Trust and Betrayals
Reputational Payoffs and Behaviors without Commitment

Harry PEI*

January 22, 2019

Abstract: I introduce a reputation model where all types of the reputation building player 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 every type of the patient player's highest equilibrium payoff. Interestingly, incomplete information affects this payoff only through the lowest benefit in the support of the prior belief. In every equilibrium that attains this highest payoff, the patient player's behavior depends non-trivially on past play. I establish bounds on her long-run action frequencies that apply to all of her equilibrium best replies. These features on behavior are essential for her to extract information rent while preserving her informational advantage. I construct a class of high-payoff equilibria in which the patient player's reputation only depends on 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: rational reputational types, lack-of-commitment problem, equilibrium behavior, reputation
JEL Codes: C73, D82, D83

1 Introduction

Trust is essential in many socioeconomic activities, yet it is also susceptible to opportunism and exploitation. To fix ideas, consider the example of a supplier who promises his clients about on-time deliveries. After the client agrees to purchase and makes a relationship-specific investment, the supplier has an incentive to delay in order to save cost.[1] Similarly, 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 also occur when incumbents deter entrants, central banks fight hyperinflation, politicians seek support from their electorate and entrepreneurs raise funding for their projects.

The common theme in these applications is a *lack-of-commitment problem* faced by the suppliers, firms, politicians, central banks and entrepreneurs. As a response, these agents build reputations for being trustworthy,

*Department of Economics, Northwestern University. I thank Daron Acemoglu, Ricardo Alonso, Guillermo Caruana, Mehmet Ekmekci, Jeff Ely, Drew Fudenberg, Bob Gibbons, Aditya Kuvalekar, Elliot Lipnowski, Debraj Ray, Bruno Strulovici, Jean Tirole, Juuso Toikka, Tristan Tomala, Alex Wolitzky, Muhamet Yildiz and my seminar audiences for helpful comments. Errors are mine.

[1] As documented by Banerjee and Duflo (2000), non-contractible delays and cost overruns are important concerns in the Indian custom software industry, and are primary motivations for suppliers to build reputations. See Dellarocas (2006) for more examples.

from which they 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.

I introduce 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*. Each short-run player observes the outcomes of all past interactions and prefers to trust the long-run player if the probability of high effort is 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 high payoffs. This includes her optimal commitment payoff (or *Stackelberg payoff*) when the lowest possible cost vanishes. I derive properties of the long-run player's behavior that apply to *all* equilibria in which she attains her highest equilibrium payoff.[4] I establish bounds on her long-term action frequencies that apply not only to all of her equilibrium strategies, but also to all of her equilibrium best replies. As a result, one can test these predictions by observing a realized path of play instead of the entire distribution. I construct a class of those high-payoff equilibria in which the long-run player's reputation only depends on 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.[5]

My first result (Theorem 1) provides a full characterization of the patient long-run player's equilibrium payoff set. It unifies the incomplete information setting with the complete information benchmark in Fudenberg, Kreps and Maskin (1990). The highlight of this result is a tractable formula for every type's highest equilibrium payoff, which is the product of her Stackelberg payoff and an *incomplete information multiplier*. The latter is a sufficient statistic for the effect of incomplete information and only depends on the *lowest cost* in the support

[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 (such as Airbnb Uber, Lyft) where buyers are unlikely to meet with the same seller twice, etc. Relaxing this assumption does not affect the *attainability* of high payoffs. Nevertheless, it can expand the long-run player's equilibrium payoff set, see Pęski (2014).

[4]For every $\varepsilon > 0$, there exist equilibria such that every type's payoff is at least her highest equilibrium payoff minus ε. By "attain her highest equilibrium payoff", I mean *every type* of the long-run player approximately attains her highest equilibrium payoff.

[5]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.

of the short-run players' prior belief. Moreover, the multiplier coincides with the maximal probability attached to the Stackelberg outcome such that the lowest cost type's payoff does not exceed her highest payoff in the repeated complete information game. My formula implies that every type except the lowest cost one benefits from incomplete information, in the sense that her highest equilibrium payoff is strictly higher compared to the complete information benchmark. Furthermore, when the lowest cost in the support vanishes to zero, the multiplier converges to one and every type's highest equilibrium payoff converges to her Stackelberg payoff.

Next, I study properties of the patient player's behavior that apply to *all* equilibria in which she approximately attains her highest equilibrium payoff (or *high payoff equilibria*). As a first step, Theorem 2 shows that no type of the patient player uses stationary strategies nor has a completely mixed best reply. This conclusion extends to a type whose cost of high effort is zero. It implies that in every high payoff equilibrium, every type of the patient player faces non-trivial incentives and cherry-picks actions based on the history of play. It also contrasts to the stationary commitment strategies in the reputation literature that prescribe the same action at every history.

To further explore these non-stationary behaviors, Theorem 3 derives bounds on the patient player's long-term action frequencies that apply to all of her best replies (to her opponents' equilibrium strategies). In particular, for every type that does not have the highest cost and according to each of her pure strategy best reply, the relative frequency of high and low effort cannot fall below the ratio between their probabilities in the Stackelberg action (call it the *critical ratio*). Similarly, for every type that does not have the lowest cost, the relative frequency of high and low effort cannot exceed the critical ratio under each of her pure strategy best reply. The two bounds together pin down the action frequencies for all types except for the ones with the highest and the lowest cost. Since these bounds apply to *all equilibrium best replies* of the patient player, they are stronger than the ones that only apply to her equilibrium strategies. This distinction is economically important as the resulting predictions can be tested by observing a realized path of the long-run player's action paths, instead of the entire distribution.

Intuitively, the properties identified in Theorems 2 and 3 allow the high-cost types to shirk occasionally without losing too much reputation, while discourage them from shirking too frequently in order to provide the short-run players the incentives to trust. To achieve the first objective, one needs to motivate the rational reputational type to shirk so that the high cost types can conceal their private information while extracting rent. To achieve the second objective, one needs to take into account the short-run players' learning. As suggested by the canonical reputation results in Fudenberg and Levine (1989,1992) and Gossner (2011), the short-run players can predict the long-run player's future actions with high precision in all except for a bounded number of periods.

The proofs of Theorems 2 and 3 exploit the rationality of all types of the long-run player as well as the supermodularity of her stage-game payoffs. For a snapshot of the argument, suppose it is optimal for the lowest cost type to exert low effort at every history (which happens when mixing at every history is her best reply). Due to the high-cost types' comparative advantage in exerting low effort, they will shirk for sure at every on-path history. However, if they behave like this in equilibrium, then the short-run players believe that low effort will occur with high enough probability in all future periods after they observe low effort for a bounded number of periods. They will then stop trusting the long-run player, leaving those high-cost types a low payoff. In general, this logic leads to an upper bound on the frequencies with which the low-cost types exert low effort.

Conversely, suppose it is optimal for a high-cost type to exert high effort at a given frequency, then according to every best reply of the lowest cost type, her frequency of exerting high effort must be weakly higher. In order for a high-cost type to hide behind the reputational type and extract information rent, the long-run frequencies of her actions cannot be too different from the equilibrium action frequencies of the reputational type. Therefore, a lower bound on the high-cost type's equilibrium payoff leads to an upper bound on the frequency of high effort not only under her equilibrium strategies but also under each of her pure strategy best reply.

My proof of Theorem 1 offers a more concrete illustration of players' behaviors by constructing a class of such high payoff equilibria. The main conceptual challenge arises from the observation that extracting information rent (i.e. shirk while winning her opponent's trust) inevitably reveals information about the long-run player's type, which undermines her informational advantage and her ability to extract information rent in the future. This tension grows as the long-run player becomes more patient, as she needs to extract information rent in unbounded number of periods to obtain a discounted average payoff strictly above her highest complete information payoff.

My main technical contribution is to overcome the above challenge while taking the reputational type's incentives into account. The class of equilibria I construct exhibit *slow learning* and *reputation cycles* that *end in finite time*. Play starts from an *active learning phase*, in which the short-run players trust and the patient player's reputation improves after high effort and deteriorates after low effort. Every high cost type plays a non-trivially mixed action unless her reputation (i.e., probability of being the lowest cost type) is sufficiently close to one, at which point she shirks for one period and extracts information rent. This reduces the magnitude of her reputation loss in the process of extracting rent. Moreover, she can rebuild her reputation after milking it, which leads to reputation cycles and allows her to extract information rent in the long run. In order to provide the right incentives for all types of the patient player, play transits to an *absorbing phase* either when reputation becomes perfect, or when the patient player has extracted enough information rent by shirking too much in the past. One of these events must happen in finite time and the long-run player's payoff in the absorbing phase depends on her realized action choices in the active learning phase.

