the reputation literature, I will replace the seller with *long-run player* and the buyers with the *short-run players*.

First, the incomplete information multiplier only depends on θ_1 but not on the other details of the prior distribution. Since it is decreasing in θ_1, a decrease in the lowest cost can improve every other type's equilibrium payoff. Intuitively, by imitating the equilibrium strategy of type $\theta \in \Theta$, any type of the long-run player can build a reputation for behaving equivalently to type θ within a bounded number of periods, where the prior probability of type θ determines that actual number. This probability does not matter when the long-run player is patient as the payoff consequence for any bounded number of periods becomes negligible.[12] The presence of other costs in

[12]This argument relies on an implicit private value assumption that θ does not affect the short-run players' payoffs. In interdependent value environments, the state matters for the patient long-run player's payoff, as shown in Pei (2017).

the support of the prior belief (aside from the lowest one) has no impact on any type's highest equilibrium payoff either, as every type is strictly better off by imitating the lowest cost type when she is patient.

Second, every type aside from the lowest cost type can strictly benefit from incomplete information. Moreover, the incomplete information multiplier converges to 1 as θ_1 vanishes to 0. As a result, for every $j \in \{1, ..., m\}$, v_j^* converges to type θ_j's Stackelberg payoff v_j^{**}. Economically, this implies that under realistic informational assumptions, the patient long-run player can overcome her lack-of-commitment problem in the repeated incomplete information game by achieving her payoff under her optimal commitment.

Corollary 1. *For every $\epsilon > 0$, there exist $\overline{\delta} \in (0,1)$ and $\overline{\theta}_1 > 0$ such that when $\delta > \overline{\delta}$ and $\theta_1 < \overline{\theta}_1$, there exists a sequential equilibrium in which type θ_j's equilibrium payoff is no less than $v_j^{**} - \epsilon$ for all $j \in \{1, ..., m\}$.*

Third, in many applications of interest, it is also important to address questions related to social welfare. In what follows, I show that every payoff on the Pareto frontier is approximately attainable in sequential equilibrium when the long-run player is patient and the lowest cost is small. To be more precise, let $\overline{v} \equiv (v_0, v_1, ..., v_m) \in \mathbb{R}^{m+1}$ where v_0 is the discounted sum of the short-run player's payoff and v_j is type θ_j's discounted average payoff for every $j \geq 1$. Similar to Fudenberg and Levine (1994), I say that $\overline{v}$ is *incentive compatible* for the short-run players if there exists $(\alpha_1, a_2) \in \Delta(A_1) \times A_2$ such that:

$$a_2 \in \arg\max_{a_2' \in A_2} u_2(\alpha_1, a_2'), \tag{3.2}$$

$u_1(\theta_j, \alpha_1, a_2) = v_j$ for every $j \in \{1, 2, ..., m\}$ and $u_2(\alpha_1, a_2) = v_0$. Let $\overline{V}^* \subset \mathbb{R}^{m+1}$ be the convex hull of the set of incentive compatible payoff vectors, we have the following corollary:

Corollary 2. *For every $\epsilon > 0$, there exist $\overline{\delta} \in (0,1)$ and $\overline{\theta}_1 > 0$ such that for every $\delta > \overline{\delta}$, $\theta_1 < \overline{\theta}_1$ and $\overline{v}$ that is on the Pareto frontier of $\overline{V}^*$, there exists $\overline{v}'$ that is within ϵ of $\overline{v}$ such that $\overline{v}'$ is attainable in sequential equilibrium in the repeated incomplete information game without commitment.*

This conclusion follows from Corollary 1 as the Pareto frontier of $\overline{V}^*$ is the straight line connecting $(b, 1 - \theta_1, ..., 1 - \theta_m)$ and $(0, v_1^*, ..., v_m^*)$. As both extreme points are approximately attainable, every payoff vector in the interior of this line is also approximately attainable.

Connections to Canonical Reputation Models: Theorem 1 and Corollary 1 are reminiscent of a well-known conclusion in Fudenberg and Levine (1989, 1992) that a patient long-run player can approximately attain her commitment payoff from playing any action (pure or mixed) if with positive probability, she is a *commitment type* who is mechanically playing that action in every period.

Formally, let $\sigma_M^* : \mathcal{H} \to \Delta(A_1)$ be a *commitment strategy* and let $\Sigma_2^*(\sigma_M^*)$ be the set of player 2's complete information best replies to σ_M^*. Type θ's commitment payoff from playing σ_M^* is:

$$U(\sigma_M^*) \equiv \inf_{\sigma_2^* \in \Sigma_2^*(\sigma_M^*)} \Big\{ \mathbb{E}^{(\sigma_M^*, \sigma_2^*)} \Big[\sum_{t=0}^{\infty} (1-\delta)\delta^t u_1(\theta, a_t) \Big] \Big\}. \tag{3.3}$$

For every $\epsilon > 0$, σ_M^* is type θ's *ϵ-Stackelberg strategy* if $U(\sigma_M^*) \geq 1 - \gamma^*\theta - \epsilon$. When ϵ is small enough, every ϵ-Stackelberg strategy is non-trivially mixed. A classic example is the following *stationary ϵ-Stackelberg strategy*:

$$\sigma_M^*(h^t)[H] = \gamma' H + (1-\gamma')L \text{ for every } h^t \in \mathcal{H} \text{ where } \gamma' \in (\gamma^*, \gamma^* + \epsilon]. \tag{3.4}$$

Fudenberg and Levine (1989,1992) show that if the short-run players' prior belief attaches strictly positive probability to a *commitment type* who is mechanically playing σ_M^*, then there exist Nash equilibria in which a sufficiently patient long-run player's payoff is at least $U(\sigma_M^*) - \epsilon$.[13] For example, if with positive probability, the long-run player is mechanically playing one of her ϵ-Stackelberg strategies, then she can approximately attain her Stackelberg payoff in equilibrium.

Despite their results identify an interesting *reputation effect*, finding good ways to rationalize those mixed strategy commitment types remains an open question. Theorem 1 and Corollary 1 provide an affirmative answer by showing that in terms of the patient long-run player's highest attainable payoff, those mixed strategy commitment types can be replaced by rational types that have standard ordinal preferences over stage-game outcomes, but have different costs to exert high effort. For example, the Stackelberg commitment type can be rationalized by a strategic type that has very low albeit positive cost to exert high effort.

This additional requirement on the long-run player's ordinal preferences is motivated by economic applications where (1) all reputation building agents are facing temptations to renege; (2) their private benefits from reneging are not publicly observed. Mapping back into the applications, these imply that the following aspects of the game's payoff structure are common knowledge:

1. Firms can benefit from consumers' purchases, governments can benefit from FDI, central banks can better stimulate the economy when citizens to expect low inflation.

2. Providing high quality is costly for the firm, governments benefit from expropriating foreign investments, central banks benefit from high unexpected inflation.

On the other hand, the firm's cost of exerting high effort, the extent to which central banks trade-off inflation

[13] When the stage game is of simultaneous-move, she can guarantee that payoff in all Nash equilibria. The commitment payoff cannot be guaranteed when the stage-game is of sequential-move, as there always exists an equilibrium in which the short-run players never trust and the long-run player can never signal her private information.

and unemployment, etc. tend to be their private information. This realistic assumption on the long-run player's ordinal preferences also introduces new challenges to the proof, as motivating a strategic player to randomize between actions is difficult when she has strict preferences over stage-game outcomes.[14]

3.2 Equilibrium Behavior

In this subsection, I study the patient long-run player's behavior in equilibria that approximately attain v^*. This includes but not limited to her behavior in equilibria that approximately attain her Stackelberg payoff when θ_1 is small. Contrast to the classic examples of stationary commitment types who are mechanically playing the same mixed action in every period, I show in Theorem 2 that no type of the long-run player will play *stationary strategies* or *completely mixed strategies* in any such equilibrium.

Theorem 2. *For every small enough $\epsilon > 0$, there exists $\overline{\delta} \in (0,1)$, such that when $\delta > \overline{\delta}$, no type of the long-run player will play a completely mixed strategy or a stationary strategy in any Nash equilibrium that attains payoff within ϵ of v^*.*

The proof is in Appendix C. For some intuition, suppose a type's equilibrium strategy is completely mixed, then both playing L at every on-path history and playing H at every on-path history are her best replies. As the types can be vertically ranked such that lower cost types enjoy a comparative advantage in playing H, every type that has strictly higher cost will play L with probability 1 at every on-path history and every type that has strictly lower cost will play H with probability 1 at every on-path history.[15]

However, the presence of pure stationary strategies are at odds with the requirement that types θ_2 to θ_m can extract information rent in the long-run.[16] To see why, first, if there exists a type θ_j that plays L with probability 1 at every on-path history, then according to the learning argument in Fudenberg and Levine (1992), the short-run players will eventually believe that L will be played with probability close to 1 in all future periods, after which they will have a strict incentive to play N, leaving type θ_j a discounted average payoff close to 0 when δ is high enough. Next, suppose there exists a type θ_j that plays H with probability 1 at every on-path history, then she can never extract any information rent. Moreover, type θ_{j+1} will be the lowest cost type in the support of the

[14]The standard techniques to construct mixed strategy equilibria in repeated games, such as the belief-free equilibrium approach in Ely, Hörner and Olzewski (2005), Hörner and Lovo (2009) is not applicable in this context. This is because (1) the uninformed players are myopic, (2) in every equilibrium that approximately attains v^* when δ is close to 1, the uninformed players need to learn about the informed player's type in unbounded number of periods. That is to say, different types are mixing with different probabilities in the stage game. Therefore, these equilibria cannot be belief-free.

[15]To be more precise, once we order the states and actions according to $T \succ N$, $H \succ L$ and $\theta_1 \succ \theta_2 \succ ... \succ \theta_m$, our stage game payoff satisfies a monotone-supermodularity condition introduced in Liu and Pei (2017). This condition is sufficient to guarantee the monotonicity of the sender's strategy with respect to the state in all the Nash equilibria of one-shot signalling games. The proof of Theorem 2 uses the implication of this result on repeated signalling games, which is developed in Pei (2017).

[16]The validity of Theorem 2 relies on the presence of incomplete information, i.e. $m \geq 2$. In a repeated complete information game, there exist sequential equilibria that attain v^* in which the patient long-run player plays a stationary mixed strategy (Appendix D.1).

short-run players' posterior after the first time she plays L. This implies that type θ_{j+1} cannot extract information rent in the continuation game. To summarize, neither type θ_j nor type θ_{j+1} can extract information rent in the long-run, contradicting the hypothesis that their equilibrium payoffs are no less than $v_j^* - \epsilon$ and $v_{j+1}^* - \epsilon$.

I conclude this subsection with several remarks. First, the conclusion in Theorem 2 remains valid when $\theta_1 = 0$. This is somewhat surprising as the long-run player cannot be mixing at every history even when she is indifferent between high and low effort. Intuitively, this is because her action choices in the stage game can affect the frequency with which her opponents trust her in the future, and therefore, will affect the other types' incentives to imitate. If the zero-cost type is indifferent between H and L in every period where she receives her opponent's trust, then playing L at every history as well as playing H at every history will result in the same frequency with which player 2 plays T. Consequently, every other type will strictly prefer to play L at every on-path history. As shown in Fudenberg and Levine (1992), the short-run players will eventually learn that L will be played with very high probability in every future period, after which they will play N. As a result, no positive-cost type can extract information rent in the long-run, leading to a contradiction.

Second, will the stationary ϵ-Stackelberg strategy be played in other equilibria, for example ones that result in low payoffs for the long-run player? In the sequential-move stage game, the answer to this question depends on the choice of solution concepts. In Appendix D.2, I show that (1) the stationary ϵ-Stackelberg strategy won't be played by any type in any *sequential equilibrium*; (2) there exist *Nash equilibria* in which some type of the long-run player adopts the stationary ϵ-Stackelberg strategy.[17]

Third, given that stationary strategies and completely mixed strategies will not be played by the lowest-cost type, how will she behave in equilibria that approximately attain v^*? To see some necessary conditions for σ_{θ_1}, notice that type θ_m's equilibrium payoff cannot exceed $1 - \gamma^* \theta_m + \epsilon$ according to the payoff upper bound result in Fudenberg and Levine (1992). This implies an upper bound on the occupation measure of outcome (T, L) along every action path played by type θ_1 in equilibrium, which converges to γ^* as θ_1 vanishes. Moreover, the ϵ-Stackelberg strategy being played in any Stackelberg equilibrium also has the following feature: the expected occupation measure of (T, L) is close to $1-\gamma^*$ once we take the weighted-average across action paths. Intuitively, when θ_1 is small enough, the lowest-cost type *cherry-picks* her actions so that the discounted average frequency of each pure action along every infinite action path matches its probability in the mixed Stackelberg action.

