Trust and Betrayals:
Reputational Payoffs and Behaviors without Commitment

Harry PEI*

October 19, 2019

Abstract: I introduce a reputation model in which all types of the reputation-building player are facing lack-of-commitment problems. I study a repeated *trust game* in which a patient player (e.g., a seller) wants to win the trust of some myopic opponents (e.g., buyers) but can strictly benefit from betraying them. Her benefit from betrayal is her persistent private information. I provide a tractable formula for the highest equilibrium payoff for every type of that patient player. 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 nontrivially on past play. I establish bounds on her long-run action frequencies that apply to all of her equilibrium best replies. These features of her behavior are essential for her to extract information rent while preserving her informational advantage. I construct a class of such high-payoff equilibria in which the patient player's reputation depends only on the number of times she has betrayed as well as the number of times she has been trustworthy in the past. The method I developed is useful for studying other repeated incomplete information games between a patient informed player and a sequence of myopic uninformed players.

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 illustrate this, consider the example of a supplier who promises his clients to make 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 of their high quality standards. But after receiving upfront payments, they may be tempted to undercut quality, especially on aspects that are hard to verify. Similar situations also occur when incumbent firms threaten to fight potential entrants but then accommodate them after

*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 custom software industry, so that suppliers who wish to build their reputations have a strong motivation to minimize or avoid them. See Dellarocas (2006) and Bar-Isaac and Tadelis (2008) for more examples of lack-of-commitment problems in business transactions.

entry happens, when politicians seek support from their electorate but later renege on their campaign promises, and when entrepreneurs raise funding for their projects but then divert those funds for other purposes.

The common theme in these examples is a *lack-of-commitment problem* faced by the firms, politicians, and entrepreneurs. As a response, these agents build reputations for being trustworthy, from which they derive benefits in the future. In practice, a key challenge to reputation building is that all agents face temptations to renege, including even 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 or not they are facing temptations. For example, all firms can save costs by undercutting quality, but their costs can be different due to different production technologies. This challenge is not addressed in the classic theories of Sobel (1985), Fudenberg and Levine (1989,1992), and Benabou and Laroque (1992), in which agents of some types are committed to playing prespecified strategies, and others can establish their reputations by imitating those committed 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 an infinite time horizon between a patient long-run player (e.g., a seller) and a sequence of myopic short-run players (e.g., buyers).[2] In every period, the long-run player wants 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 exerting 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 although 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 (mixed) *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.[3] I establish tight bounds on her action frequencies that apply not only to all of her equilibrium strategies, but also to all of her equilibrium best replies. This merit of my result is meaningful from the perspective of applications, since one can estimate the parameters of the model (e.g., the lowest cost in the support of the short-run players' belief) by observing a realized path of play instead of the entire distribution over plays. I construct a class of those high-payoff equilibria in which the long-run player's reputation depends only 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 and Yelp (Dellarocas 2006, Dai et al. 2018), in which a seller's score depends only on the number of times she

[2] The assumption about myopia is motivated by the applications, such as in durable good markets in which each buyer has unit demand and online platforms (such as Airbnb, Uber, Lyft), where buyers are unlikely to meet with the same seller twice. Relaxing this assumption does not affect the attainability of high payoffs. Nevertheless, it can expand the long-run player's equilibrium payoff set.

[3] 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.

has received each rating, instead of on other more complicated metrics.[4]

My first result (Theorem 1) characterizes the highest equilibrium payoff for every type of the patient long-run player. It unifies the incomplete information setting with the complete information benchmark in Fudenberg, et al. (1990). Every type's highest equilibrium payoff admits a tractable formula, 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 depends only on the *lowest cost* in the support of the short-run players' prior belief. Moreover, the multiplier coincides with the maximal probability attached to the Stackelberg outcome such that the payoff of the lowest-cost type does not exceed her highest payoff in the repeated complete information game where her cost is common knowledge. My formula implies that every type except the lowest-cost type 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 or 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 nontrivial incentives and chooses actions based on the history of play. It also contrasts with the stationary commitment strategies in the reputation literature that prescribe the same action at every history.

To further explore these nonstationary 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 for every one of her pure-strategy best replies, the relative frequency of high and low effort cannot fall below the ratio between their probabilities in the Stackelberg action (or *critical ratio*). Similarly, for every type that does not have the lowest cost, the relative frequency of high and low effort cannot exceed the above critical ratio under each of her pure-strategy best replies. 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 apply only to her equilibrium strategies. This distinction is economically important since the resulting predictions can be tested by observing a realized path of the long-run player's actions, instead of the entire distribution.