Related Literature: This paper relates to the literatures on reputations and repeated incomplete information games. The conceptual contributions lie both in the approach taken and in the research questions.

The canonical approach to study reputations, which has been adopted by Fudenberg and Levine (1989,1992), Gossner (2011) and many others, *fixes* the behavior of at least one type of the long-run player (or *commitment* *type*). In presence of this commitment type, the strategic long-run player can attain her commitment payoff when her discount factor approaches 1. This includes payoffs that are not attainable in the repeated complete information game (or *attainability*). When the long-run player's action choices are identifiable, which is the case in the simultaneous-move trust game but not in the sequential-move game, she can guarantee her commitment payoff in all equilibria (or *refinement*).

I take a complementary approach in which all types of the long-run player are rational, have reasonable payoff functions and can flexibly choose their strategies to maximize their payoffs. I characterize the patient player's highest equilibrium payoff in this repeated incomplete information game and study the *common properties of her behavior* in all equilibria that (approximately) attain this highest payoff. A drawback of my approach is that it can only lead to the *attainability* part of the reputational payoff result but not the *refinement* part, which is a loss in the simultaneous-move trust game as it cannot rule out equilibria with low payoffs.

My approach has the advantage of raising and answering novel questions related to the long-run player's behavior, such as, how will the reputational types behave when they are also rational and how will the other types behave to take advantage of this rational reputational type. The canonical reputation models are not well-suited to answer these questions as the long-run player's behavior is sensitive to the reputational types' strategies due to the other types' incentives to imitate.[6] However, the commitment strategies in the canonical models are usually viewed as modeling short-hands to attain reputational payoffs instead of capturing realistic aspects of people's behaviors. Moreover, there are many commitment strategies with different behavioral implications but lead to the same commitment payoff, and the choice among those commitment strategies does not respond to changes in the payoff environment, which includes the discount factor and the long-run player's initial reputation.

My approach addresses these concerns by treating the reputational type as a strategic player who responds to changes in the payoff environment.[7] By exploiting the incentives of all types of the long-run player, I derive properties of her equilibrium behaviors and best replies that are true for all high payoff equilibria. These results are also potentially useful for future applied work both in the computation of optimal equilibria and in identifying bounds for a seller's cost of supplying high quality. In terms of complementing the canonical approach to study reputations, the implications of my results on the lowest-cost type's behavior can evaluate which of the many commitment strategies are more reasonable. From this perspective, they help to refine the commitment types when commitment types are modeling short-hands for rational types whose temptation to deviate is low or zero.

My characterization of the patient player's equilibrium payoffs is related to the study of repeated incomplete information games. Instead of restricting attention to zero-sum games (Aumann and Maschler 1965) or games

[6]These concerns are raised by Weinstein and Yildiz (2007) that one can rationalize almost every outcome by introducing types that have qualitatively different payoffs and beliefs. Their finding calls for a careful selection of types in setting up incomplete information models, including those that occur with low probability. My approach follows this spirit by requiring all types to have *reasonable* payoffs.

[7]Other approaches to endogenize commitment in dynamic games include introducing switching costs (Caruana and Einav 2008), perturbing players' belief hierarchies and introducing interdependent values (Weinstein and Yildiz 2016), etc.

with equally patient long-run players (Hart 1985, Cripps and Thomas 2003, Pęski 2014), I study non-zero sum games with unequal discount factors. This class of games is important for understanding the dynamic interactions between firms and consumers, governments and citizens, etc. Common in these scenarios, there are gains from cooperation and one class of players, such as the consumers and citizens, lack intertemporal incentives.

I develop a tractable method to construct high payoff equilibria in models with *multiple strategic types* and *no commitment type*, which can be applied to future studies of marketplaces with active learning. The belief-free equilibrium approach in Hörner, Lovo and Tomala (2011) is not applicable in this setting, as at every history where some types can extract information rent, the short-run player's best reply depends on the long-run player's type. In the equilibria I construct, despite learning and rent extraction persist in the long run, player's equilibrium reputation is linear in account and is simple to compute. Using this tractable property, I show a result that verifies the attainability of continuation payoffs taken into account the changes in beliefs (Lemma A.1). It overcomes the difficulties of applying Abreu et al.(1990)'s self-generation argument to incomplete information games.

The equilibria I construct also exhibit interesting reputation dynamics that are different from the behavioral patterns in models with commitment types such as Barro (1986), Phelan (2006), Ekmekci (2011), Liu (2011), Jehiel and Samuelson (2012) and Liu and Skrzypacz (2014), and in models where one of the rational type's strategy is exogenously assumed and acts like a commitment type such as Sobel (1985) and Schmidt (1993). These differences in the dynamic reputation building-milking behaviors highlight the role of rational reputational types in shaping players' incentives. I will elaborate on the details in subsection 4.3.

2 The Baseline Model

I introduce a repeated *trust game* that captures the lack-of-commitment problem in many economic applications. Different from the canonical reputation models with commitment types, all types of the reputation building player are rational and have qualitatively similar payoff functions, i.e., they share the same ordinal preferences over stage-game outcomes. My approach is motivated by the realistic concern that no agent is immune to reneging temptations. It helps to answer novel questions, such as how will rational reputational types behave in equilibrium, how will the other types behave to take advantage of this rational reputational type, etc.

Stage Game: Consider the following trust game between a seller (player 1, she) and a buyer (player 2, he). The buyer moves first, deciding whether to trust the seller (action T) or not (action N). If he chooses N, then both players' payoffs are normalized to 0. If he chooses T, then the seller chooses between high effort (action H) and low effort (action L).[8] If the seller chooses L, then her payoff is 1 and the buyer's payoff is $-c$. If the

[8] Despite the baseline model examines a sequential-move trust game in Bower et al.(1997) and Chassang (2010), my analysis and results also apply to the product choice game in Mailath and Samuelson (2001), Ekmekci (2011) and Liu (2011) where players move

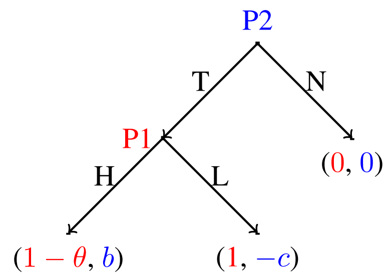

Figure 1: The stage game, where $\theta \in (0,1)$, $b > 0$, $c > 0$

seller chooses H, then her payoff is $1-\theta$ and the buyer's payoff is b, where:

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

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

The unique equilibrium outcome in the stage game is N and the seller's payoff 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.

Consider a benchmark scenario in which the seller commits to an action $\alpha_1 \in \Delta(A_1)$ before the buyer moves. When the seller can optimally choose which action to commit to, every type's optimal commitment is to play H with probability $\gamma^* \equiv \frac{c}{b+c}$ and L with probability $1-\gamma^*$. 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 called her *Stackelberg payoff*, $\gamma^* H + (1-\gamma^*)L$ is called her *Stackelberg action* and $\gamma^*(T,H) + (1-\gamma^*)(T,L)$ is called the *Stackelberg outcome*.

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, corruption (Tirole 1996) and corporate finance. The rest of this article explores the extent to which the seller can overcome this lack-of-commitment problem by building reputations and moreover, different types of the seller's payoffs and behaviors in those reputational equilibria.

simultaneously. My results also extend to stage games with imperfect monitoring. See Section 5 for more details.

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 the order described above and is depicted in Figure 1.[9]

The cost of high effort θ is the seller's private information and is perfectly persistent. The buyers' prior belief is $\pi_0 \in \Delta(\Theta)$, which is assumed to have full support. Outcomes of all past interactions are perfectly observed. Let $a_t \in \{N, H, L\}$ be outcome in period t. Let $h^t = \{a_s\}_{s=0}^{t-1} \in \mathcal{H}^t$ 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 $\sigma_1 \equiv (\sigma_\theta)_{\theta \in \Theta}$ be the seller's strategy, where $\sigma_\theta : \mathcal{H} \to \Delta(A_1)$ specifies type θ seller's action choice after the buyer plays T. A strategy σ_θ is *stationary* if it takes the same value for all $h^t \in \mathcal{H}$. A strategy σ_θ is *pure* if $\sigma_\theta(h^t)$ is a degenerate distribution for all $h^t \in \mathcal{H}$.

The seller's discount factor is $\delta \in (0, 1)$. 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 $(\sigma_\theta, \sigma_2)$.