[17] When the stage game is of simultaneous move, the stationary ϵ-Stackelberg strategy won't be played by any type in any Nash equilibrium for all small enough ϵ. This is because every type of the long-run player can guarantee payoff $1 - \gamma^* \theta - 2\epsilon$ when there exists a type that is playing an ϵ-Stackelberg strategy and δ is sufficiently high. Applying Theorem 2, one can obtain a contradiction.

4 Proof of Theorem 1: Intuition and Ideas

Theorem 1 is implied by the following pair of statements. First, every payoff that is bounded away from V^* is not attainable in any Nash equilibrium when δ exceeds some cutoff, i.e. $\overline{V}(\pi_0) \subset V^*$. Second, every payoff in the interior of V^* is attainable in some sequential equilibria when δ is high enough, i.e. $\underline{V}(\pi_0) \supset V^*$.

I explain the ideas behind the proof using an example with two types. The first statement hinges on understanding the necessity of the two constraints characterizing V^*. The second statement is shown by constructing equilibria that approximately attain v^* when δ close to 1. These equilibria feature slow learning and reputation building-milking cycles. Such arrangements enable her to extract information rent while preserving her informational advantage, which allow learning and rent extraction to occur in unbounded number of periods.

4.1 Necessity of Constraints

Recall that the limiting equilibrium payoff set V^* is characterized by two linear constraints:

1. The equilibrium payoff of type θ_1 cannot exceed $1-\theta_1$.
2. The ratio between the convex weight of (T, H) and that of (T, L) is no less than $\gamma^*/(1-\gamma^*)$.

Let $\sigma \equiv (\sigma_{\theta_1}, \sigma_{\theta_2}, \sigma_2)$ be a Nash equilibrium. To understand the necessity of the first constraint, it is instructive to define the long-run player's *highest action path*. Formally, let $\mathcal{H}(\sigma)$ be the set of on-path histories. For every $h^t \in \mathcal{H}(\sigma)$ such that $\sigma_2(h^t)[T] > 0$, let $\Theta(h^t)$ be the support of the short-run player's posterior belief at h^t. The highest action path is:

$$\overline{\sigma}_1(h^t) \equiv \begin{cases} H & \text{if } H \in \bigcup_{\theta \in \Theta(h^t)} \operatorname{supp}\big(\sigma_\theta(h^t)\big) \\ L & \text{otherwise .} \end{cases} \tag{4.1}$$

By definition, the short-run player has an incentive to play T at h^t only when $\overline{\sigma}_1(h^t) = H$. By construction, $\overline{\sigma}_1$ is at least one type's best reply to σ_2. Consider two cases separately: (1) If $\overline{\sigma}_1$ is type θ_1's best reply, then type θ_1's payoff in every period cannot exceed $1-\theta_1$, which implies that her discounted average payoff is no more than $1-\theta_1$. (2) If $\overline{\sigma}_1$ is type θ_2's best reply, then type θ_2's payoff in every period cannot exceed $1-\theta_2$. Since the difference between type θ_1 and type θ_2's payoff is at most $\theta_2-\theta_1$, type θ_1's discounted average payoff cannot exceed $1-\theta_1$. The necessity of the first constraint is obtained by unifying these cases.

Next, I explain the necessity of the second constraint. Recall from the conclusion in Fudenberg and Levine (1992) that if the long-run player plays according to the equilibrium strategy of type θ, then the short-run players' predictions about her actions will be close to type θ's strategy in all but a bounded number of periods. If the ratio between the occupation measure of (T, H) and that of (T, L) is strictly less than $\gamma^*/(1-\gamma^*)$, then the short-run players will learn about the long-run player's true strategy in finite time and will not trust her with high enough frequency in the future to attain the target payoff.

In what follows, I provide an alternative and more intuitive interpretation of this second constraint based on the relative speed of reputation building to reputation milking. For this purpose, I introduce an alternative version of the highest action path based on type θ_2's equilibrium strategy:

$$\overline{\sigma}_{\theta_2}(h^t) \equiv \begin{cases} H & \text{if } H \in \text{supp}\big(\sigma_{\theta_2}(h^t)\big) \\ L & \text{otherwise,} \end{cases} \tag{4.2}$$

By construction, $\overline{\sigma}_{\theta_2}$ is type θ_2's best reply to σ_2. If type θ_2 plays according to $\overline{\sigma}_{\theta_2}$, then her stage game payoff exceeds $1-\theta_2$ only at histories where the short-run player plays T but $\overline{\sigma}_{\theta_2}$ prescribes L. The short-run player's incentive constraint implies that at those histories, σ_{θ_1} needs to prescribe H with sufficiently high probability. Let $\eta(h^t)$ be the probability of type θ_1 at h^t, which I call *the long-run player's reputation*. The above argument implies that when type θ_2 plays according to $\overline{\sigma}_{\theta_2}$, she can only extract information rent (playing L) at the expense of her reputation, i.e. $\eta(h^t, L) < \eta(h^t)$. But nevertheless, she can rebuild her reputation in periods where $\overline{\sigma}_{\theta_2}$ prescribes H, i.e. $\eta(h^t, H) > \eta(h^t)$.

Now comes the key question: what is the *maximal frequency* of L relative to H under strategy profile $(\overline{\sigma}_{\theta_2}, \sigma_2)$? The answer depends on the relative speed with which the long-run player can rebuild her reputation (by playing H) to the speed with which her reputation deteriorates (by playing L). The short-run players' incentives to trust require that H to be played with probability at least γ^*, which bounds the relative speed of learning from above:

$$\frac{\eta(h^t, H) - \eta(h^t)}{\eta(h^t) - \eta(h^t, L)} \leq \frac{1-\gamma^*}{\gamma^*}. \tag{4.3}$$

According to (4.3), $(1-\gamma^*)/\gamma^*$ is the threshold relative frequency between L and H such that there exists a long-run player's strategy and its induced belief system under which:

1. The short-run players have the incentives to play T in every period.

2. Regardless of how the long-run player times her actions, her posterior reputation is no less than her initial reputation as long as L is played with relative frequency below this threshold.

This provides an economic interpretation of the second constraint characterizing V^*.

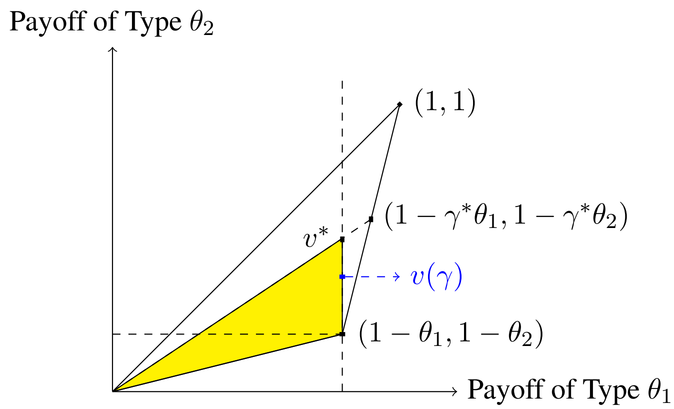

Figure 3: V^* in yellow and $v(\gamma)$ in blue for some $\gamma \in (\gamma^*, 1)$.

4.2 Overview of Equilibrium Construction

In this subsection, I construct a class of sequential equilibria that approximately attain payoff v^* when δ is large enough. In these equilibria, the long-run player can extract information rent only when her actions are informative about her type. Moreover, her reputation is gained and lost *gradually* in periods with active learning. For every $j \in \{1, 2, ..., m\}$ and $\gamma \in [\gamma^*, 1]$, let

$$v_j(\gamma) \equiv (1 - \gamma\theta_j)\frac{1-\theta_1}{1-\gamma\theta_1} \tag{4.4}$$

and $v(\gamma) \equiv \Big(v_j(\gamma)\Big)_{j=1}^{m}$. An example of $v(\gamma)$ is shown in Figure 3. By definition, $v_j(\gamma^*) = v_j^*$ and $v_j(1) = 1 - \theta_j$. My proof hinges on the following Proposition:

Proposition 4.1. *For every $\overline{\eta} \in (0, 1)$ and $\gamma \in (\gamma^*, 1)$, there exists $\overline{\delta} \in (0, 1)$, such that for every $\delta > \overline{\delta}$ and $\pi_0 \in \Delta(\Theta)$ with $\pi_0(\theta_1) \geq \overline{\eta}$, there exists a sequential equilibrium in which player 1's payoff is $v(\gamma)$.*

Since the other two vertices of V^* are attainable via the replication of stage-game Nash equilibrium and grim-trigger strategies, respectively, Proposition 4.1 also implies that every payoff vector in the interior of V^* is attainable when δ is large enough, i.e. $\underline{V}(\pi_0) \supset V^*$. The rest of this subsection consists of two parts. Part I provides an overview of players' strategies and the resulting systems of beliefs. Part II summarizes the ideas and economic intuitions behind the construction. The technical details can be found in Appendix A.

Part I: Equilibrium Strategies The constructed equilibrium has three phases: an *active learning phase* and two *absorbing phases*. I keep track of two state variables: (1) the probability with which player 2's posterior belief attaches to type θ_1, which I call *the long-run player's reputation*, denoted by $\eta(h^t)$; (2) the remaining occupation measure of each stage game outcome, denoted by $p^N(h^t)$, $p^H(h^t)$ and $p^L(h^t)$, respectively. The initial values of these state variables are given by:

$$\eta(h^0) = \pi_0(\theta_1)\,,\, p^N(h^0) = \frac{\theta_1(1-\gamma)}{1-\gamma\theta_1}\,,\, p^H(h^0) = \frac{(1-\theta_1)\gamma}{1-\gamma\theta_1} \text{ and } p^L(h^0) = \frac{(1-\theta_1)(1-\gamma)}{1-\gamma\theta_1}.$$

One can verify that for every $j \in \{1, 2, ..., m\}$,

$$v_j(\gamma) = p^H(h^0)(1-\theta_j) + p^L(h^0).$$

That is to say, the initial value of p^a is the convex weight of outcome a according to the target payoff $v(\gamma)$.

Play starts from the *active learning phase*, in which player 2 plays T in every period and player 1 is mixing between H and L with probabilities pinned down by the following belief-updating formulas:

$$\eta(h^t, L) - \eta^* = (1-\lambda\gamma^*)(\eta(h^t) - \eta^*) \tag{4.5}$$

and

$$\eta(h^t, H) - \eta^* = \min\Big\{1-\eta^*, \big(1+\lambda(1-\gamma^*)\big)\big(\eta(h^t)-\eta^*\big)\Big\}, \tag{4.6}$$

where η^* is an arbitrary number within $\Big(\gamma^*\eta(h^0), \eta(h^0)\Big)$ and λ is some positive number. Intuitively, η^* is the lower bound on the long-run player's reputation in the active learning phase, which is to satisfy player 2's incentive constraint by the end of this phase, with details explained in Part I of Appendix A.3. The parameter λ measures the *absolute speed* of learning, which is required to be small enough with details specified in (A.5). Intuitively, according to the belief updating formulas in (4.5) and (4.6), if the relative frequencies between L and H is strictly less than $(1-\gamma^*)/\gamma^*$, then the Taylor's theorem implies that the long-run player's reputation will increase compared to her initial reputation when learning is sufficiently slow, i.e. λ is small enough.

The other state variable, $p^a(h^t)$, evolves in response to the realized stage game outcome, with

$$p^a(h^t, a_t) \equiv \begin{cases} p^a(h^t) & \text{if } a_t \neq a \\ p^a(h^t) - (1-\delta)\delta^t & \text{if } a_t = a \end{cases} \tag{4.7}$$

with $a, a_t \in \{N, H, L\}$. Intuitively, the long-run player has three separate accounts, each of them represents the occupation measure of an individual stage game outcome. If outcome a is realized in period t, then her account for outcome a is deducted by $(1-\delta)\delta^t$. The state variable $p^a(h^t)$ is then interpreted as the remaining credit in the account for outcome a.

Play transits to the first absorbing phase when $\eta(h^t)$ reaches 1, after which the continuation value of type θ_j

is

$$v_1(h^t)\frac{1-\theta_j}{1-\theta_1} \text{ for every } j \in \{1,2,...,m\},$$

with

$$v_1(h^t) \equiv \frac{p^H(h^t)(1-\theta_1)+p^L(h^t)}{p^H(h^t)+p^L(h^t)+p^N(h^t)}. \tag{4.8}$$

The resulting payoff vector can be delivered by randomizing between outcomes N and (T,H).