Intuitively, the properties of behavior identified in Theorems 2 and 3 allow the high-cost types to shirk occasionally without losing too much reputation, while discouraging them from shirking too frequently in order to provide the short-run players with the incentive to trust. To achieve the first objective, the rational reputational

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

type needs to have an incentive 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 learning results in Fudenberg and Levine (1992), the short-run players can predict the long-run player's future actions with high precision in all periods except for a bounded number of them.

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 if she has a completely mixed 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 will believe that low effort will occur with a 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 with 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 each of her equilibrium strategies but also under each of her pure-strategy best replies.

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 (that is, shirking during which she wins her opponent's trust) inevitably reveals information about the long-run player's type, which undermines her informational advantage as well as her ability to extract information rent in the future. This tension grows as the long-run player becomes more patient, since she needs to extract information rent in an increasing number of periods to obtain a discounted average payoff strictly above her highest equilibrium payoff in the repeated complete information game.

My main technical contribution is to overcome the above challenge while taking the reputational type's incentives into account. The equilibria I construct exhibit *slow learning* and *reputation cycles* that *end in finite time*. Play starts from an *active learning phase*, in which (1) the short-run players trust, and (2) the long-run player's reputation improves after high effort and deteriorates after low effort. Every high-cost type plays a nontrivially mixed action unless her reputation (which is the probability of her being the lowest-cost type) is sufficiently close to one, at which point she shirks for one period and extracts information rent. The above arrangement that high-cost type shirks for sure only at histories where low-cost type occurs with sufficiently high probability reduces the magnitude of her reputation loss in the process of extracting information 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. Play transits to one of the two *absorbing phases* either when the long-run player has extracted enough information rent by shirking too much in the past, or when her reputation becomes perfect by exerting too much effort in the past. One of these events must happen in finite time. Intuitively, the first absorbing phase discourages the long-run player from shirking too frequently. The second absorbing phase discourages the high-cost types from exerting high effort too frequently, which is important to ensure that the high-cost types' continuation payoffs are low enough so that they can be attained in some equilibria of the continuation game.

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 for studying 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 (call these types *commitment types*). These papers show that every type of the strategic long-run player can attain her commitment payoff when her discount factor approaches 1 (call this *attainability*). This includes commitment payoffs that are not attainable in the repeated complete information game. 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 (call this *refinement*).

I take a complementary approach in which all types of the long-run player are rational, have reasonable stage-game payoff functions (that is, they share the same ordinal preferences over stage-game outcomes), and can flexibly choose their strategies in order to maximize their payoffs.

Compared to the canonical approach, my approach has the advantage of raising and answering novel questions related to the long-run player's *behavior*, such as (1) how the reputational types behave when they are also rational, and (2) how the other types behave to take advantage of this rational reputational type. The canonical approach is not well-suited to answer these questions since the long-run player's behaviors are sensitive to the commitment types' strategies, and those commitment behaviors are usually viewed as modeling shortcuts to attain reputational payoffs instead of capturing realistic aspects of economic agents' behaviors.[5] Moreover, there are many commitment behaviors that lead to the same commitment payoff, and the modeling specification of commitment types does not respond to changes in the payoff environment. This includes, but not limited to, changes in the long-run player's discount factor and her initial reputation.

My approach addresses these concerns by treating the reputational type as a strategic player whose behavior

[5]Weinstein and Yildiz (2016) point out the sensitivity of the game's predictions to the specifications of the commitment types' strategies. Their findings call for a careful selection of types in setting up incomplete information game models, including types that occur with very low probability. My approach follows this spirit by requiring all types to have *reasonable* payoff functions.

responds to changes in the payoff environment.[6] Focusing on trust games, which are classic examples to demonstrate the value of reputations, I characterize the patient long-run player's highest equilibrium payoff in this repeated incomplete information game. By exploiting the incentives of all types of the long-run player, I derive properties of her behaviors that are true for all high-payoff equilibria. In terms of complementing the canonical reputation models, 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, it helps to refine the commitment strategies when commitment types are simplifications for rational types whose temptation to deviate is low or zero. My 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 goods.

The main drawback of my approach is that it only leads to the attainability part of the reputational payoff result, but not the refinement part. This is a loss in the simultaneous-move trust game since it cannot rule out equilibria with low payoffs but the commitment type approach can, but not in the sequential-move trust game as neither approach can rule out low-payoff equilibria and the only role of reputation is to attain higher payoffs.