3 Results

I present three results on a patient seller's equilibrium payoffs and behaviors. Theorem 1 provides a tractable formula for every type's *highest equilibrium payoff*, which converges to her Stackelberg payoff when the lowest possible cost vanishes. Theorem 2 shows that no type uses stationary strategies or completely mixed strategies in any equilibrium that approximately attains the seller's highest equilibrium payoff. Theorem 3 puts bounds on the seller's action frequencies, which apply not only to all of her equilibrium strategies but also to all of her equilibrium best replies. The behaviors of types that have low or even zero cost contrast to the stationary commitment strategies in the reputation literature that prescribe the same action at every history.

3.1 Equilibrium Payoffs

The seller's *payoff* in the incomplete information game is a vector $v = (v_1, v_2, ..., v_m) \in \mathbb{R}^m$, where v_j is the discounted average payoff of type θ_j. This implies that her *limiting equilibrium payoff set* is a subset of $\mathbb{R}^m$.

To ensure the robustness of my characterization against the choice of solution concepts and the ways of

[9] In the simultaneous move stage game, my results remain robust in presence of public randomizations. In the sequential move stage game, one can allow for public randomizations *before* player 2 moves. When public randomizations happen *after* player 2 moves, the attainability part of my results go through aside from the payoff upper bound in the complete information benchmark.

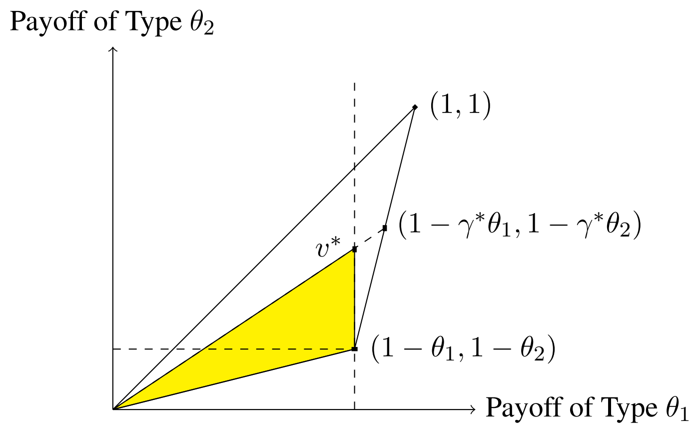

Figure 2: The limiting equilibrium payoff set V^* (in yellow) when $m = 2$.

taking the limit, I introduce a *lower bound* and an *upper bound* of the seller's limiting equilibrium payoff set such that other reasonable notions of it contains the lower bound and is contained in the upper bound. Formally, let $\underline{V}(\pi_0, \delta) \subset \mathbb{R}^m$ be the set of *sequential equilibrium* payoffs under parameter configuration (π_0, δ).[10] Let $\text{clo}(\cdot)$ be the closure of a set. The lower bound of the limiting equilibrium payoff set is given by:[11]

$$\underline{V}(\pi_0) \equiv \text{clo}\Big(\liminf_{\delta\to 1} \underline{V}(\pi_0, \delta)\Big). \tag{3.1}$$

Similarly, let $\overline{V}(\pi_0, \delta) \subset \mathbb{R}^m$ be the set of *Nash equilibrium* payoffs under (π_0, δ). The upper bound is given by:

$$\overline{V}(\pi_0) \equiv \text{clo}\Big(\limsup_{\delta\to 1} \overline{V}(\pi_0, \delta)\Big). \tag{3.2}$$

Theorem 1 shows that the two bounds coincide and provides a characterization of the seller's limiting equilibrium payoff set that unifies the incomplete information game with the complete information benchmark:

Theorem 1. *If π_0 has full support, then $\overline{V}(\pi_0) = \underline{V}(\pi_0) = V^*$ where V^* is the convex hull of the following three points:* $(0, 0, ..., 0), (1-\theta_1, ..., 1-\theta_m)$ *and* $v^* \equiv (v_1^*, ..., v_m^*)$ *with*

$$v_j^* \equiv \underbrace{(1-\gamma^*\theta_j)}_{\textit{Type } \theta_j\textit{'s Stackelberg payoff}} \underbrace{\frac{1-\theta_1}{1-\gamma^*\theta_1}}_{\textit{incomplete information multiplier}}. \tag{3.3}$$

The proof is in Appendices A and B with an intuitive explanation in Section 4. I depict V^* in Figure 1. Two extreme points of V^* are straightforward: $(0, 0, ..., 0)$ is attained by replicating the stage-game equilibrium and $(1-\theta_1, ..., 1-\theta_m)$ is attained via grim-trigger strategies. The interesting part is the attainability of $v^* \equiv$

[10] Since this is an infinitely repeated game, I adopt the notion of sequential equilibrium defined in Pęski (2014), page 658.

[11] For a family of sets $\{E_\delta\}_{\delta\in(0,1)}$, let $\liminf_{\delta\to1} E_\delta \equiv \bigcup_{\delta'\in(0,1)} \bigcap_{\delta\geq\delta'} E_\delta$ and $\limsup_{\delta\to1} E_\delta \equiv \bigcap_{\delta'\in(0,1)} \bigcup_{\delta\geq\delta'} E_\delta$.

$(v_1^*, v_2^*, ..., v_m^*)$, where v_j^* is a patient type θ_j seller's *highest equilibrium payoff*.

Equation (3.3) provides a tractable formula for v_j^*, which is the product of type θ_j's Stackelberg payoff and an *incomplete information multiplier*. This multiplier summarizes the effects of incomplete information, which is strictly below 1 and common for all types. Moreover, it only depends on the *lowest cost* in the support of the buyers' prior belief, but not on the other costs in the support and the probability distribution. Intuitively, this is because by adopting the equilibrium strategy of type θ, the seller can build a reputation for behaving equivalently to type θ. The prior probabilities only affect the time it takes to build a reputation, which has negligible payoff consequences when the seller is patient. The presence of other costs in the support has no significant impact either, as the lowest cost type is the best one available for the other types to imitate.[12]

An interesting observation is that the multiplier coincides with the *maximal probability* attached to the Stackelberg outcome such that type θ_1's payoff is no more than $1-\theta_1$, her highest equilibrium payoff under complete information. This is driven by the two linear constraints that pin down v^*. The first constraint is: the lowest cost type (i.e., type θ_1) cannot receive payoff strictly higher than $1-\theta_1$. This comes from the result in Fudenberg, et al.(1990) that in the repeated *complete information* game where θ is common knowledge, the seller's equilibrium payoff is at most $1-\theta$. To see why this is true, in every period where a myopic buyer plays T, the seller needs to play H with positive probability. Therefore, playing H in every period where the buyer plays T is one of the seller's best replies to the buyer's strategy, from which her payoff in every period is at most $1-\theta$.[13] Back to the repeated *incomplete information* game, since type θ_1 seller has the least temptation to renege, she has no good candidate to imitate and her highest payoff cannot increase compared to the complete information benchmark.

The second constraint is, according to every type's *equilibrium strategy*, the discounted average frequency of outcome (T, H) divided by that of (T, L) is no less than $\gamma^*/(1-\gamma^*)$. This follows from Gossner (2011) that the buyers are able to predict the seller's *average action* with high precision in all but a bounded number of periods, and they will cease to play T after learning that H will be played with probability less than γ^*.

Theorem 1 has several implications. First, every type aside from the lowest cost one can strictly benefit from incomplete information. That is, the gap between v_j^* and $1-\theta_j$ does not vanish as the seller becomes patient. This is interesting as a patient seller needs to extract information rent (plays L for sure while the buyer plays T) in unbounded number of periods to get a payoff strictly greater than $1-\theta_j$. However, extracting information rent inevitably reveals information about her type, which undermines her ability to extract rent in the future.

Second, the multiplier converges to 1 as θ_1 vanishes to 0. That is, every type's highest equilibrium payoff

[12] This is similar to models of Coasian bargaining (Gul et al.1986) that the informed player's information rent only depends on the best type she can imitate when bargaining friction vanishes. Nevertheless, both insights rely on a private value assumption that θ does not directly affect the buyers' payoffs. Under interdependent values, the probabilities and the other types matter (Pei 2018).

[13] This argument breaks down in repeated *incomplete information* games, as the short-run player's incentive to trust does not imply that any particular type of the long-run player exerts high effort with positive probability. This indicates that some types of the long-run player may obtain strictly higher payoffs.

converges to her Stackelberg payoff. This suggests that despite no type is immune to reneging temptations, every type of the long-run player can approximately attain her optimal commitment payoff.