Play transits to the second absorbing phase when $p^L(h^t)$ is less than $(1-\delta)\delta^t$, or intuitively, the remaining credit in the account for (T,L) is low enough such that playing L for another period will make it negative. For illustration purposes, I will focus on the ideal situation in which $p^L(h^t)=0$ and will relegate the discussions on the complications related to $p^L(h^t) \in \Big(0,(1-\delta)\delta^t\Big)$ to Appendix A. Type θ_j's continuation payoff then equals to:

$$v_j(h^t) \equiv \frac{p^H(h^t)(1-\theta_j)+p^L(h^t)}{p^H(h^t)+p^L(h^t)+p^N(h^t)}. \tag{4.9}$$

The resulting payoff vector $\Big(v_j(h^t)\Big)_{j=1}^m$ can be delivered by randomizing between N and (T,H).

Part II: Ideas & Intuitions This equilibrium is designed to maximize the frequency of outcome (T,L) while at the same time, providing incentives for the short-run players to trust and all types of the long-run player to randomize with the desired probabilities. For this purpose, this construction has two key features:

1. The rent extraction outcome (T,L) only occurs in the active learning phase.

2. Throughout the active learning phase, the long-run player is facing a trade-off between extracting information rent (by playing L) and building up her reputation (by playing H).

The first feature is necessary as learning is required for any type of the long-run player to extract information rent when her opponents are myopic. The second feature enables the long-run player to *rebuild* her reputation after extracting information rent. Importantly, rent extraction and reputation rebuilding occur in unbounded number of periods, which enables the long-run player to obtain information rent when δ is arbitrarily close to 1.

My construction of the active learning phase raises two issues, which motivate the design of the two absorbing phases. The first concern is that type θ_2 may have incentives to front-load rent extraction by rarely playing H. This explains the presence of the second absorbing phase. In particular, if she plays L too frequently, then she will reach the second absorbing phase at an earlier date after which there will be no information rent in the future.

The second concern is that the play of H can be too front-loaded, after which there is too much L remaining in the long-run player's account and the resulting continuation payoff cannot be delivered in an incentive compatible way. The first absorbing phase is designed to address this issue: if she front-loads the play of H, then play will transit to the first absorbing phase, after which every high-cost type's continuation payoff is strictly less compared to her payoff from playing L at the transition history. In general, the presence of the first absorbing phase ensures that at every history of the active learning phase, the ratio between the remaining occupation measures of L and H cannot exceed some cutoff. As a result, the long-run player's continuation value is always within V^*. This is stated as Lemma A.1, with the challenges and proof explained in Appendix A.

4.3 Comparisons on Equilibrium Dynamics

In this subsection, I compare the equilibrium dynamics in my model to the ones in the related literature. This includes reputation models with behavioral biases (Jehiel and Samuelson 2012), models of reputation building-milking cycles (Sobel 1985, Phelan 2006, Liu 2011, Liu and Skryzpacz 2014) and models that exhibit gradual learning (Benabou and Laroque 1992, Ekmekci 2011).

Analogical-Based Reasoning Equilibria: The behavioral patterns according to the high-cost type's highest action path ($\overline{\sigma}_{\theta_2}$ defined in subsection 4.1) is reminiscent of those in the analogy-based reasoning equilibria of Jehiel and Samuelson (2012). That is, the strategic long-run player alternates between two of her actions in order to manipulate her opponents' beliefs about her type.

In their model, there are multiple commitment types of the long-run player that are playing stationary mixed strategies and one strategic type who can flexibly choose her actions. The short-run players adopt an analogy-based reasoning procedure, namely, they mistakenly believe that the strategic long-run player is playing a stationary strategy. Their results imply, in context of the trust game, that the strategic long-run player can attain her Stackelberg payoff and her equilibrium behavior will experience a *reputation building phase* (or a *reputation consumption phase*) in which she plays H (or L) for a bounded number of periods, followed by a *reputation manipulation phase* that resembles the active learning phase in my model where she alternates between H and L. Moreover, the short-run players' posterior belief will fluctuate within a small neighborhood of the threshold belief, implying that the long-run player's type is never fully revealed.

In my model, despite type θ_2's behavior following her highest action path exhibits a similar pattern, there are two qualitative differences that highlight the distinctions between rational and analogical-based short-run players. First, the expected duration of the reputation manipulation phase is finite in my model due to the constraint that type θ_1's equilibrium payoff cannot exceed $1-\theta_1$. This constraint is driven by the rational short-run players' ability to correctly predict the long-run player's average action in *every period*, while analogy-based ones can only correctly predict the long-run player's average action *across all periods*. Second, the short-run players can learn the true state with positive probability in every equilibrium that approximately attains v^*, while in Jehiel

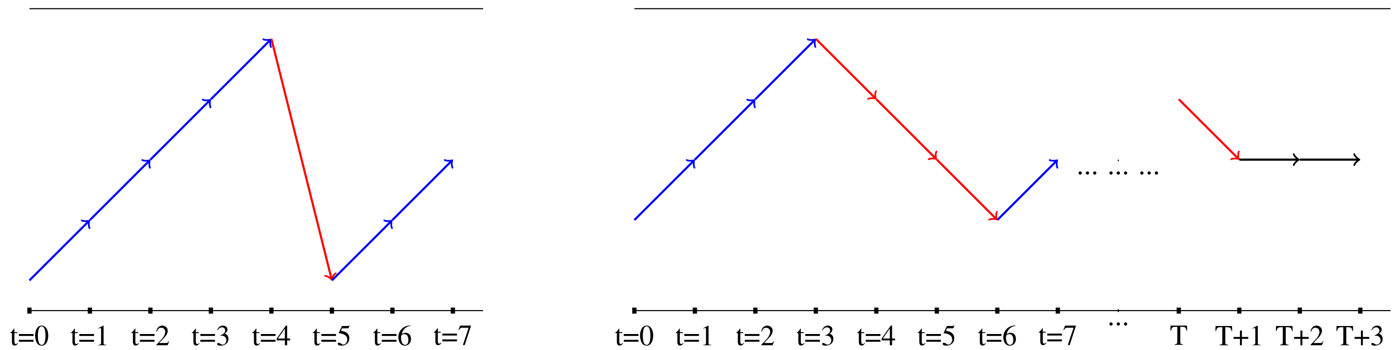

Figure 4: The horizontal axis represents the timeline and the vertical axis measures the informed player's reputation, i.e. probability of the commitment type or lowest cost type. Left: Reputation cycles in Phelan (2006). Right: A sample path of the reputation cycle in my model.

and Samuelson (2012), the probability with which they learn about the state is zero. This is because analogy-based short-run players' posterior beliefs only depend on the empirical frequencies of the observed actions. As a result, their posterior beliefs are not responsive enough to each individual observation.

Reputation Building-Milking Cycles: The behavioral pattern that a patient agent builds her reputation in order to milk it in the future has been identified in existing models of reputation building with commitment types, such as Sobel (1985), Phelan (2006), Liu (2011) and Liu and Skrzypacz (2014). Two differences in terms of the reputation dynamics emerge once comparing those models to mine. First, as in Jehiel and Samuelson (2012), the reputation cycles in Phelan (2006), Liu (2011) and Liu and Skrzypacz (2014) can last forever while the expected duration of the active learning phase is finite in mine.

Second and more importantly, reputations are built and milked *gradually* in my model while in theirs, the agent's reputation falls to its lower bound every time she milks it.[18] This is because the commitment types in their models never betray so one misbehavior will reveal the strategic long-run player's type. In my model, the comparison between good and bad types is less stark in the sense that all types are sharing the same ordinal preferences over stage-game outcomes and therefore, are tempted to betray their opponents' trust. In the long-run player's optimal equilibrium, the lowest-cost type will betray with positive probability for unbounded number of periods, which does not hurt her own payoff while at the same time, covering up the other types when they are milking reputations.

This feature of gradual learning is supported empirically by several studies of online markets. For example as

[18]In models with perfectly persistent type such as Sobel (1985), the agent loses her reputation forever after cheating. In models where types are changing over time such as Phelan (2006), the agent's reputation falls to its lower bound after cheating but it can be rebuilt over time. In models with limited memories such as Liu (2011) and Liu and Skrzypacz (2014), the agent's reputation is cleaned up after the bad record disappears from the her opponents' memory.

documented in Dellarocas (2006), consumers judge the quality of sellers based on their reputation scores, which are usually obtained via averaging the ratings they obtained in the past. In particular, one recent negative rating will neither significantly affect the amount of sales nor the prices of a reputable seller who has obtained many positive ratings in the past. This observation is closer to the equilibrium dynamics of my model compared to the aforementioned models of reputation cycles.

Benabou and Laroque (1992) and Ekmekci (2011) study games with commitment types and the long-run player's actions are imperfect monitored.[19] In their equilibria, learning also happens gradually as the short-run players cannot tell the difference between intended cheating and exogenous noise. In contrast, my model has perfect monitoring but no commitment type. Gradual learning occurs as the reputational type also cheats with positive probability in unbounded number of periods. The different driving forces behind gradual learning also lead to different long-run learning outcomes. For example, the short-run players can never fully learn the state in their models, while in mine, they will perfectly learn the state with positive probability in every equilibrium where some types of the patient long-run player can benefit from her persistent private information.

5 Extensions

In this section, I explore two alternative versions of the trust game that can be mapped into a variety of applications. This includes stage-game with simultaneous moves, games in which the long-run player's actions are imperfectly monitored, etc. The common features shared by these extensions and the baseline model are (1) all types of the patient long-run player are facing lack-of-commitment problems, including the reputational types and (2) the long-run player has persistent private information about her benefit from betrayal.

Trust Games with Continuum of Effort and Imperfect Monitoring: Player 1 is an agent, for example, a worker, supplier or private contractor. In every period, a principal (player 2, e.g. employer, final good producer) is randomly matched with the agent and decides whether to incur a fixed cost and interact with her or skip the interaction.[20] The agent chooses her effort from a closed interval unbeknownst to the principal and the probability with which the service quality being high increases with her effort. In line with the literature on incomplete contracts, the service quality is not contractible but is observable to the agent and all the subsequent principals.

[19]Benabou and Laroque (1992) study an extension of Sobel (1985) in which a reputation building sender receives noisy signals about an i.i.d. state in every period. Noisy monitoring arises as the receiver cannot distinguish between the sender's intentional deceptions and her unintentional mistakes. Ekmekci (2011) constructs a rating system that censors the short-run players' information about the long-run player's past actions. He constructs equilibrium in which the Stackelberg payoff is attained and reputations are sustained in the long-run.

[20]Interpretations of this fixed cost includes, an upfront payment made by the final good producer to his supplier, a relationship specific investment the principal needs to make in order to collaborate with the agent, etc.

The cost of effort is linear and the marginal cost of effort is the agent's persistent private information.[21]

Formally, the stage game is of sequential-move as in the baseline model, with the difference that after player 2 choosing T, player 1 chooses among a continuum of effort levels $e \in [0, 1]$ and the output being produced ($y \in \{G, B\}$) is good (i.e. $y = G$) with probability e. The cost of effort for type θ_i is $\theta_i e$. Player 1's benefit from her opponent's trust is normalized to 1. Therefore, her stage game payoff under outcome N is 0 and that under outcome (T, e) is $1 - \theta_i e$. Player 2's payoff is 0 if he chooses N. His benefit from good output is b while his loss from bad output is c, with $b, c > 0$. Therefore, player 2 is willing to trust only when player 1's expected effort is no less than $\gamma^* \equiv \frac{c}{b+c}$.

Consider the repeated version of this game in which the public history consists of player 2's actions, the realized outputs and the realizations of public randomization devices. In another word, player 1's effort choice is her private information. Formally, let $a_{1,t}$, $a_{2,t}$ and y_t be player 1's action, player 2's action and the realization of public signal in period t, respectively. Let $h^t = \{a_{2,s}, y_s, \xi_s\}_{s=0}^{t-1} \in \mathcal{H}^t$ be a public history with $\mathcal{H} \equiv \bigcup_{t=0}^{+\infty} \mathcal{H}^t$ the set of public histories. Let $h_1^t = \{a_{1,s}, a_{2,s}, y_s, \xi_s\}_{s=0}^{t-1} \in \mathcal{H}_1^t$ be player 1's private history with $\mathcal{H}_1 \equiv \bigcup_{t=0}^{+\infty} \mathcal{H}_1^t$ the set of private histories. Let $\sigma_2 : \mathcal{H} \to \Delta(A_2)$ be player 2's strategy and let $\sigma_\theta : \mathcal{H}_1 \to \Delta(A_1)$ be type θ player 1's strategy, with $\sigma_1 \equiv (\sigma_\theta)_{\theta \in \Theta}$.

In this setting, the long-run player's Stackelberg commitment payoff, $1 - \gamma^* \theta$, has two interpretations. First, as in the baseline model, it is her equilibrium payoff when she can commit. Second, it is also her highest equilibrium payoff in the repeated complete information game when her past action choices are *perfectly monitored*. However, in the repeated complete information game where θ is common knowledge but player 1's action choices are imperfectly observed, her highest equilibrium payoff is $1 - \theta$, which follows from Fudenberg and Levine (1994).

The characterization of the patient long-run player's equilibrium payoff set in Theorem 1 directly carries over to this setting. Intuitively, this is because one can substitute each type of long-run player's mixed action in the baseline model with a deterministic effort level, equals to the probability of high effort. Given the two interpretations of the Stackelberg payoff, my result implies that persistent private information can overcome the lack-of-commitment problem and/or the imperfect monitoring problem, which enables the patient long-run player to achieve the same payoff as if she can commit or when her actions are perfectly observed. In terms of the behavior, one can show a result similar to Theorem 2 that no type will play stationary strategies in any equilibrium that approximately attains v^*.

[21] Chassang (2010) studies a game with similar incentive structures, besides that the agent's cost of effort is common knowledge but the set of actions that are available in each period (which is i.i.d. across different periods) is the agent's private information. Tirole (1996) uses a similar model to study the collective reputations for commercial firms and the corruption of bureaucrats.

Simultaneous-Move Stage Games: I examine the simultaneous-move version of the trust game with stage-game payoffs given by:

-	T	N
H	$1-\theta, b$	$-d(\theta), 0$
L	$1, -c$	$0, 0$

where $b, c > 0$, $\theta \in \Theta \equiv \{\theta_1, \theta_2, ..., \theta_m\} \subset (0,1)$ is player 1's persistent private information and $d(\theta) > 0$ measures player 1's loss when she exerts high effort while player 2 does not trust. In what follows, I offer three interpretations of this payoff matrix in different contexts.

1. **Product Choice Game:** Consider the following game introduced by Mailath and Samuelson (2001), Ekmekci (2011), etc. Player 1 is a seller choosing between high quality and low quality. Player 2 is a buyer choosing between buy and not buy. Players' stage game payoffs are given by:

-	Buy	Not Buy
High Effort/Quality	$1-\theta, b$	$-d(\theta), 0$
Low Effort/Quality	$1, -c$	$0, 0$

   As in the baseline model, θ and $d(\theta)$ are the costs of providing high quality when the buyer buys and not buy, respectively. My model incorporates the *separable payoff case* in Ekmekci (2011) where $\theta = d(\theta)$.

2. **Entry Deterrence/Limit Pricing Game:** Player 1 is an incumbent choosing between a low price (or *fight*) and a normal price (or *accommodate*). Player 2 is the entrant deciding between entering the market and staying out. Players' payoffs are given by:

-	Out	Enter
Low Price	$1-\theta, 0$	$-d(\theta), -b$
Normal Price	$1, 0$	$0, c$

   where θ and $d(\theta)$ are the incumbent's costs from limit pricing (if the entrant stays out) and predation (if the entrant enters), respectively. As in Milgrom and Roberts (1982), θ depends on the efficiency of her production technology, which tends to be her private information. When studying the incumbent's payoff and equilibrium behavior, this is equivalent to the previous payoff matrix as the short-run players' incentives only depend on the their gains from staying out conditional on θ and the incumbent's action.

3. **Monetary Policy:** Player 1 is a central bank that is facing a continuum of households (player 2s), each has negligible mass. In every period, the central bank chooses the inflation level while at the same time,

households form their expectations about inflation. To simplify matters, I assume that both the actual inflation and the expected inflation are binary variables. In line with the classic work of Barro (1986), players' stage game payoffs are given by:

-	Low Expectation	High Expectation
Low Inflation	$1-\theta, x_1$	$-d(\theta), -y_1$
High Inflation	$1, -y_2$	$0, x_2$

where $x_1, x_2, y_1, y_2 > 0$ are parameters, $\theta \in \Theta \subset (0,1)$ is the central bank's private information. To make sense of this payoff matrix, households want to match their expectations with the actual inflation. The central bank's payoff decreases with the actual inflation and increases with the amount of surprised inflation (defined as actual inflation minus expected inflation). As argued in Barro (1986), the central bank can strictly benefit from surprised inflation as it can increase real economic activities, decrease unemployment rate and increase governmental revenue. How the central bank trades-off these benefits with the costs of inflation is captured by θ, which depends on the central banker's ideology and tends to be her persistent private information. Economically, the parametric assumption that $\theta < 1$ implies that inflation is costly for the central bank if it is perfectly anticipated by the households.

In what follows, I analyze the repeated version of this simultaneous-move stage game that is played over the infinite time horizon. As in the baseline model, players' past action choices are perfectly monitored and the public history $h^t \equiv \{a_{1,s}, a_{2,s}, \xi_s\}_{s=0}^{t-1}$ consists of players' past action choices and the past realizations of the public randomization device. Other features of the game remain the same as in the baseline model. Recall the definition of $v^* \equiv (v_j^*)_{j=1}^m$ in (3.1) and $\underline{V}(\pi_0)$ in (2.3), we have the following result, which shows that the patient long-run player can attain v^* in the repeated incomplete information game.

Corollary 3. *If π_0 has full support, then $v^* \in \underline{V}(\pi_0)$.*

The proof of Corollary 3 follows along the same line as Theorem 1, which can be found in Appendix A. Intuitively, this is because in the equilibrium I construct, only outcomes (T, H), (T, L) and (N, L) occur with positive probability along the equilibrium path. An implication of this corollary is the attainability of every type of long-run player's Stackelberg payoff when the lowest cost in the support of π_0 vanishes.

Corollary 4. *For every $\epsilon > 0$, there exists $\overline{\delta} \in (0,1)$ and $\overline{\theta}_1 \in (0,1)$, such that when $\delta > \overline{\delta}$ and $\theta_1 < \overline{\theta}_1$, there exists a sequential equilibrium in which type θ_j's payoff is no less than $v_j^{**} - \epsilon$ for all $j \in \{1, 2, ..., m\}$.*

Furthermore, under a supermodularity condition on player 1's stage game payoff function, i.e. an assumption on $d(\cdot)$, one can provide a full characterization on the game's limiting equilibrium payoff set.

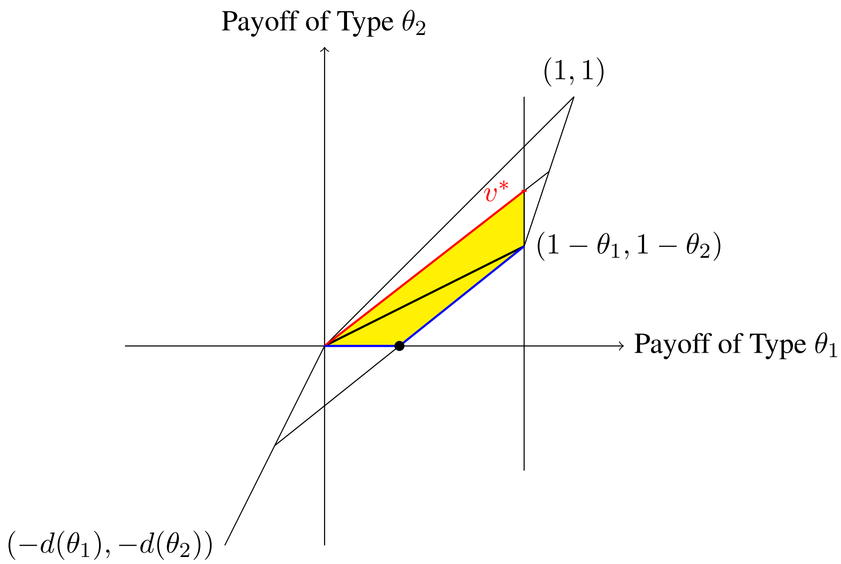

Figure 5: A patient player 1's equilibrium payoff set when the stage game is of simultaneous-move and $m = 2$.

Condition 1. *u_1 is supermodular if $0 \leq d(\theta_j) - d(\theta_i) \leq \theta_j - \theta_i$ for every $j < i$.*

Intuitively, when we rank the types and players' actions according to $\theta_1 \succ \theta_2 \succ ... \succ \theta_m$, $H \succ L$ and $T \succ N$, Condition 1 implies that u_1 is supermodular in θ and (a_1, a_2). For some concrete examples of u_1 that satisfies supermodularity, first, when $d(\theta_j) = 0$ for every $j \in \{1, 2, ..., m\}$, then the stage game payoff function is the same as in the sequential move game; second, when $d(\theta_j) = \theta_j$ for every $j \in \{1, 2, ..., m\}$, then the long-run player's cost of playing H is independent of the short-run player's actions, which is the case analyzed in Ekmekci (2011). Let $\overline{V}_j(\pi_0)$ be the projection of $\overline{V}(\pi_0)$ on the jth-coordinate, we have the following payoff upper bound for each type of the long-run player:

Corollary 5. *If u_1 is supermodular, then $\max \overline{V}_j(\pi_0) \leq v_j^*$ for every $j \in \{1, 2, ..., m\}$.*

Corollary 5 is shown in Appendix B together with the necessity part of Theorem 1. For the characterization of a patient long-run player's equilibrium payoff set, recall that a payoff vector $v \in \mathbb{R}^m$ is *incentive compatible* if there exists $(\alpha_1, a_2) \in \Delta(A_1) \times A_2$ such that:

$$a_2 \in \arg\max_{a_2' \in A_2} u_2(\alpha_1, a_2') \tag{5.1}$$

and $u_1(\theta_j, \alpha_1, a_2) = v_j$ for every $j \in \{1, 2, ..., m\}$. The patient long-run player's equilibrium payoff set V^* can be obtained via the following algorithm:

1. Take the convex hull of the set of payoffs that are incentive compatible.

2. Intersect this convex hull with two sets of constraints: $v_1 \leq 1-\theta_1$ and $v_j \geq 0$ for every $j \in \{1, ..., m\}$.

An example of this set is depicted in Figure 5. The proof is similar to that of Theorem 1. The key step is to show the attainability of the payoff vector marked as the black dot, which uses outcomes (T, H), (N, H) and (N, L) in a similar way as we used outcomes (T, H), (N, H) and (N, L) to construct equilibria that approximately attains v^* in the sufficiency proof of Theorem 1. The idea is to maximize the play of H while the short-run players do not trust. The details are available upon request.

6 Conclusion

This paper introduces a model of reputation building that has three important features: (1) the reputation building player values her opponent's trust but has a strict incentive to betray; (2) all types of the reputation building player are rational and share the same ordinal preferences over stage-game outcomes; (3) the reputation building player has persistent private information about her benefit from betrayals. This framework addresses several realistic concerns when studying trust building between buyers and sellers, citizens and governments, etc. in which lack-of-commitment problems are universal and the private benefits from betrayals are hard to observe.

My results address a patient reputation building player's equilibrium payoff and behavior in this repeated incomplete information game. Despite no type is immune to reneging temptations, the long-run player can still overcome her lack-of-commitment problem and attain her commitment payoff, which includes her Stackelberg payoff when the lowest cost in the support of the prior belief vanishes. This provides a partial strategic foundation for those mixed strategy commitment types in the reputation literature. My formula for the patient long-run player's highest equilibrium payoff cleanly separates the effects of incomplete information from those of her true cost of effort. The former is summarized by a multiplier that only depends on the lowest cost in the support of the short-run players' prior belief. In every equilibrium that approximately attains this highest equilibrium payoff, the patient long-run player's strategy must be non-stationary and reputations are built and milked gradually. This arrangement allows her to extract information rent in unbounded number of periods while preserving her informational advantage, which increases her payoff in long-term relationships.

A Proof of Theorem 1: Sufficiency

In this Appendix, I show that $V^* \subset \underline{V}(\pi_0)$ by constructing sequential equilibria that can attain any payoff vector in the interior of V^* when δ is sufficiently high. Recall that for every $\gamma \in [0,1]$,

$$v_j(\gamma) \equiv (1-\gamma\theta_j)\frac{1-\theta_1}{1-\gamma\theta_1}, \tag{A.1}$$

and $v(\gamma) \equiv \Big(v_j(\gamma)\Big)_{1\leq j\leq m}$. The rest of this subsection shows Proposition 4.1, which claims that for every $\gamma \in (\gamma^*, 1)$, $v(\gamma)$ is attainable in sequential equilibrium when δ is large enough. In subsection A.1, I define several variables that are key to my construction. In subsection A.2, I describe players' strategies and belief systems. In subsection A.3, I verify players' incentive constraints and the consistency of their beliefs.