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 terms of social welfare, every payoff on the Pareto frontier is approximately attainable when the seller is patient and the lowest possible cost is small. Formally, let $\overline{v} \equiv (v_0, v_1, ..., v_m) \in \mathbb{R}^{m+1}$ where v_0 is the buyers' discounted average payoff and v_j is type θ_j's payoff. I say that $\overline{v}$ is *incentive compatible* for the buyers if there exists $(\alpha_1, a_2) \in \Delta(A_1) \times A_2$ such that: $a_2 \in \arg\max_{a_2' \in A_2} u_2(\alpha_1, a_2')$, $u_1(\theta_j, \alpha_1, a_2) = v_j$ 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 incentive compatible payoffs. The Pareto frontier of $\overline{V}^*$ is a line connecting $(b, 1-\theta_1, ..., 1-\theta_m)$ and $(0, v_1^*, ..., v_m^*)$, which according to Theorem 1 and Corollary 1, can be approximated by equilibrium payoffs when θ_1 is small and δ is large.

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 equilibrium.*

Fourth, in terms of the buyers' learning, the lowest cost type seller fully reveals her private information for unboundedly many times in every equilibrium where the high-cost types can extract information rent. Formally, let $\mathcal{H}^{(\sigma_\theta, \sigma_2)}$ be the set of histories that occur with positive probability under $(\sigma_\theta, \sigma_2)$. A subset of histories $\mathcal{H}'$ is an *independent set* if no pair of elements in $\mathcal{H}'$ can be ranked via the predecessor-successor relationship. This implication is stated as the following corollary, the proof of which can be found in Online Appendix B.

Corollary 3. *For every $N \in \mathbb{N}$ and $v = (v_1, ..., v_m) \in V^*$ such that $v_j > 1 - \theta_j$ for every $j \in \{2, ..., m\}$, there exists $\overline{\delta}$ such that in every Nash equilibrium that attains v when $\delta > \overline{\delta}$, there exists an independent set $\mathcal{H}'$ with $|\mathcal{H}'| > N$ and $\mathcal{H}' \subset \mathcal{H}^{(\sigma_{\theta_1}, \sigma_2)}$, such that player 2's belief attaches probability 1 to θ_1 for every $h^t \in \mathcal{H}'$.*

3.2 Equilibrium Behaviors under Incomplete Information

This subsection focuses on environments where incomplete information is non-trivial, namely, $m \geq 2$. I derive properties on the patient seller's behavior that are true for *all* Nash equilibria that can approximately attain the seller's highest equilibrium payoff v^*. That being said, my conclusions also extend to other solution concepts such as Perfect Bayesian equilibrium and sequential equilibrium.

As a first step, I show that in the repeated incomplete information game, no type of the long-run player uses stationary strategies or has completely mixed best replies, no matter how low her cost is. This contrasts to the stationary commitment strategies in the reputation literature that prescribe the same action at every history.

Theorem 2. *When $m \geq 2$, 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 uses stationary strategies or completely mixed strategies in any Nash equilibrium that attains payoff within ϵ of v^*. Moreover, no type has a completely mixed equilibrium best reply.*[14]

The proof is in Online Appendix C, which exploits the interplay between supermodularity and incomplete information. For a heuristic proof, suppose towards a contradiction that type θ_j's best reply is to mix whenever she is trusted. Then both playing L at every on-path history and playing H at every on-path history are her best replies. As lower cost types enjoy a comparative advantage in playing H and vice versa, every type that has strictly higher cost plays L with probability 1 at every on-path history and every type that has strictly lower cost plays H with probability 1 at every on-path history.[15] However, none of these pure stationary strategies are compatible with the requirement that type θ_j's payoff is strictly above $1 - \theta_j$ for every $j \geq 2$.

First, suppose there exists a type θ_k that plays L with probability 1 at every on-path history. According to the learning argument in Fudenberg and Levine (1992), if the short-run players were to face type θ_k, they will eventually believe that L is played with probability greater than $1-\gamma^*$ in all future periods, after which they have a strict incentive to play N, leaving type θ_k a discounted average payoff close to 0 when δ is high enough.

Next, suppose $j \geq 2$ and types θ_1 to θ_{j-1} play H with probability 1 at every on-path history. Then after type θ_j plays L for one period, she becomes the lowest cost type in the support of the short-run players' posterior belief. This implies that type θ_j cannot extract information rent in the continuation game,[16] in which case her discounted average payoff cannot exceed $(1 - \delta) + \delta(1 - \theta_j)$. The latter converges to $1 - \theta_j$ as $\delta \to 1$. This contradicts the hypothesis that type θ_j's equilibrium payoff is close to v_j^*, which is strictly greater than $1 - \theta_j$.

The above argument suggests that the conclusion of Theorem 2 remains valid for a type whose cost of high effort is zero. This is somewhat surprising as the long-run player's best reply cannot be completely mixed despite she is indifferent between high and low effort in the stage game. It also implies that every type (including the zero cost one) faces *non-trivial intertemporal incentives* in all high-payoff equilibria. That is to say, her stage-game action choice has a non-trivial impact on the frequencies with which her opponents trust her in the future. To see this, suppose towards a contradiction that the zero-cost type is indifferent between H and L at every history, then playing L at every history and playing H at every history lead to the same frequency with which player 2 plays T. Consequently, every positive cost type plays L with probability 1 at every on-path history. Conditional on facing such a type, the short-run players will stop playing T after observing L for a bounded number of periods,

[14]The conclusion in Theorem 2 fails when $m = 1$. I show in Online Appendix E.1 that in the repeated complete information game, there exists a sequential equilibrium in which player 1 receives payoff v^* and her equilibrium strategy is stationary.

[15]If we order the states and actions according to $T \succ N$, $H \succ L$ and $\theta_1 \succ \theta_2 \succ ... \succ \theta_m$, then the stage game payoff satisfies a monotone-supermodularity condition in Liu and Pei (2017). This is sufficient to guarantee the monotonicity of the sender's strategy with respect to her type in one-shot signalling games. I use an implication of this result on repeated signalling games (Pei 2018).

[16]This does not follow from the previous conclusion that type θ_1's payoff is less than $1 - \theta_1$. This is because under Nash equilibrium, a type that occurs with zero probability is not equivalent to a type that is excluded in the type space. To address this concern, I show in Proposition B.1 that the lowest cost type *in the support of the posterior belief* cannot extract information rent in the continuation game.

leaving every positive-cost type a discounted average payoff close to 0.

To further explore the properties of these non-stationary behaviors, I establish bounds on the long-run player's action frequencies that can be applied to all of her equilibrium best replies.

Theorem 3. *For every small enough $\varepsilon > 0$, there exists $\overline{\delta}$ such that when $\delta > \overline{\delta}$, in every Nash equilibrium $\big((\sigma_\theta)_{\theta\in\Theta}, \sigma_2\big)$ that attains payoff within ε of v^*,*

1. *For every $\theta \neq \theta_m$ and under every pure strategy best reply $\widehat{\sigma}_\theta$ of type θ against σ_2,*

$$\frac{\mathbb{E}^{(\widehat{\sigma}_\theta,\sigma_2)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t \mathbf{1}\{a_t=(T,H)\}\Big]}{\mathbb{E}^{(\widehat{\sigma}_\theta,\sigma_2)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t \mathbf{1}\{a_t=(T,L)\}\Big]} \geq \frac{\gamma^*-\varepsilon}{1-(\gamma^*-\varepsilon)}. \tag{3.4}$$

2. *For every $\theta \neq \theta_1$ and under every pure strategy best reply $\widehat{\sigma}_\theta$ of type θ against σ_2,*[17]

$$\frac{\mathbb{E}^{(\widehat{\sigma}_\theta,\sigma_2)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t \mathbf{1}\{a_t=(T,H)\}\Big]}{\mathbb{E}^{(\widehat{\sigma}_\theta,\sigma_2)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t \mathbf{1}\{a_t=(T,L)\}\Big]} \leq \frac{\gamma^*+\varepsilon}{1-(\gamma^*+\varepsilon)}. \tag{3.5}$$

The proof is in Appendix C. According to Theorem 3, when the patient player's cost is not the highest one in the support, the relative (discounted average) frequency of high effort and low effort under each of her pure strategy best reply cannot be lower than $\gamma^*/(1-\gamma^*)$. Conversely, when her cost is not the lowest one in the support, this relative frequency under each of her pure strategy best reply cannot be greater than $\gamma^*/(1-\gamma^*)$. For types ranging from θ_2 to θ_{m-1}, the two bounds together *pin down* the long-run frequency of each stage-game outcome under each of their pure strategy best reply in each high payoff equilibrium, namely,

$$\mathbb{E}^{(\widehat{\sigma}_\theta,\sigma_2)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t \mathbf{1}\{a_t=a\}\Big] \text{ for every } a \in \{(T,L),(T,H),N\} \text{ and every } \widehat{\sigma}_\theta \text{ that best replies } \sigma_2.$$

This is because type θ_j long-run player's payoff is approximately v_j^* in every high payoff equilibrium, which together with the tight bounds on relative frequencies pin down the frequency of each stage-game outcome.