A.1 Defining the Variables

In this subsection, I define several variables that are critical for my construction. I will also specify how large $\overline{\delta}$ needs to be for every given $\gamma \in (\gamma^*, 1)$ and $\pi_0(\theta_1)$, i.e. the long-run player's initial reputation.

Fixing $\gamma \in (\gamma^*, 1)$, there exists a rational number $\widehat{n}/\widehat{k} \in (\gamma^*, \gamma)$ with $\widehat{n}, \widehat{k} \in \mathbb{N}$. Moreover, there exists an integer $j \in \mathbb{N}$ such that

$$\frac{\widehat{n}}{\widehat{k}} = \frac{\widehat{n}j}{\widehat{k}j} < \frac{\widehat{n}j}{\widehat{k}j-1} < \gamma.$$

Let $n \equiv \widehat{n}j$ and $k \equiv \widehat{k}j$. Let

$$\widetilde{\gamma} \equiv \frac{1}{2}\Big(\frac{n}{k} + \frac{n}{k-1}\Big), \tag{A.2}$$

and

$$\widehat{\gamma} \equiv \frac{1}{2}\Big(\frac{n}{k} + \gamma^*\Big). \tag{A.3}$$

Let $\overline{\delta}_1 \in (0,1)$ to be large enough such that for every $\delta > \overline{\delta}_1$,

$$\frac{\delta+\delta^2+...+\delta^n}{\delta+\delta^2+...+\delta^k} < \widetilde{\gamma} < \frac{\delta^{k-n-1}(\delta+\delta^2+...+\delta^n)}{\delta+\delta^2+...+\delta^{k-1}}. \tag{A.4}$$

By construction, $\gamma^* < \widehat{\gamma} < \frac{n}{k} < \widetilde{\gamma} < \frac{n}{k-1} < \gamma$. Let $\eta(h^0) \equiv \pi_0(\theta_1)$, which is the probability of type θ_1 according to player 2s' prior belief. Let η^* be an arbitrary real number satisfying:

$$\eta^* \in \Big(\gamma^*\eta(h^0), \eta(h^0)\Big).$$

Let $\lambda > 0$ be small enough such that:

$$\Big(1+\lambda(1-\gamma^*)\Big)^{\widehat{\gamma}}\Big(1-\lambda\gamma^*\Big)^{1-\widehat{\gamma}} > 1. \tag{A.5}$$

Given $\gamma^* < \widehat{\gamma}$, the existence of such λ is implied by the Taylor's Theorem. Let $X \in \mathbb{N}$ be a large enough integer such that

$$\Big(1+\lambda(1-\gamma^*)\Big)^{X-1} > \frac{1-\eta^*}{\eta(h^0)-\eta^*}. \tag{A.6}$$

Let

$$Y \equiv \frac{1}{2}\underbrace{\left(\gamma-(1-\gamma)\frac{\widetilde{\gamma}}{1-\widetilde{\gamma}}\right)}_{>0}\frac{1-\theta_1}{1-\gamma\theta_1}, \tag{A.7}$$

which is strictly positive. Let $\overline{\delta}_2 \in (0,1)$ be large enough such that for every $\delta > \overline{\delta}_2$,

$$Y > \max\left\{1-\delta^X, \frac{1-\delta}{1-\gamma}\right\} \text{ and } \frac{\delta-\theta_1}{1-\theta_1} > \frac{1-\delta}{1-\gamma}. \tag{A.8}$$

The existence of such $\overline{\delta}_2$ is implied by $\widetilde{\gamma} < \gamma$.

Let $\overline{\delta} \equiv \max\{\overline{\delta}_1, \overline{\delta}_2\}$, which will be referred to as the *cutoff discount factor*. Let v^L, v^H and $v^N \in \mathbb{R}^m$ be player 1's payoff vectors from terminal outcomes L, H and N, respectively. The target payoff vector $v(\gamma)$ can be written as the following convex combination of v^L, v^H and v^N:

$$v(\gamma) = \underbrace{\frac{\theta_1(1-\gamma)}{1-\gamma\theta_1}}_{\equiv p^N} v^N + \underbrace{\frac{(1-\theta_1)\gamma}{1-\gamma\theta_1}}_{\equiv p^H} v^H + \underbrace{\frac{(1-\theta_1)(1-\gamma)}{1-\gamma\theta_1}}_{\equiv p^L} v^L, \tag{A.9}$$

with p^N, p^H and p^L being the convex weights of outcomes N, H and L, respectively.

Importantly, for every $\overline{\delta}$ that meets the above requirements under $\eta(h^0)$, it also meets all the requirements under every $\eta'(h^0) \geq \eta(h^0)$. This is because the required X decreases with $\eta(h^0)$, so an increase in $\eta(h^0)$ only slackens inequality (A.8) while having no impact on the other requirements.

A.2 Three-Phase Equilibrium

In this subsection, I describe players' strategies and player 2s' belief system. Players' sequential rationality constraints and the consistency of their beliefs are verified in the next step. Every type other than type θ_1 follows the same strategy, which is called *high cost types*, while type θ_1 is called the *low cost type*. Let $\eta(h^t)$ be the probability player 2s' posterior belief at h^t attaches to type θ_1. Recall the definition of η^*, which I will refer to as the *belief lower bound*. Let

$$\Delta(h^t) \equiv \eta(h^t) - \eta^*, \tag{A.10}$$

which is the gap between player 2s' posterior belief and the belief lower bound.

State Variables: The equilibrium keeps track of the following set of state variables: $\Delta(h^t)$ as well as $p^a(h^t)$ for $a \in \{N, H, L\}$ such that

$$p^a(h^0) = p^a \text{ and } p^a(h^{t+1}) \equiv \begin{cases} p^a(h^t) & \text{if } h^t \neq (h^t, a) \\ p^a(h^t) - (1-\delta)\delta^t & \text{if } h^t = (h^t, a). \end{cases} \tag{A.11}$$

Intuitively, $p^a(h^t)$ is the remaining occupation measure of outcome a at history h^t, while $p^a(h^0) - p^a(h^t)$ is the occupation measure of a from period 0 to $t-1$. Player 1's continuation value at h^t is

$$v(h^t) \equiv \delta^{-t} \sum_{a \in \{N,H,L\}} p^a(h^t) v^a. \tag{A.12}$$

Equilibrium Phases: The constructed equilibrium consists of three phases: an *active learning phase*, an *absorbing phase* and a *reshuffling phase*.

Play starts from the *active learning phase*, in which player 2 always plays T. Every type of player 1's mixed strategy at every history can be uniquely pinned down by player 2's belief updating process:

$$\Delta(h^t, L) = (1 - \lambda\gamma^*)\Delta(h^t) \quad \text{and} \quad \Delta(h^t, H) = \min\Big\{1 - \eta^*, \Big(1 + \lambda(1-\gamma^*)\Big)\Delta(h^t)\Big\}. \tag{A.13}$$

Since $\eta(h^0) > \eta^*$, we know that $\Delta(h^t) > 0$ for every h^t in the active learning phase.

Play transits to the *absorbing phase* permanently when $\Delta(h^t)$ reaches $1 - \eta^*$ for the first time. Recall that $v(h^t) \in \mathbb{R}^m$ is player 1's continuation value at h^t. Let $v_i(h^t)$ be the projection of $v(h^t)$ on the i-th dimension. After reaching the absorbing phase, player 2s' learning stops and the continuation outcome is either (T, H) in all subsequent periods or N in all subsequent periods, depending on the realization of a public randomization device, with the probability of (T, H) being $v_1(h^t)/(1 - \theta_1)$.

Play transits to the *reshuffling phase* at h^t if $\Delta(h^t) < 1 - \eta^*$ and $p^L(h^t) \in [0, (1-\delta)\delta^t)$.

1. If $p^L(h^t) = 0$, then the continuation play starting from h^t randomizes between N and (T, H), depending on the realization of the public randomization device, with the probability of (T, H) being $\frac{v_1(h^t)}{1-\theta_1}$.

2. If $p^L(h^t) \in (0, (1-\delta)\delta^t)$, then the continuation payoff vector can be written as a convex combination of

v^H, v^N and

$$(1-\delta)v^L + \widetilde{Q}v^H + (\delta - \widetilde{Q})v^N, \tag{A.14}$$

for some

$$\widetilde{Q} \in \Big[\min\{Y, \frac{\delta-\theta_1}{1-\theta_1}\}, \frac{\delta-\theta_1}{1-\theta_1}\Big]$$

and Y being defined in (A.7). I will show in the next subsection that $\widetilde{Q}$ indeed belongs to this range for every history reaching the reshuffling phase.

If player 1's realized continuation value at h^t takes the form in (A.14), then player 2 plays T at h^t, type θ_1 player 1 plays H for sure while other types mix between H and L with the same probabilities (could be degenerate) such that:

$$\Delta(h^t, L) = -\eta^* \text{ and } \Delta(h^t, H) = \begin{cases} \Delta(h^0) & \text{if } \Delta(h^t) \leq \Delta(h^0) \\ \Delta(h^t) & \text{if } \Delta(h^t) > \Delta(h^0). \end{cases} \tag{A.15}$$

If player 2 observes L at h^t, then he attaches probability 0 to type θ_1 and player 1's continuation value is

$$\delta^{-1}\widetilde{Q}v^H + \delta^{-1}(\delta - \widetilde{Q})v^N, \tag{A.16}$$

which can be delivered by randomizing between outcomes (T, H) and N, with probabilities $\delta^{-1}\widetilde{Q}$ and $1 - \delta^{-1}\widetilde{Q}$, respectively.

If player 2 observes H at h^t, then he attaches probability $\Delta(h^t, H) + \eta^*$ to type θ_1 and player 1's continuation value is:

$$\frac{1-\delta}{\delta}v^L + \frac{\widetilde{Q} - (1-\delta)}{\delta}v^H + \frac{\delta - \widetilde{Q}}{\delta}v^N, \tag{A.17}$$

which can be written as a convex combination of v^N and

$$v\Big(1 - \frac{1-\delta}{\widetilde{Q}}\Big). \tag{A.18}$$

According to (A.8) and the range of $\widetilde{Q}$,

$$\gamma < 1 - \frac{1-\delta}{\widetilde{Q}} < 1, \tag{A.19}$$

which implies that (A.18) can further be written as a convex combination of v^H and $v(\gamma)$.

If the continuation value is v^H or v^N, then the on-path outcome is (T, H) in every subsequent period or is N in every subsequent period. If the continuation value is $v(\gamma)$, then play switches back to the active learning phase with belief $\max\{\Delta(h^0), \Delta(h^t)\}$, which is no less than $\Delta(h^0)$.

A.3 Verifying Constraints

In this subsection, I verify that the strategy profile and the belief system indeed constitute a sequential equilibrium by verifying players' sequential rationality constraints and the consistency of beliefs. This consists of two parts. In Part I, I verify player 2's incentive constraints. In Part II, I verify the range of $\widetilde{Q}$ in Subsection A.2. In particular, at every history of the active learning phase or reshuffling phase, the ratio between the occupation measure of H and the occupation measure of L must exceed some cutoff.

Part I: Player 2's incentive constraints consist of two parts: the active learning phase and the reshuffling phase. If play remains in the active learning phase at h^t, then (A.13) implies that the unconditional probability with which H being played is at least γ^*, implying that player 2 has an incentive to play T. If play reaches the reshuffling phase at h^t and at this history, player 1 is playing a non-trivial mixed action, then according to (A.15) and the requirement that $\eta^* > \gamma^* \eta(h^0)$, the unconditional probability with which H is played is at least γ^*. This verifies player 2's incentives to play T.

Part II: In this part, I establish bounds on player 1's continuation value at every history in the active learning phase or in the beginning of the reshuffling phase. In particular, I establish a lower bound on the ratio between the convex weight of H and the convex weight of L at such histories, or equivalently, a lower bound on the depleted occupation measure of H and the depleted occupation measure of L. Recall the definitions of n and k in Subsection A.1. The conclusion is summarized in the following Lemma:

Lemma A.1. *If $\delta > \overline{\delta}$ and $T \geq 2k + X$, then for every $h^T = (a_0, ..., a_{T-1})$, if play remains in the active learning phase for every $h^t \preceq h^T$, then*

$$\underbrace{(1-\delta)\sum_{t=0}^{T-1}\delta^t \mathbf{1}\{a_t = H\}}_{\textit{depleted occupation measure of H}} - \underbrace{(1-\delta^X)}_{\textit{weight of initial } X \textit{ periods}} \leq \underbrace{(1-\delta)\sum_{t=0}^{T-1}\delta^t \mathbf{1}\{a_t = L\}}_{\textit{depleted occupation measure of L}} \cdot \underbrace{\frac{\widetilde{\gamma}}{1-\widetilde{\gamma}}}_{\textit{multiplier}}. \tag{A.20}$$

Lemma A.1 implies that when play first reaches the reshuffling phase, the remaining occupation measure of H is at least $\widetilde{Q}$. This implies that player 1's continuation value after reshuffling also attaches sufficiently high convex weight on v^H compared to the convex weight of v^L. Adapting the self-generation arguments in Abreu, Pearce and Stacchetti (1990) to an environment with persistent private information, one can conclude that payoff vector $v(\gamma)$ is attainable in sequential equilibrium when $\delta > \overline{\delta}$.

The challenge to prove this lemma is that we are seeking to establish a bound on the *discounted number of periods* with which the long-run player plays L relative to H while the constraints leading to this bound is expressed in terms of the *absolute number of periods*. In particular, notice that on one hand, player 2's belief, measured by $\eta(h^t)$, depends on the number of periods with which H and L are being played. On the other hand, player 1's continuation value, summarized by $p^H(h^t)$ and $p^L(h^t)$, depend on the discounted number of the periods with which H and L are being played. The translations between the above two constraints are difficult, even when δ is arbitrarily close to 1. This is because the occupation measure of the active learning phase is strictly positive even in the $\delta \to 1$ limit. Therefore, the discounted value of a period in the beginning of the active learning phase and that by the end is significantly different.

Proof of Lemma A.1: For every $t \in \mathbb{N}$, let $N_{L,t}$ and $N_{H,t}$ be the number of periods in which L and H are played from period 0 to $t-1$, respectively. The proof is done by induction on $N_{L,t}$.

When $N_{L,t} \leq 2(k-n)$, then the conclusion holds as $N_{H,t} \geq 2n + X$. According to (A.5) and (A.6), we know that $\Delta(h^T)$ will reach $1-\eta^*$ before period T (or equivalently, play reaches the absorbing phase).

Suppose the conclusion holds for when $N_{L,t} \leq N$ with $N \geq 2(k-n)$, and suppose towards a contradiction that there exists h^T with $T \geq k + X$ and $N_{L,T} = N+1$, such that play remains in the active learning phase for every $h^t \preceq h^T$ but

$$(1-\delta)\sum_{t=0}^{T-1}\delta^t \mathbf{1}\{a_t = H\} - (1-\delta^X) > (1-\delta)\sum_{t=0}^{T-1}\delta^t \mathbf{1}\{a_t = L\} \cdot \frac{\widetilde{\gamma}}{1-\widetilde{\gamma}}, \tag{A.21}$$

I will obtain a contradiction in three steps.

Step 1: I show that for every $s < T$,

$$(1-\delta)\sum_{t=s}^{T-1}\delta^t \mathbf{1}\{a_t = H\} \geq (1-\delta)\sum_{t=s}^{T-1}\delta^t \mathbf{1}\{a_t = L\}\frac{\widetilde{\gamma}}{1-\widetilde{\gamma}}. \tag{A.22}$$

Suppose towards a contradiction that the opposite of (A.22) holds, then (A.22) and (A.21) together imply that:

$$(1-\delta)\sum_{t=0}^{s-1}\delta^t \mathbf{1}\{a_t = H\} - (1-\delta^X) > (1-\delta)\sum_{t=0}^{s-1}\delta^t \mathbf{1}\{a_t = L\}\frac{\widetilde{\gamma}}{1-\widetilde{\gamma}} \tag{A.23}$$

and

$$(1-\delta)\sum_{t=s}^{T-1}\delta^t \mathbf{1}\{a_t = L\} > 0. \tag{A.24}$$

According to (A.24), $N_{L,s} < N_{L,T}$. Since $N_{L,T} = N + 1$, we have $N_{L,s} \leq N$. Applying the induction hypothesis and (A.23), we know that play reaches the absorbing phase before h^s, leading to a contradiction.

Step 2: I show that for every k consecutive periods

$$\{a_r, ..., a_{r+k-1}\} \subset h^T,$$

the number of H in this sequence is at least $n+1$. According to (A.22) shown in the previous step and (A.4), H occurs at least $n+1$ times in the last k periods, i.e. $\{a_{T-k+1}, ..., a_T\}$.

Suppose towards a contradiction that there exists k consecutive periods in which H occurs no more than n times, then the conclusion above that H occurs at least $n+1$ times in the last k periods implies that there exists k consecutive periods $\{a_r, ..., a_{r+k-1}\}$ in which H occurs exactly n times and L occurs exactly $k-n$ times. According to (A.4), we have

$$(1-\delta)\sum_{t=r}^{r+k-1}\delta^t \mathbf{1}\{a_t = H\} < (1-\delta)\sum_{t=r}^{r+k-1}\delta^t \mathbf{1}\{a_t = L\}\frac{\widetilde{\gamma}}{1-\widetilde{\gamma}}, \tag{A.25}$$

but according to (A.5) and the definition of $\widehat{\gamma}$ in (A.3), we also know that

$$\Delta(h^{r+k}) > \Delta(h^{r+1}). \tag{A.26}$$

Next, let us consider the following new sequence with length $T-k$:

$$\widetilde{h}^{T-k} \equiv \{\widetilde{a}_0, \widetilde{a}_1, ..., \widetilde{a}_{T-k-1}\} \equiv \{a_0, a_1, ..., a_{r-1}, a_{r+k}, ..., a_{T-1}\}$$

which is obtained by removing $\{a_r, ..., a_{r+k-1}\}$ from the original sequence and front-loading the subsequent play $\{a_{r+k}, ..., a_{T-1}\}$. The number of L in this new sequence is at most $N+1-(n-k)$, which is no more than N. According to the conclusion in Step 1:

$$(1-\delta)\sum_{t=r+k}^{T-1}\delta^t \mathbf{1}\{a_t = H\} > (1-\delta)\sum_{t=r+k}^{T-1}\delta^t \mathbf{1}\{a_t = L\}\frac{\widetilde{\gamma}}{1-\widetilde{\gamma}}. \tag{A.27}$$

This together with (A.25) and (A.21) imply that

$$(1-\delta)\sum_{t=0}^{T-k-1}\delta^t \mathbf{1}\{\widetilde{a}_t = H\} - (1-\delta^X) > (1-\delta)\sum_{t=0}^{T-k-1}\delta^t \mathbf{1}\{\widetilde{a}_t = L\}\frac{\widetilde{\gamma}}{1-\widetilde{\gamma}}.$$

According to the induction hypothesis, play will reach the absorbing phase before period $T-k$ if player 1 plays according to $\{\widetilde{a}_0, \widetilde{a}_1, ..., \widetilde{a}_{T-k-1}\}$.

1. Suppose $\widetilde{h}^{T-k}$ reaches the absorbing phase before period r, then play will also reach the absorbing phase before period r according to the original sequence.

2. Suppose $\widetilde{h}^{T-k}$ reaches the absorbing phase in period s, with $s > t$, then according to (A.26), we have $\Delta(\widetilde{h}^s) \leq \Delta(h^{s+k})$, implying that play will reach the absorbing phase in period $s+k$ according to the original sequence.

This contradicts the hypothesis that play has never reached the absorbing phase before period T if play proceeds according to h^T.

Step 3: For every history $h^T \equiv \{a_0, a_1, ..., a_{T-1}\} \in \{H, L\}^T$ and $t \in \{1, ..., T-1\}$, define the operator $\Omega_t : \{H, L\}^T \rightarrow \{H, L\}^T$ as:

$$\Omega_t(h^T) = (a_0, ..., a_{t-2}, a_t, a_{t-1}, a_{t+1}, ..., a_{T-1}), \tag{A.28}$$

in another word, swapping the order between a_{t-1} and a_t. Recall the belief updating formula in (A.13) and let

$$\mathcal{H}^{T,*} \equiv \left\{h^T \middle| \Delta(h^t) < 1 - \eta^* \text{ for all } h^t \prec h^T\right\}. \tag{A.29}$$

If $h^T \in \mathcal{H}^{T,*}$, then $\Omega_t(h^T) \in \mathcal{H}^{T,*}$ unless:

- $a_{t-1} = L, a_t = H.$
- *and*, $\Big(1 + \lambda(1 - \gamma^*)\Big)\Delta(h^{t-1}) \geq 1 - \eta^*.$

Next, I show that the above situation cannot occur besides in the last k periods. Suppose towards a contradiction that there exists $t \leq T-k$ such that $h^T \in \mathcal{H}^{T,*}$ but $\Omega_t(h^T) \notin \mathcal{H}^{T,*}$. Then according to the conclusion in step 2, H occurs at least $n+1$ times in $\{a_t, ..., a_{t+k-1}\}$. Now, consider the sequence $\{a_{t-1}, ..., a_{t+k-1}\}$, in which H

occurs at least $n+1$ times and L occurs at most $k-n$ times. This implies that:

$$\begin{aligned}
\Delta(h^{t+k}) &\geq \Delta(h^{t-1})\Big(1+\lambda(1-\gamma^*)\Big)^{n+1}\Big(1-\lambda\gamma^*\Big)^{k-n} \\
&= \Delta(h^{t-1})\underbrace{\Big(1+\lambda(1-\gamma^*)\Big)^{n}\Big(1-\lambda\gamma^*\Big)^{k-n}}_{\geq 1}\Big(1+\lambda(1-\gamma^*)\Big) \\
&\geq \Delta(h^{t-1})\Big(1+\lambda(1-\gamma^*)\Big) \\
&\geq 1-\eta^*, \qquad\qquad (A.30)
\end{aligned}$$

where 2nd inequality follows from $n/k > \widehat{\gamma}$ and (A.5), and the 3rd inequality follows from the hypothesis that $\Omega_t(h^T) \notin \mathcal{H}^{T,*}$. Inequality (A.30) implies that play reaches the high phase before period $t+k \leq T$, contradicting the hypothesis that $h^T \in \mathcal{H}^{T,*}$.

To summarize, for every $t \leq T-k$, if $h^T \in \mathcal{H}^{T,*}$, then $\Omega_t(h^T) \in \mathcal{H}^{T,*}$. For every $t > T-k$, if $h^T \in \mathcal{H}^{T,*}$, then $\Omega_t(h^T) \in \mathcal{H}^{T,*}$ unless $a_{t-1} = L$ and $a_t = H$. Therefore, one can freely front-load the play of H from period 0 to $T-k-1$ and obtain the following revised sequence:

$$\{H, H, .., H, L, L, ..., L, a_{T-k}, ..., a_{T-1}\}, \qquad (A.31)$$

which meets the following two requirements: (1) the revised sequence (A.31) still belongs to set $\mathcal{H}^{T,*}$; (2) sequence (A.31) satisfies (A.21).

According to the conclusion in Step 2: (1) the number of L from period 0 to $T-k-1$ cannot exceed $k-n-1$; (2) the number of L from period $T-k$ to $T-1$ cannot exceed $k-n-1$. This is because otherwise, there exists a sequence of length k that has at most n periods of H, contradicting the two conditions the revised sequence in (A.31) satisfies. Therefore, the total number of L in this sequence is at most $2(k-n-1)$, which contradicts the induction hypothesis that the number of L exceeds $2(k-n)$.

Summary: Steps 1-3 together imply that there exists no sequence with length greater than $X+2k$ such that play remains in the active learning phase and satisfies inequality (A.21). This leads to the conclusion in Lemma A.1 which concludes the proof of $V^* \subset \underline{V}(\pi_0)$. □

B Proof of Theorem 1: Necessity

In this Appendix, I show that $\overline{V}(\pi_0) \subset V^*$. In subsection B.1, I establish a payoff upper bound for the lowest cost type that uniformly applies across all discount factors. In subsection B.2, I establish a payoff upper bound for other types that applies in the $\delta \to 1$ limit. To accommodate applications where players move simultaneously, I prove the result under the following simultaneous-move stage game:[22]

$\theta = \theta_i$	T	N
H	$1-\theta_i, b$	$-d(\theta_i), 0$
L	$1, -c$	$0, 0$

I assume that players' payoffs are monotone-supermodular (Liu and Pei 2017). In the context of this game, once the states and players' actions are ranked according to $\theta_1 \succ \theta_2 \succ ... \succ \theta_m$, $H \succ L$ and $T \succ N$, monotone-supermodularity implies that $d(\theta_i) \geq 0$ for every $\theta_i \in \Theta$ and $|\theta_i - \theta_j| \geq |d(\theta_i) - d(\theta_j)|$ for every $i < j$. The proof consists of two parts that establish the necessity of the two linear constraints characterizing V^*, respectively.