To better understand the result's economic implications, it is helpful to distinguish the bounds in Theorem 3 and the following implication of Fudenberg and Levine (1992)'s learning result, that the relative discounted average frequency according to every type's *equilibrium strategy* is no less than $\gamma^*/(1-\gamma^*)$. The lower and upper bounds in Theorem 3 apply to *all pure strategies* in the support of any given type's equilibrium strategy,

[17] In some high payoff equilibria, playing H at every history is type θ_1's equilibrium best reply, despite it cannot be her equilibrium strategy. Examples include the constructed equilibria in the proof of Theorem 1.

not just on average. A direct implication is that they apply to all equilibrium best replies, pure or mixed. As an illustrative example, the strategy of playing the Stackelberg action at every history satisfies the requirement in Fudenberg and Levine (1992) but has been ruled out by Theorem 3. This is because for some pure strategies in its support (e.g. playing L at every history), the frequency of L exceeds the bound in (3.4), and for others in its support (e.g. playing H at every history), the frequency of L falls below the bound in (3.5).

Theorem 3 has more implications than just ruling out completely mixed best replies. First, it sheds light on how different types of the long-run player cherry-pick actions based on the history of play. For the lowest cost type, she has a strict incentive to play H when she has already played L with occupation measure greater than:

$$\underbrace{\frac{1-\theta_1}{1-\gamma^*\theta_1}}_{\text{incomplete info multiplier}} \qquad (1-\gamma^*). \tag{3.6}$$

In general, this lower bound on the occupation measure of H helps to evaluate which of the many commitment strategies in the canonical reputation models are more reasonable.

For types ranging from θ_2 to θ_{m-1}, all of their equilibrium best replies lead to the same action frequency, the only difference between them is in the timing of actions. As a result, these types have a strict incentive to play H when the occupation measure of L exceeds (3.6) and have a strict incentive to play L when the occupation measure of H exceeds

$$\underbrace{\frac{1-\theta_1}{1-\gamma^*\theta_1}}_{\text{incomplete info multiplier}} \qquad \gamma^*. \tag{3.7}$$

Second, the bounds that apply to all pure best replies have stronger testable implications compared to the ones that only apply to the equilibrium strategies. This distinction is economically important as in many markets, researchers can only observe a few *realized paths* of the seller's dynamic behavior instead of an entire distribution of action paths. To fix ideas, consider a scenario in which the equilibrium being played is approximately optimal for the seller, a researcher knows γ^* and observes the frequencies of outcomes under a realized path of play. However, he does not observe the entire distribution over the paths of play and does not know which type of seller he is facing as well as her cost. Theorem 3 can identify the position of the seller in the cost distribution (e.g. whether she is the highest or the lowest cost or a type with intermediate cost) as well as bounds on θ_1:

- If the relative frequency of H and L is approximately $\gamma^*/(1-\gamma^*)$, then the researcher can identify θ_1 via the frequency of N. This is because the observed frequency of N is approximately

$$\frac{(1-\gamma^*)\theta_1}{1-\gamma^*\theta_1} \tag{3.8}$$

under each pure best reply, which is a strictly increasing function of θ_1.

- If the relative frequency of H and L is significantly greater than $\gamma^*/(1-\gamma^*)$, then the researcher concludes that the long-run player's cost is the lowest in the support. He can then identify a lower bound on θ_1 using the observed frequency of N. This is because the observed frequency of N under the lowest cost type's best reply provides a lower bound on (3.8), which translates into a lower bound on θ_1.
- If the relative frequency of H and L is significantly smaller than $\gamma^*/(1-\gamma^*)$, then the researcher concludes that the long-run player's cost is the highest in the support. He can then identify an upper bound on θ_1 using the observed frequency of N. This is because the observed frequency of N under the highest cost type's best reply provides an upper bound on (3.8), which translates into an upper bound on θ_1.

4 Ideas Behind the Proof of Theorem 1

I explain the ideas behind the proof using an example with two types, that is $\Theta = \{\theta_1, \theta_2\}$. I decompose the theorem into two statements. First, every payoff that is bounded away from V^* is not attainable in any Nash equilibrium when δ exceeds some cutoff, namely $\overline{V}(\pi_0) \subset V^*$. Second, every payoff in the interior of V^* is attainable in sequential equilibrium when δ is high enough, namely $\underline{V}(\pi_0) \supset V^*$.

4.1 Necessity of Constraints

Let $\sigma \equiv (\sigma_{\theta_1}, \sigma_{\theta_2}, \sigma_2)$ be a generic Nash equilibrium. Recall that V^* is characterized by two linear constraints. First, the equilibrium payoff of type θ_1 cannot exceed $1-\theta_1$. Second, the ratio between the weight attached to (T,H) and that attached to (T,L) is no less than $\gamma^*/(1-\gamma^*)$.

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 $\sigma_2(h^t)$ attaches positive probability to T, let $\Theta(h^t)$ be the support of the short-run player's posterior belief at h^t. The highest action path is defined as:

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

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 two cases.

For the second constraint, I provide an alternative explanation based on the *relative rate* of reputation building and milking. The motivation is to offer a new perspective that complements the arguments based on the canonical reputation logic (see subsection 3.1). First, I introduce another 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 player 2 plays T but $\overline{\sigma}_{\theta_2}$ prescribes L. Player 2'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 *reputation*. The above argument implies that when type θ_2 plays according to $\overline{\sigma}_{\theta_2}$, she can only extract information rent 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. $\eta(h^t, H) > \eta(h^t)$.

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 player 1 rebuilds her reputation (by playing H) to the speed with which her reputation deteriorates (by playing L). Player 2's incentive to trust requires H to be played with probability at least γ^*. The martingale property of beliefs bounds this relative speed from above:

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

In summary, $(1-\gamma^*)/\gamma^*$ is the *maximal relative frequency* between L and H such that player 1 has an equilibrium strategy that satisfies: first, player 2 has an incentive to play T in every period; and second, as long as player 1 plays L with relative frequency below this threshold, her posterior reputation is no less than her initial reputation.

4.2 Overview of Equilibrium Construction

To approximate v^*, let $v(\gamma) \equiv \big(v_j(\gamma)\big)_{j=1}^m$ with

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

for every $j \in \{1, 2, ..., m\}$ and $\gamma \in [\gamma^*, 1]$. An example of $v(\gamma)$ is shown in Figure 3. By definition, $v_j(\gamma^*) = v_j^*$ and $v_j(1) = 1 - \theta_j$. The sufficiency part of Theorem 1 is implied by the following proposition:[18]

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 $\pi_0 \in \Delta(\Theta)$ satisfying $\pi_0(\theta_1) \geq \overline{\eta}$, there exists an equilibrium in which player 1's payoff is $v(\gamma)$.*

[18]This directly implies Theorem 1 when players have access to a public randomization device in the beginning of each period. As actions can be perfectly monitored, the public randomization device can be dispensed according to Fudenberg and Maskin (1991).

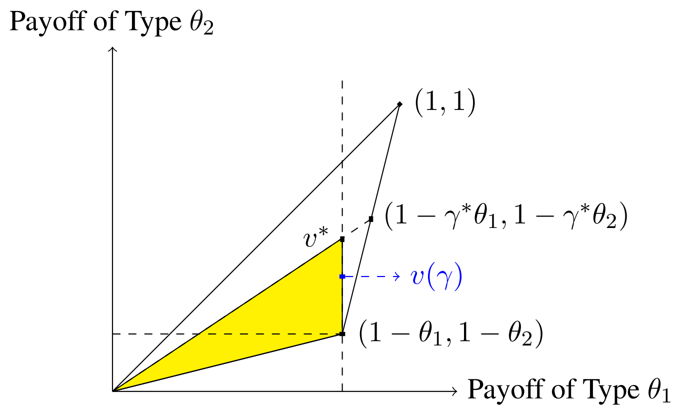

Figure 3: V^* in yellow and $v(\gamma)$ in blue for some $\gamma \in (\gamma^*, 1)$.

The proof is in Appendix A. The rest of this subsection provides an overview of the ideas behind the construction. Part I describes players' strategies and their induced beliefs. Part II summarizes how my construction overcomes the aforementioned challenges, such as how to ensure that the promised continuation payoffs can be delivered in the continuation game, how to satisfy the rational reputational type's incentives, how to make the high-cost types extracting information rent while preserving their informational advantages, etc.

Part I: Equilibrium Strategies The equilibrium consists of three phases. Play starts from an *active learning phase* and transits into one of the two *absorbing phases* in finite time. Phase transition happens either when the long-run player has played H too many times or has played L too many times. The high-cost type extracts information rent only in the active learning phase and the long-run player's continuation payoff in the absorbing phase depends on her action choices in the active learning phase before transition happens. Moreover, the long-run player's reputation is built and milked gradually. In another word, the absolute speed of learning is slow.

I keep track of two state variables. First, the probability that player 2's posterior attaches to type θ_1, which I refer to as *reputation*, denoted by $\eta(h^t)$. Second, the remaining weight of each stage-game outcome, denoted by $p^N(h^t)$, $p^H(h^t)$ and $p^L(h^t)$. The initial values of these state variables are:

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

Play starts from an *active learning phase*, in which player 2 plays T in every period. Since the belief process is a martingale, each type of player 1's mixed action at h^t is pinned down by the following pair of belief-updating formulas:

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

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

where η^* is an arbitrary number between $\gamma^*\eta(h^0)$ and $\eta(h^0)$ and λ is a small positive number. According to (4.5) and (4.6), H is played with probability at least γ^* at history h^t. This implies that player 2 has an incentive to play T. The parameter η^* is player 1's reputation lower bound in the active learning phase, which is to satisfy player 2's incentive constraint by the end of this phase (Appendix A.3). The parameter λ measures the *absolute speed* of learning, which needs to be small enough due to the following implication of the Taylor's Theorem.

Lemma 4.1. *For every $\epsilon > 0$, there exists $\overline{\lambda} > 0$ such that if the ratio between the number of times with which L and H are played is strictly less than $(1-\gamma^*-\epsilon)/(\gamma^*+\epsilon)$, then player 1's reputation is no less than $\eta(h^0)$ when the belief updating formulas are given by (4.5) and (4.6) with $\lambda \in (0, \overline{\lambda})$.*

The other three state variables evolve according to the realized stage game outcome, according to:

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

for every $a, a_t \in \{N, H, L\}$. Intuitively, player 1 has three credit lines. Each of them represents the occupation measure of a 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

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

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 a convex combination of outcomes N and (T, H). Intuitively, when type θ_1 reaches the first absorbing phase, her continuation payoff equals her promised payoff in the active learning phase. However, if types other θ_1 reaches this phase, then their continuation payoffs are less than what they have been promised before, which discourages them from doing so. This inequality is driven by the supermodularity of stage-game payoffs, namely the high-cost types having stronger preferences for (T, L).

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 makes it negative. For illustration purposes, I focus on the simple situation in which $p^L(h^t) = 0$ and relegate the complicated case in which $p^L(h^t) \in \big(0, (1-\delta)\delta^t\big)$ to Appendix A. Type θ_j's continuation payoff is:

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

The resulting payoff vector $\big(v_j(h^t)\big)_{j=1}^m$ can be delivered by a combination of outcomes N and (T, H).

Part II: Summary of Main Ideas The ideas behind my construction are summarized as follows. First, introduce reputation cycles so that the long-run player can rebuild her reputation after extracting information rent. Second, design the absorbing phases to provide incentives to all types of the long-run player, including but not limited to the rational reputational type. They also ensure that the promised continuation payoff to the high-cost type can be credibly delivered. Third, lowering the absolute speed of learning reduces the amount of reputation loss after fixing the relative frequencies of actions and the relative rate of learning as defined in (4.3).

The first idea is incorporated in the design of the active learning phase. The long-run player can flexibly choose when to build and milk reputations, and her equilibrium reputation only depends on the number of times she has played each action in the past. As extracting information rent requires learning (Corollary 3), the high-cost and low-cost types need to mix with different probabilities, and therefore, one needs to provide them incentives to mix at each of those histories. This incentive issue is resolved by the transition rules to the absorbing phases. In particular, if the long-run player front-loads the play of L, then the second absorbing phase is reached at an earlier date after which it becomes impossible for her to extract information in the future. This is because whenever she plays L at histories where she is supposed to play H, she will be facing a grim-trigger punishment.

A hidden issue remains, which is, what will happen if the high-cost type front-loads the play of H? This is problematic as her continuation payoff will become so high that it cannot be delivered in any continuation equilibrium. The first absorbing phase is designed to address this issue: if she front-loads the play of H, then a transition to the first absorbing phase happens 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 first absorbing phase ensures that at every history of the active learning phase, the ratio between the weight of L and that of H in the long-run player's continuation value does not exceed some cutoff. As a result, her continuation value is always within V^*. This is stated as Lemma A.1 in Appendix A and the proof can be found in Online Appendix A.

Given the above building blocks, what remains to be designed is the detailed learning process. In order to maximize the long-run player's equilibrium payoff, one needs to maximize the frequency of outcome (T, L) while simultaneously providing incentives for the short-run players to trust. This leads to the role of slow learning. This is because first, the short-run player's incentive to trust translates into an upper bound on the relative rate of learning defined in (4.3). Second, fixing the relative rate of learning and the long-run frequencies of H and L, the amount of reputation loss per period vanishes as the *absolute speed of learning* goes to zero (Lemma 4.1). As a result, lowering the absolute speed of learning improves the patient player's long-term reputation without compromising on the short-run players' willingness to trust. This allows for an increase in the long-run frequency of low effort without sacrificing the patient player's long-term reputation, which increases her payoff.[19]

4.3 Comparisons on Equilibrium Dynamics

I compare the equilibrium dynamics in my model to those in reputation models with behavioral biases (Jehiel and Samuelson 2012), models of reputation cycles (Sobel 1985, Phelan 2006, Liu 2011, Liu and Skryzpacz 2014), models that exhibit gradual learning (Benabou and Laroque 1992, Ekmekci 2011) and the capital-theoretic models of reputation building (Board and Meyer-ter-Vehn 2013, Bohren 2018, Dilmé 2018).

Analogical-Based Reasoning Equilibria: The long-run player alternates between her actions in order to manipulate her opponents' belief is reminiscent of the analogy-based reasoning equilibria of Jehiel and Samuelson (2012). 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 myopic players mistakenly believe that the strategic long-run player is playing a stationary strategy. In the trust game, their results imply that the strategic long-run player's behavior experiences a *reputation building phase* (or a *reputation milking 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 according to the Stackelberg frequencies. The myopic players' posterior belief fluctuates within a small neighborhood of the cutoff belief, implying that the long-run player's type is never fully revealed.

Comparing my model to theirs, there are two qualitative differences in the reputation dynamics that highlight the distinctions between rational and analogical-based myopic players. First, the expected duration of the reputation manipulation phase is finite in my model while it is infinite in theirs. This is driven by the constraint that type θ_1's equilibrium payoff cannot exceed $1-\theta_1$, which comes from the rational myopic players' ability to correctly predict the long-run player's average action in *every period*. This constraint is absent when myopic players use analogy-based reasoning as they can only correctly predict the average action *across all periods*. Second, the myopic players can learn the true state with positive probability in every equilibrium that approximately attains v^* (Corollary 3), while in Jehiel and Samuelson (2012), the probability with which they fully learn the state is zero. This is because analogy-based myopic players' posterior beliefs only depend on the empirical frequencies of the observed actions. As a result, their beliefs are not responsive enough to each individual observation.

[19]This argument only applies to a patient long-run player. This is because decreasing the absolute speed of learning also increases the number of periods required to build a reputation (i.e. reaching the point at which the high cost types can shirk for sure in one period), which has non-negligible payoff consequences when the long-run player is impatient.