B.1 Necessity of Constraint One: Payoff Upper Bound for Type θ_1

I start with recursively defining the set of *high histories*. Let $\overline{\mathcal{H}}^0 \equiv \{h^0\}$ and

$$\overline{a}_1(h^0) \equiv \max\Big\{ \bigcup_{\theta \in \Theta} \text{supp}\Big(\sigma_\theta(h^0)\Big)\Big\}.$$

Let

$$\overline{\mathcal{H}}^1 \equiv \{h^1 | \exists h^0 \in \overline{\mathcal{H}}^0 \text{ s.t. } h^1 \succ h^0 \text{ and } \overline{a}_1(h^0) \in h^1\}.$$

For every $t \in \mathbb{N}$ and $h^t \in \overline{\mathcal{H}}^t$, let $\Theta(h^t) \subset \Theta$ be the set of types that occur with positive probability at h^t. Let

$$\overline{a}_1(h^t) \equiv \max\Big\{ \bigcup_{\theta \in \Theta(h^t)} \text{supp}\Big(\sigma_\theta(h^t)\Big)\Big\} \tag{B.1}$$

and

$$\overline{\mathcal{H}}^{t+1} \equiv \{h^{t+1} | \exists h^t \in \overline{\mathcal{H}}^t \text{ s.t. } h^{t+1} \succ h^t \text{ and } \overline{a}_1(h^t) \in h^{t+1}\}. \tag{B.2}$$

Let $\overline{\mathcal{H}} \equiv \bigcup_{t=0}^{\infty} \overline{\mathcal{H}}^t$ be the set of high histories. The main result in this subsection is the following Proposition, which shows that at every history, the lowest cost type in the support of player 2s' posterior belief cannot receive a continuation payoff higher than her pure Stackelberg commitment payoff.

Proposition B.1. *For every $h^t \in \overline{\mathcal{H}}$, if $\theta_i = \min \Theta(h^t)$, then type θ_i's continuation payoff at h^t is no more than $1 - \theta_i$ in any Nash equilibrium.*

[22] Having players move simultaneously in the stage game and letting future short-run players observing their predecessors' actions introduce new challenges. As discussed in Pei (2017), the predecessor-successor relationship is incomplete on the set of histories where player 1 has always played H. Nevertheless, the adaptation of my proof to sequential-move stage games in the baseline model is straightforward.

Since $h^0 \in \overline{\mathcal{H}}$ and $\theta_1 = \min \Theta(h^0)$, a corollary of Proposition B.1 is that type θ_1's payoff cannot exceed $1 - \theta_1$ in any Nash equilibrium.

Proof of Proposition B.1: For every $\theta \in \Theta$, let $\overline{\mathcal{H}}(\theta)$ be a subset of $\overline{\mathcal{H}}$ (could be empty) such that $h^t \in \overline{\mathcal{H}}(\theta)$ if and only if both of the following conditions are satisfied:

1. For every $h^s \succeq h^t$ with $h^s \in \overline{\mathcal{H}}$, we have $\theta \in \Theta(h^s)$.

2. If $h^{t-1} \prec h^t$, then for every $\tilde{\theta} \in \Theta(h^{t-1})$, there exists $h^s \in \overline{\mathcal{H}}$ with $h^s \succ h^{t-1}$ such that $\tilde{\theta} \notin \Theta(h^s)$.

Let $\overline{\mathcal{H}}(\Theta) \equiv \bigcup_{\theta \in \Theta} \overline{\mathcal{H}}(\theta)$. By definition, $\overline{\mathcal{H}}(\Theta)$ possesses the following two properties:

1. $\overline{\mathcal{H}}(\Theta) \subset \overline{\mathcal{H}}$.

2. For every $h^t, h^s \in \overline{\mathcal{H}}(\Theta)$, neither $h^t \succ h^s$ nor $h^t \prec h^s$.

For every $h^t \in \overline{\mathcal{H}}(\theta_i)$, at the subgame starting from h^t, type θ_i's stage game payoff is no more than $1-\theta_i$ in every period if she plays $\overline{a}_1(h^s)$ for every $h^s \succsim h^t$ and $h^s \in \overline{\mathcal{H}}$. Since $h^t \in \overline{\mathcal{H}}(\theta_i)$ implies that doing so is optimal for type θ_i, her continuation payoff at h^t cannot exceed $1 - \theta_i$. When the stage game payoff is supermodular, for every $j < i$, the payoff difference between type θ_j and type θ_i in any period is at most $|\theta_i - \theta_j|$. This implies that for every $\theta_j \in \Theta(h^t)$ with $\theta_j < \theta_i$, type θ_j's continuation payoff at h^t cannot exceed $1 - \theta_j$.

In what follows, I show Proposition B.1 by induction on $|\Theta(h^t)|$. When $|\Theta(h^t)| = 1$, i.e. there is only one type (call it type θ_i) that can reach h^t. The above argument implies that type θ_i's payoff cannot exceed $1 - \theta_i$.

Suppose the conclusion in Proposition B.1 holds for every $|\Theta(h^t)| \leq n$, consider the case when $|\Theta(h^t)| = n + 1$. Let $\theta_i \equiv \min \Theta(h^t)$. Next, I introduce the definition of set $\overline{\mathcal{H}}^B(h^t)$: For every $h^s \succeq h^t$ with $h^s \in \overline{\mathcal{H}}$, $h^s \in \overline{\mathcal{H}}^B(h^t)$ if and only if:

- $\theta_i \in \Theta(h^s)$ but $\theta_i \notin \Theta(h^{s+1})$ for every $h^{s+1} \succ h^s$ and $h^{s+1} \in \overline{\mathcal{H}}$.

In another word, type θ_i has a strict incentive not to play $\overline{a}_1(h^s)$ at h^s. A useful property of $\overline{\mathcal{H}}^B(h^t)$ is:

- For every $h^\infty \in \overline{\mathcal{H}}$ with $h^\infty \succ h^t$, either there exists $h^s \in \overline{\mathcal{H}}^B(h^t)$ such that $h^s \prec h^\infty$, or there exists $h^s \in \overline{\mathcal{H}}(\theta_i)$ such that $h^s \prec h^\infty$.

which means that play will eventually reach either a history in $\overline{\mathcal{H}}^B(h^t) \bigcup \overline{\mathcal{H}}(\theta_i)$ if type θ plays $\overline{a}_1(h^\tau)$ before that for every $t \leq \tau \leq s$. In what follows, I examine type θ_i's continuation value.

1. For every $h^s \in \overline{\mathcal{H}}^B(h^t)$, at every h^{s+1} satisfying $h^{s+1} \succ h^s$ and $h^{s+1} \in \overline{\mathcal{H}}$, we have:

$$|\Theta(h^{s+1})| \leq n.$$

Let $\theta_j \equiv \min \Theta(h^{s+1})$. According to the induction hypothesis, type θ_j's continuation payoff at h^{s+1} is at most $1-\theta_j$. Since this applies to every such h^{s+1}, type θ_j's continuation value at h^s also cannot exceed $1-\theta_j$ since she is playing $\overline{a}_1(h^s)$ with positive probability at h^s, and her stage game payoff from doing so is at most $1-\theta_j$. Therefore, type θ_i's continuation value at h^s is at most $1-\theta_i$.

2. For every $h^s \in \overline{\mathcal{H}}(\theta_i)$, playing $\overline{a}_1(h^\tau)$ for all $h^\tau \succeq h^s$ and $h^\tau \in \overline{\mathcal{H}}$ is a best reply for type θ_i. Her stage game payoff from this strategy cannot exceed $1-\theta_i$, which implies that her continuation value at h^s also cannot exceed $1-\theta_i$.

Starting from h^t, consider the strategy in which player 1 plays $\overline{a}_1(h^\tau)$ at every $h^\tau \succ h^t$ and $h^\tau \in \overline{\mathcal{H}}$ until play reaches $h^s \in \overline{\mathcal{H}}^B(h^t)$ or $h^s \in \overline{\mathcal{H}}(\theta_i)$. By construction, this is type θ_i's best reply. Under this strategy, type θ_i's stage game payoff cannot exceed $1-\theta_i$ before reaches h^s. Moreover, her continuation payoff after reaching h^s is also bounded above by $1-\theta_i$, which establishes the conclusion of Proposition B.1 when $|\Theta(h^t)| = n+1$. $\square$

B.2 Necessity of Constraint Two: Maximal Relative Frequency Between L and H

Suppose towards a contradiction that there exists $v = (v_1, ..., v_m) \in \overline{V}(\pi_0)$ and $j \in \{1, 2, ..., m\}$ such that $v_j > v_j(\gamma^*)$. Then given the constraint established in the first part that $v_1 \leq 1-\theta_1$, we know that $j > 1$. Under the probability measure over $\mathcal{H}$ induced by $(\sigma_{\theta_j}, \sigma_2)$, let $X^{(\sigma_{\theta_j}, \sigma_2)}$ be the occupation measure of outcome (T, H) and let $Y^{(\sigma_{\theta_j}, \sigma_2)}$ be the occupation measure of outcome (T, L). As $v_j > v_j(\gamma^*)$, we have:

$$\frac{X^{(\sigma_{\theta_j}, \sigma_2)}}{Y^{(\sigma_{\theta_j}, \sigma_2)}} < \frac{\gamma^*}{1-\gamma^*}. \tag{B.3}$$

Let the value of the left-hand-side be $\frac{\gamma}{1-\gamma}$ for some $\gamma \in [0, \gamma^*)$.

For every $h^\tau \in \mathcal{H}$, let $\sigma_{\theta_j}(h^\tau) \in \Delta(A_1)$ be the (mixed) action prescribed by σ_{θ_j} at h^τ and let $\alpha_1(\cdot|h^\tau)$ be player 2's expected action of player 1's at h^τ. Let $d(\cdot\|\cdot)$ be the Kullback-Leibler divergence between two action distributions. Suppose player 1 plays according to σ_{θ_j}, the result in Gossner (2011) implies that:

$$\mathbb{E}^{(\sigma_{\theta_j}, \sigma_2)}\Big[\sum_{\tau=0}^{+\infty} d(\sigma_{\theta_j}(h^\tau)||\alpha_1(\cdot|h^\tau))\Big] \leq -\log \pi_0(\theta_j). \tag{B.4}$$

This implies that the expected number of periods such that $d(\sigma_{\theta_j}(h^\tau)||\alpha_1(\cdot|h^\tau)) > \epsilon$ is no more than

$$T(\epsilon) \equiv \Big\lceil \frac{-\log \pi_0(\theta_j)}{\epsilon} \Big\rceil. \tag{B.5}$$

Let

$$\epsilon \equiv d\Big(\frac{\gamma+2\gamma^*}{3}H+(1-\frac{\gamma+2\gamma^*}{3})L\Big\|\gamma^*H+(1-\gamma^*)L\Big), \tag{B.6}$$

and let δ be large enough such that:

$$\frac{X^{(\sigma_{\theta_j},\sigma_2)}}{Y^{(\sigma_{\theta_j},\sigma_2)}-(1-\delta^{T(\epsilon)})}<\frac{2\gamma+\gamma^*}{3-2\gamma-\gamma^*}. \tag{B.7}$$

According to (B.4) and (B.5), if type θ_j plays according to her equilibrium strategy, then there are at most $T(\epsilon)$ periods in which player 2's expectation over player 1's action differs from σ_{θ_j} by more than ϵ. According to (B.6), aside from $T(\epsilon)$ periods, player 2 will trust player 1 at h^t only when $\sigma_{\theta_j}(h^t)$ assigns probability at least $\frac{\gamma+2\gamma^*}{3}$ to H. Therefore, under the probability measure induced by $(\sigma_{\theta_j},\sigma_2)$, the occupation measure with which player 2 trusts player 1 is at most:

$$\underbrace{(1-\delta^{T(\epsilon)})}_{\text{periods s.t. player 2's prediction is wrong}}+\underbrace{\Big(X^{(\sigma_{\theta_j},\sigma_2)}+Y^{(\sigma_{\theta_j},\sigma_2)}-(1-\delta^{T(\epsilon)})\Big)\frac{2\gamma+\gamma^*}{\gamma+2\gamma^*}}_{\text{maximal frequency with which player 2 trusts after he learns}}, \tag{B.8}$$

which is strictly less than $X^{(\sigma_{\theta_j},\sigma_2)}+Y^{(\sigma_{\theta_j},\sigma_2)}$ when δ is close enough to 1, leading to a contradiction.