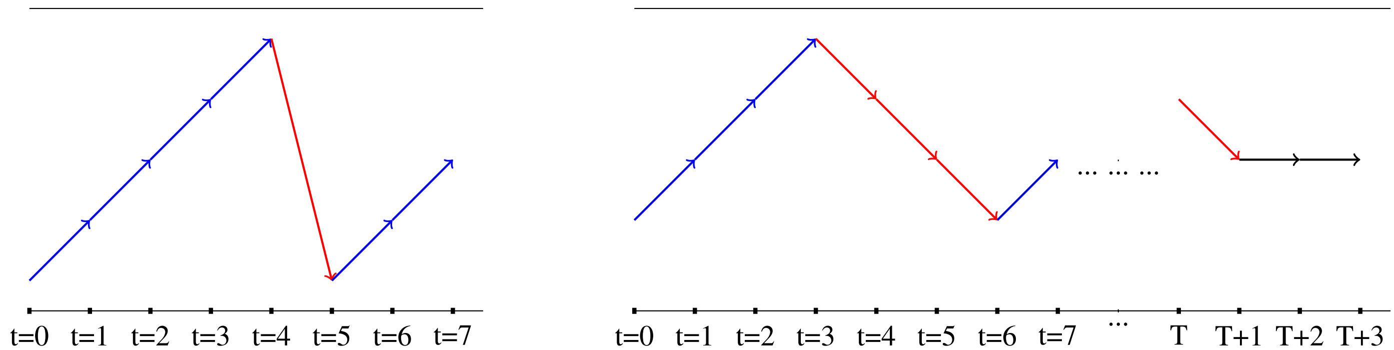

Figure 4: The horizontal axis represents the timeline and the vertical axis measures the informed player's reputation, i.e. probability of the commitment type or lowest cost type. Left: Reputation cycles in Phelan (2006). Right: A sample path of the reputation cycle in my model.

Reputation Building-Milking Cycles: The behavioral pattern that a patient player builds her reputation in order to milk it in the future has been identified in commitment type models with either changing types (Phelan 2006) or limited memories (Liu 2011, Liu and Skrzypacz 2014). Two differences emerge once comparing the reputation dynamics in my model to theirs.

First, 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. This is driven by the constraint that the reputational type's equilibrium payoff cannot exceed $1 - \theta_1$, which arises only when the reputational type is rational and faces a strict temptation to betray.

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. This is because the commitment types in their models never betray so one misbehavior reveals that the long-run player's rationality. In my model, the comparison between good and bad types is not that stark as all types share the same ordinal preferences over stage-game outcomes and have strict temptations to betray. In the long-run player's optimal equilibrium, the lowest-cost type betrays with positive probability for unbounded number of periods, which does not reduce 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. As 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 neither significantly affects the amount of sales nor the prices of a reputable seller who has obtained many positive ratings in the past. This observation is better explained by the reputation dynamics in my model.

Benabou and Laroque (1992) and Ekmekci (2011) study games with commitment types and imperfect monitoring. 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. The different driving forces behind gradual learning also lead to different long-run learning outcomes. In my model, the reputation building-milking cycles will stop in finite time while in theirs, reputation cycles can last forever. This is because the rational reputational type faces strict incentives to betray, leading to an upper bound on her equilirium payoff. Moreover, the myopic players never fully learns the state in their models, while in mine, they perfectly learn the lowest-cost type with positive probability in every high payoff equilibrium.

Capital-Theoretic Models of Reputation: Reputation cycles also occur in the Poisson good news models of Board and Meyer-ter-Vehn (2013) and Dilmé (2018).[20] They characterize Markov equilibria in which the long-run player exerts effort if and only if her belief is below a cutoff. Different from my model, reputation jumps up immediately after the arrival of good news. Moreover, the long-run player's reputation only depends on the most recent time of news arrival in their models while it depends on the frequencies of her past actions in mine. These distinctions are caused by the differences in the source of learning. In my model, learning arises from the differences in different types' behaviors, while in their models, all types adopt the same behavior but face different news arrival rates. In terms of the applications, my model fits into online platforms where feedback arrives frequently while their models fit into markets with infrequent news arrivals.

5 Extensions

Simultaneous-Move Stage Game: Consider a simultaneous-move 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. Three interpretations of this game are provided in Online Appendix D.

In the repeated version of this game, players' past action choices are perfectly monitored and the public history $h^t \equiv \{a_{1,s}, a_{2,s}\}_{s=0}^{t-1}$ consists of players' past action choices. Players can also have access to a public randomization device. Other features of the game remain the same as in the baseline model.

[20] In the Poisson bad news model of those papers and in the Brownian model of Bohren (2018), the long-run player's effort is increasing in her reputation, which differs from the high-cost types' behaviors of my model that cheat for sure when reputation is close to 1.

For the results on equilibrium payoffs, recall the definition of $v^* \equiv (v_j^*)_{j=1}^m$ in (3.3). Using the idea described in subsection 4.2, one can show that v^* is approximately attainable as $\delta \to 1$. The constructed equilibria have the feature that only outcomes (T,H), (T,L) and (N,L) occur with positive probability. Under a supermodularity condition on the stage-game payoffs, namely:

$$0 \leq d(\theta_j) - d(\theta_i) \leq \theta_j - \theta_i \quad \text{for every} \quad j < i, \tag{5.1}$$

one can show that for every $j \in \{1, 2, ..., m\}$, v_j^* is type θ_j patient long-run player's highest equilibrium payoff.

Under the supermodularity condition in (5.1), the conclusions in Theorem 2 and Theorem 3 extend when the stage-game is of simultaneous move. In particular, no type of the long-run player uses a stationary strategy nor has a completely best reply in any equilibrium that approximately attains v^*. For the bounds on the long-run player's action frequencies that apply to all of her pure strategy best replies, one needs to replace $a_t = (T,H)$ and $a_t = (T,L)$ in (3.4) and (3.5) with $a_{1,t} = H$ and $a_{1,t} = L$, respectively. The driving force behind these results is the monotone-supermodularity of players' stage-game payoffs, which is a sufficient condition for the monotonicity of all equilibria in one-shot signalling games (Liu and Pei 2017). Their result implies that in every Nash equilibrium of the repeated game and compare every best reply of type θ_i with every best reply of type θ_j with $i > j$, the discounted average frequency of H under the former must be weakly lower compared to the discounted average frequency under the latter.

Stage Game with 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.[21] 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. The cost of effort is linear and the marginal cost of effort is the agent's persistent private information.[22]

The stage game is of sequential-move. Different from the baseline model, 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

[21] 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.

[22] 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.

$1-\theta_i e$. 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 exceeds $\gamma^* \equiv \frac{c}{b+c}$.

Consider the repeated version of this game in which the public history consists of player 2's actions and the realized outputs. 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 $h^t = \{a_{2,s}, y_s\}_{s=0}^{t-1} \in \mathcal{H}^t$ be a public history with $\mathcal{H} \equiv \bigcup_{t=0}^{+\infty} \mathcal{H}^t$ the set of public histories. Let $h_1^t = \{a_{1,s}, a_{2,s}, y_s\}_{s=0}^{t-1} \in \mathcal{H}_1^t$ 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 $\sigma_2 : \mathcal{H} \rightarrow \Delta(A_2)$ be player 2's strategy and let $\sigma_\theta : \mathcal{H}_1 \rightarrow \Delta(A_1)$ be type θ player 1's strategy, with $\sigma_1 \equiv (\sigma_\theta)_{\theta\in\Theta}$.

The above game with a continuum of effort, linear effort cost and imperfect monitoring is equivalent to the baseline model with binary effort and perfect monitoring. To see this, choosing effort level e under imperfect monitoring is equivalent to choosing a mixed action $eH + (1-e)L$ under perfect monitoring. In terms of the results on payoffs, one can show that v_j^* is type θ_j's highest equilibrium payoff when she is patient, and payoff vector v^* is approximately attainable when δ is close to 1. In terms of behaviors, the bounds on the relative frequencies can be applied to realized paths of public signals, namely, one needs to replace $a_t = (T, H)$ and $a_t = (T, L)$ in (3.4) and (3.5) with $y_t = G$ and $y_t = B$, respectively.

6 Conclusion

This paper introduces a new approach to study reputations in which all types of the long-run player are rational and are facing lack-of-commitment problems. Motivated by various applications of reputation effects, I require all types of the long-run player to have reasonable payoffs, that is, they share the same ordinal preferences over stage-game outcomes. This captures, for example, that people are likely to understand that firms can benefit from consumers' purchases, providing high quality and making timely deliveries are costly for firms, etc. In contrast, a firm's cost of providing high quality and making timely deliveries tend to be its private information.

I characterize the patient long-run player's attainable payoffs when she can build a reputation about her payoff type. My formula for every type's highest equilibrium payoff identifies a sufficient statistic for the role of incomplete information and helps to compute each type of the long-run player's gain from persistent private information. Focusing on equilibria that approximately attain the long-run player's highest equilibrium payoff, I characterize the properties of the long-run player's behavior that applies to all of her equilibrium strategies and even all of her equilibrium best replies.

The implications of my results on the lowest-cost type's behavior help to select among the many commitment strategies in the reputation literature. My results deliver behavioral predictions that can be tested by observing a realized path of the long-run player's actions, rather than the cross-section distributions. I develop a tractable method to construct high-payoff equilibria which is potentially useful for future work, both in the equilibrium analysis of dynamic markets with active learning, and in the theoretical studies of repeated incomplete information games where not all parties are patient.

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. The proof of a technical lemma (Lemma A.1) is relegated to Online Appendix A.

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*

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

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 proof is in Online Appendix A. The challenge 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.

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:

$\theta = \theta_i$	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 strictly higher than her highest payoff in the repeated complete information game.

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.*

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 3

Statement 1: Suppose there exists type $\theta_i \neq \theta_m$ and a pure strategy $\widehat{\sigma}_{\theta_i}$ that is type θ_i's best reply to σ_2, such that

$$\frac{\mathbb{E}^{(\widehat{\sigma}_{\theta_i},\sigma_2)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t \mathbf{1}\{a_t=(T,H)\}\Big]}{\mathbb{E}^{(\widehat{\sigma}_{\theta_i},\sigma_2)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t \mathbf{1}\{a_t=(T,L)\}\Big]} = \frac{\gamma_i}{1-\gamma_i} \tag{C.1}$$

for some $\gamma_i < \gamma^*$. Let p_i be the discounted average frequency with which player 2 plays T under $(\widehat{\sigma}_{\theta_i}, \sigma_2)$.

Let $\widehat{\sigma}_{\theta_m}$ be an arbitrary pure strategy best reply of type θ_m to σ_2. Let p_m be the discounted average frequency with which player 2 plays T under $(\widehat{\sigma}_{\theta_m}, \sigma_2)$ and let γ_m be pinned down via:

$$\frac{\mathbb{E}^{(\widehat{\sigma}_{\theta_m},\sigma_2)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t \mathbf{1}\{a_t=(T,H)\}\Big]}{\mathbb{E}^{(\widehat{\sigma}_{\theta_m},\sigma_2)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t \mathbf{1}\{a_t=(T,L)\}\Big]} = \frac{\gamma_m}{1-\gamma_m}. \tag{C.2}$$

Type θ_i prefers $\widehat{\sigma}_{\theta_i}$ to $\widehat{\sigma}_{\theta_m}$ and type θ_m prefers $\widehat{\sigma}_{\theta_m}$ to $\widehat{\sigma}_{\theta_i}$ imply that $p_i \geq p_m$ and $\gamma_i \geq \gamma_m$. This implies that according to type θ_m's equilibrium strategy σ_{θ_m},

$$\frac{\mathbb{E}^{(\sigma_{\theta_m},\sigma_2)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t\mathbf{1}\{a_t=(T,H)\}\Big]}{\mathbb{E}^{(\sigma_{\theta_m},\sigma_2)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t\mathbf{1}\{a_t=(T,L)\}\Big]} \leq \frac{\gamma_i}{1-\gamma_i},$$

or equivalently,

$$\gamma_i\mathbb{E}^{(\sigma_{\theta_m},\sigma_2)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t\mathbf{1}\{a_t=(T,L)\}\Big]-(1-\gamma_i)\mathbb{E}^{(\sigma_{\theta_m},\sigma_2)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t\mathbf{1}\{a_t=(T,H)\}\Big]>0. \quad \text{(C.3)}$$

Since type θ_m's payoff from σ_{θ_m} is at least $v_m^*-\varepsilon$, which is strictly greater than $1-\theta_m$ when ε is small enough. This places a lower bound on p_m. If type θ_m plays according to σ_{θ_m}, then the learning arguments in Fudenberg and Levine (1992) and Gossner (2011) imply that for every $\varepsilon>0$, there exists $\overline{\delta}$ such that when $\delta>\overline{\delta}$,

$$\gamma^*\mathbb{E}^{(\sigma_{\theta_m},\sigma_2)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t\mathbf{1}\{a_t=(T,L)\}\Big]-(1-\gamma^*)\mathbb{E}^{(\sigma_{\theta_m},\sigma_2)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t\mathbf{1}\{a_t=(T,H)\}\Big]<\varepsilon. \quad \text{(C.4)}$$

This contradicts (C.3) once we pick ε to be small enough, which establishes the lower bound on the relative frequencies.

Statement 2: Suppose towards a contradiction that according to one of type θ ($\neq \theta_1$)'s pure strategy best reply to σ_2, denoted by $\widehat{\sigma}_\theta$,

$$\frac{\mathbb{E}^{(\widehat{\sigma}_\theta,\sigma_2)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t\mathbf{1}\{a_t=(T,H)\}\Big]}{\mathbb{E}^{(\widehat{\sigma}_\theta,\sigma_2)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t\mathbf{1}\{a_t=(T,L)\}\Big]} = \frac{\gamma}{1-\gamma} \quad \text{(C.5)}$$

where $\gamma>\gamma^*$. Let

$$p \equiv \mathbb{E}^{(\widehat{\sigma}_\theta,\sigma_2)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t\mathbf{1}\{a_t=(T,H)\}\Big]+\mathbb{E}^{(\widehat{\sigma}_\theta,\sigma_2)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t\mathbf{1}\{a_t=(T,L)\}\Big].$$

If type θ_1 plays according to $\widehat{\sigma}_\theta$, her payoff is $p(1-\gamma\theta_1)$. According to Theorem 1,

$$p(1-\gamma\theta_1) \leq 1-\theta_1. \quad \text{(C.6)}$$

If type θ plays according to $\widehat{\sigma}_\theta$, her payoff is her equilibrium payoff, equal to $p(1-\gamma\theta)$. The equilibrium payoff is within ε of v^* implies that:

$$p(1-\gamma\theta) \geq \frac{1-\theta_1}{1-\gamma^*\theta_1}(1-\gamma^*\theta)-\varepsilon. \quad \text{(C.7)}$$

Inequalities (C.6) and (C.7) together imply that:

$$\varepsilon > (1-\theta_1)\Big\{\frac{1-\gamma^*\theta}{1-\gamma^*\theta_1}-\frac{1-\gamma\theta}{1-\gamma\theta_1}\Big\}. \quad \text{(C.8)}$$

The RHS is strictly positive as $\gamma>\gamma^*$ and $\theta>\theta_1$. As a result, inequality (C.8) cannot hold for ε smaller than the RHS. For every $\gamma>\gamma^*$, take ε to be smaller than

$$\min_{\theta\neq\theta_1}(1-\theta_1)\Big\{\frac{1-\gamma^*\theta}{1-\gamma^*\theta_1}-\frac{1-\gamma\theta}{1-\gamma\theta_1}\Big\},$$

we obtain a contradiction. This establishes the upper bound on the relative frequencies.

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 (1965) *Repeated Games with Incomplete Information*, MIT Press.

[3] Banerjee, Abhijit and Esther Duflo (2000) “Reputation Effects and the Limits of Contracting: A Study of the Indian Software Industry,” *Quarterly Journal of Economics*, 115(3), 989-1017.

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

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

[6] Board, Simon and Moritz Meyer-ter-Vehn (2013) “Reputation for Quality,” *Econometrica*, 81(6), 2381-2462.

[7] Bohren, Aislinn (2018) “Using Persistence to Generate Incentives in a Dynamic Moral Hazard Problem,” *Theoretical Economics*, forthcoming.

[8] Bower, Anthony, Steven Garber and Joel Watson (1997) “Learning About a Population of Agents and the Evolution of Trust and Cooperation,” *International Journal of Industrial Organization*, 15(2), 165-190.

[9] Caruana, Guillermo and Liran Einav (2008) “A Theory of Endogenous Commitment,” *Review of Economic Studies*, 75(1), 99-116.

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

[11] 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.

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

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

[14] Dilmé, Francesc (2018) “Reputation Building through Costly Adjustment,” *Journal of Economic Theory*, forthcoming.

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

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

[17] 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.

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

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

[20] Fudenberg, Drew and Eric Maskin (1991) "On the Dispensability of Public Randomization in Discounted Repeated Games," *Journal of Economic Theory*, 53(2), 428-438.

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

[22] Gul, Faruk, Hugo Sonnenschein and Robert Wilson (1986) "Foundations of Dynamic Monopoly and the Coase Conjecture," *Journal of Economic Theory*, 39, 155-190.

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

[24] Hörner, Johannes, Stefano Lovo and Tristan Tomala (2011) "Belief-free Equilibria in Games with Incomplete Information: Characterization and Existence," *Journal of Economic Theory*, 146, 1770-1795.

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

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

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

[28] Liu, Shuo and Harry Pei (2017) "Monotone Equilibria in Signalling Games," Working Paper.

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

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

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

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

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

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

[35] 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.

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

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