C Proof of Theorem 2

Notation: For every $\sigma_\theta:\mathcal{H}\rightarrow\Delta(A_1)$ and $\sigma_2:\mathcal{H}\rightarrow\Delta(A_2)$, let $\mathcal{H}(\sigma_\theta,\sigma_2)$ be the set of histories that occur with positive probability under the measure induced by (σ_θ,σ_2). Let $\overline{\sigma}_\theta:\mathcal{H}\rightarrow\Delta(A_1)$ be such that $\overline{\sigma}_\theta(h^t)=H$ for every $h^t\in\mathcal{H}$. Let $\underline{\sigma}_\theta:\mathcal{H}\rightarrow\Delta(A_1)$ be such that $\underline{\sigma}_\theta(h^t)=L$ for every $h^t\in\mathcal{H}$.

Completely Mixed Strategies: Suppose towards a contradiction that there exists a Nash equilibrium $\sigma=\Big((\sigma_\theta)_{\theta\in\Theta},\sigma_2\Big)$ that attains payoff within ϵ of v^* but some type $\widehat{\theta}$ is playing a completely mixed strategy. Then both $\overline{\sigma}_\theta$ and $\underline{\sigma}_\theta$ are type $\widehat{\theta}$'s best reply to σ_2. Since the stage game payoff is monotone-supermodular according to the orders $T\succ N$, $H\succ L$ and $\theta_1\succ\theta_2\succ...\succ\theta_m$ (Liu and Pei 2017), Lemma C.1 in Pei (2017) implies that:

1. For every $\theta_j\succ\widehat{\theta}$, type θ_j will play H with probability 1 at every $h^t\in\mathcal{H}(\overline{\sigma}_\theta,\sigma_2)$.
2. For every $\theta_k\prec\widehat{\theta}$, type θ_k will play L with probability 1 at every $h^t\in\mathcal{H}(\underline{\sigma}_\theta,\sigma_2)$.

Consider two cases. First, suppose $\widehat{\theta}\neq\theta_m$, then type θ_m will play L with probability 1 at every $h^t\in\mathcal{H}(\underline{\sigma}_\theta,\sigma_2)$, from which she is supposed to receive payoff no less than $v_m^*-\epsilon$. On the other hand, the argument in Fudenberg and Levine (1992) implies that in every Nash equilibrium, there are at most

$$T_{\theta_m} \equiv \log \pi_0(\theta_m) \Big/ \log(1-\gamma^*) \tag{C.1}$$

periods in which player 2 plays T. That is to say, for every $\epsilon > 0$, there exists $\overline{\delta} \in (0,1)$ such that when $\delta > \overline{\delta}$, type θ_m's payoff is less than ϵ in every Nash equilibrium. Pick ϵ to be small enough such that $\epsilon < v_m^*/2$, we obtain a contradiction.

Second, suppose $\widehat{\theta} = \theta_m$, then types $\theta_1 \sim \theta_{m-1}$ will play H with probability 1 at every $h^t \in \mathcal{H}(\overline{\sigma}_\theta, \sigma_2)$. Therefore, after playing L for the first time, type θ_m will reveal her type so her continuation payoff is at most $1-\theta_m$. Hence, her discounted average payoff in the repeated game cannot exceed $(1-\delta)+\delta(1-\theta_m)$. Let ϵ be small enough such that $(1-\delta)+\delta(1-\theta_m) < v_m^* - \epsilon$, we have a contradiction.

Stationary Strategies: The above argument rules out completely mixed strategies. To rule out stationary s-trategies, one only needs to show that no type will play stationary pure strategies. First, suppose towards a contradiction that type $\widehat{\theta}$ plays L in every period, then in every Nash equilibrium, there are at most

$$T_{\widehat{\theta}} \equiv \log \pi_0(\widehat{\theta}) \Big/ \log(1-\gamma^*) \tag{C.2}$$

periods in which player 2 plays T. Therefore, her equilibrium payoff vanishes to 0 as δ approaches 1, contradict-ing the fact that $v_{\widehat{\theta}} \geq 1 - \widehat{\theta} > 0$.

Second, suppose towards a contradiction that type $\widehat{\theta}$ plays H in every period. If $\widehat{\theta} \neq \theta_1$, then her equilibrium payoff is at most $1-\widehat{\theta}$, which is strictly less than $v_{\widehat{\theta}}^*$, leading to a contradiction. If $\widehat{\theta} = \theta_1$, then type θ_2 will be separated from type θ_1 the first time she plays L, after which her continuation payoff is no more than $1-\theta_2$. Therefore, type θ_2's equilibrium payoff is at most $(1-\delta)+\delta(1-\theta_2)$, which is strictly less than v_2^* when δ is close enough to 1, leading to a contradiction.

D Miscellaneous

D.1 Counterexample under Complete Information

I provide a counterexample showing that when $|\Theta| = 1$, there exists a sequential equilibrium in which player 1 plays a stationary mixed strategy and attains payoff v^*, equal to $1-\theta$ according to (3.1).

The long-run player plays H with probability γ^* at every history. The short-run player plays T in period 0.

His action in period $t(\geq 1)$ depends on the game's outcome in period $t-1$ (denoted by a_{t-1}):

$$\text{Player 2's action in period } t = \begin{cases} N & \text{if } a_{t-1} = N \\ T & \text{if } a_{t-1} = H \\ p^*T + (1-p^*)N & \text{if } a_{t-1} = L \end{cases} \tag{D.1}$$

where

$$p^* \equiv \frac{1-\theta/\delta}{1-\theta}, \tag{D.2}$$

which is strictly between 0 and 1 when δ is close enough to 1.

D.2 ϵ-Stackelberg Strategies

First, I construct a Nash equilibrium in which some type of the long-run player plays a stationary ϵ-Stackelberg strategy but the long-run player's equilibrium payoff is 0. Consider the following strategy profile:

– The short-run player plays N at every history.

– Type θ_1 plays H with probability $\gamma^* + \epsilon$ at every history.

– Types other than θ_1 plays L at every history.

Let ϵ be small enough such that:

$$\epsilon < \frac{c}{b+c}\frac{1-\pi_0(\theta_1)}{\pi_0(\theta_1)}, \tag{D.3}$$

the above strategy profile is a Nash equilibrium for every $\delta \in (0,1)$.

Next, I show that no type will play any stationary ϵ-Stackelberg strategy in any sequential equilibrium when ϵ is small enough so that every stationary ϵ-Stackelberg strategy is completely mixed.

Proposition D.1. *For every ϵ small enough, there exists no sequential equilibrium in which some type of the long-run player plays a stationary ϵ-Stackelberg strategy.*

Proof. Let ϵ be small enough so that every stationary ϵ-Stackelberg strategy is completely mixed. Suppose towards a contradiction that there exists a sequential equilibrium (σ, π) with $\sigma = \Big((\sigma_\theta)_{\theta\in\Theta}, \sigma_2\Big)$ and $\pi : \mathcal{H} \to \Delta(\Theta)$ such that $\sigma_{\widehat{\theta}}$ is a stationary ϵ-Stackelberg strategy. Consider the subgame after player 2 plays T in period $t \in \mathbb{N}$. Both $\overline{\sigma}_\theta$ and $\underline{\sigma}_\theta$ are type $\widehat{\theta}$'s best reply to σ_2 in that subgame. Lemma C.1 in Pei (2017) implies that:

1. For every $\theta < \widehat{\theta}$, type θ will play H with probability 1 in period t.

2. For every $\theta > \widehat{\theta}$, type θ will play L with probability 1 in period t.

Therefore, after observing H in period 0, player 2's posterior attaches probability 1 to the event that $\theta \leq \widehat{\theta}$. For every $\theta \leq \widehat{\theta}$, we have shown before that she will play H with probability strictly greater than γ^* at every history where T is played with positive probability. Hence, T is the short-run player's strict best reply after he observes H in period 0 and regardless of player 1's action choices after period 0. As a result, in the subgame following the short-run player plays T in period 0, type $\widehat{\theta}$ can obtain continuation payoff $(1-\delta)(1-\widehat{\theta})+\delta$ by playing H in period 0 and playing L in all subsequent periods, which is strictly more than her payoff by playing H in every period, which is $1-\widehat{\theta}$. This contradicts the previous claim that $\overline{\sigma}_\theta$ is her best reply. □

References

[1] Abreu, Dilip, David Pearce and Ennio Stacchetti (1990) "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica*, 58(5), 1041-1063.

[2] Aumann, Robert and Michael Maschler (1995) *Repeated Games with Incomplete Information*, MIT Press.

[3] Benabou, Roland and Guy Laroque (1992) "Using Privileged Information to Manipulate Markets: Insiders, Gurus, and Credibility," *Quarterly Journal of Economics*, 107(3), 921-958.

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

[5] Chassang, Sylvain (2010) "Building Routines: Learning, Cooperation, and the Dynamics of Incomplete Relational Contracts," *American Economic Review*, 100(1), 448-465.

[6] Cole, Harold, James Dow and Willam English (1995) "Default, Settlement and Signalling: Lending Resumption in a Reputational Model of Sovereign Debt," *International Economic Review*, 36(2), 365-385.

[7] Cripps, Martin and Jonathan Thomas (2003) "Some Asymptotic Results in Discounted Repeated Games of One-Sided Incomplete Information," *Mathematics of Operations Research*, 28, 433-462.

[8] Dai, Weijia, Ginger Jin, Jungmin Lee and Michael Luca (2018) "Aggregation of Consumer Ratings: An Application to Yelp.com," *Quantitative Marketing and Economics*, forthcoming.

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

[10] Ekmekci, Mehmet (2011) "Sustainable Reputations with Rating Systems," *Journal of Economic Theory*, 146(2), 479-503.

[11] Ely, Jeffrey, Johannes Hörner and Wojciech Olszewski (2005) "Belief-Free Equilibria in Repeated Games," *Econometrica*, 73(2), 377-415.

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

[13] 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.

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

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

[16] Fudenberg, Drew and David Levine (1994) "Efficiency and Observability with Long-Run and Short-Run Players," *Journal of Economic Theory*, 62(1), 103-135.

[17] Ghosh, Parikshit and Debraj Ray (1996) "Cooperation and Community Interaction Without Information Flows," *Review of Economic Studies*, 63(3), 491-519.

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

[19] Hart, Sirgiu (1985) "Nonzero-Sum Two-Person Repeated Games with Incomplete Information," *Mathematics of Operations Research*, 10(1), 117-153.

[20] Horner, Johannes and Stefano Lovo (2009) "Belief-Free Equilibria in Games With Incomplete Information," *Econometrica*, 77(2), 453-487.

[21] Jehiel, Philippe and Larry Samuelson (2012) "Reputation with Analogical Reasoning," *Quarterly Journal of Economics*, 127(4), 1927-1969.

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

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

[24] Liu, Shuo and Harry Pei (2017) "Monotone Equilibria in Signalling Games," Working Paper, MIT and University of Zurich.

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

[26] Milgrom, Paul and John Roberts (1982) "Limit Pricing and Entry under Incomplete Information: An Equilibrium Analysis," *Econometrica*, 50(2), 443-459.

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

[28] Pei, Harry (2017) "Reputation Effects under Interdependent Values," Working Paper, MIT.

[29] Pęski, Marcin (2014) "Repeated Games with Incomplete Information and Discounting," *Theoretical Economics*, 9(3), 651-694.

[30] Schmidt, Klaus (1993) "Commitment through Incomplete Information in a Simple Repeated Bargaining Game," *Journal of Economic Theory*, 60(1), 114-139.

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

[32] Tirole, Jean (1996) "A Theory of Collective Reputations (with applications to the persistence of corruption and to firm quality)," *Review of Economic Studies*, 63(1), 1-22.

[33] Tirole, Jean (2006) "The Theory of Corporate Finance," Princeton University Press.

[34] Weinstein, Jonathan and Muhamet Yildiz (2007) "A Structure Theorem for Rationalizability with Application to Robust Predictions of Refinements," *Econometrica*, 75(2), 365-400.

[35] Weinstein, Jonathan and Muhamet Yildiz (2016) "Reputation without Commitment in Finitely Repeated Games," *Theoretical Economics*, 11(1), 157-185.