My work is also related to the study of repeated incomplete information games. Instead of restricting attention to zero-sum games (as in Aumann and Maschler 1965) or games with equally patient long-run players (as in Hart 1985, Cripps and Thomas 2003, and Pęski 2014), I study non-zero-sum games with unequal discount factors. This class of games is important for understanding dynamic interactions like those between firms and consumers, or between governments and citizens. In these scenarios, it is common to see gains from cooperation, as well as a lack of intertemporal incentives for one class of players, such as the consumers or the citizens.

I develop a tractable method to construct high-payoff equilibria in models with *multiple strategic types* and *no commitment type*. It generalizes to other repeated incomplete information games with monotone-supermodular payoffs, including those with interdependent values (Pei 2018). In these settings with short-run players, the belief-free equilibrium approach in Hörner, Lovo, and Tomala (2011) is not applicable. This is because at every history in which some types of the long-run player can extract information rent, the short-run player's best reply depends on her belief about the long-run player's type.

The class of equilibria I construct start from active learning phases followed by absorbing phases. Different from the learning phases in Wiseman (2005, 2012), during which players experiment and learn about their payoffs, the role of the learning phase in my model is to allow the informed player to extract information rent. As a result, the length of the learning phase in my model increases with the discount factor.

To make things tractable when learning and rent extraction persist in the long run, the informed player's equilibrium reputation can be computed via the number of times he has played the high action and the low action.

[6]This differs from other approaches that endogenize commitment in dynamic games, which include introducing switching costs (Caruana and Einav 2008), and perturbing players' belief hierarchies and introducing interdependent values (Weinstein and Yildiz 2016).

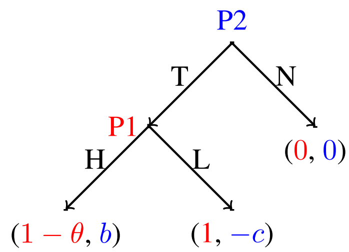

Figure 1: The stage game, where $\theta \in (0, 1)$, $b > 0$, $c > 0$

Using this tractable property, I show a result that verifies the attainability of continuation payoffs which takes into account the changes in the short-run players' beliefs (Lemma A.1). It overcomes the following difficulty of applying the self-generation argument in Abreu, Pearce, and Stacchetti (1990) to repeated games with incomplete information, that is, the set of equilibrium payoffs depends on the uninformed players' beliefs about the informed player's type.

Those equilibria also exhibit interesting reputation dynamics that are different from the behavioral patterns in models with commitment types such as those in Sobel (1985), Barro (1986), Schmidt (1993), Phelan (2006), Ekmekci (2011), Liu (2011), Jehiel and Samuelson (2012), and Liu and Skrzypacz (2014). These differences in dynamic reputation building and reputation milking behaviors highlight the role played by rational reputational types in shaping players' incentives. I will explain these differences in behavior in subsection 4.3.

2 The Baseline Model

I introduce a repeated *trust game* that captures the lack-of-commitment problem in many situations. Different from the canonical reputation models with commitment types, all types of the reputation-building player are rational and have qualitatively similar payoff functions, that is, they share the same ordinal preferences over stage-game outcomes. My approach is motivated by realistic concerns that no agent is immune to reneging temptations. It helps to answer novel questions, such as (1) how the rational reputational type behaves in equilibrium, and (2) how the other types behave to take advantage of this rational reputational type.

Stage Game: Consider a *trust game* described in Figure 1 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).[7] If the seller chooses L, then her payoff is 1 and the buyer's payoff is $-c$.

[7] Although 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 simultaneous-move product choice games in Mailath and Samuelson (2001), Ekmekci (2011) and Liu (2011). My results also extend to stage games with imperfect monitoring. See Section 5 for more details.

If the 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, Ekmekci 2011), but also in fiscal and monetary policies (Barro 1986, Phelan 2006) and political economy (Tirole 1996). The rest of this article sets up a repeated version of this game and explores (1) the extent to which the seller can overcome this lack-of-commitment problem by building reputations, and (2) different types of the seller's *payoffs* as well as *behaviors* in those reputational equilibria.

Repeated Game: Time is discrete, indexed by $t=0,1,2,....$ A long-lived seller is interacting with an infinite sequence of buyers, arriving one in each period and each buyer plays the trust game only once.

Both b and c are common knowledge. The cost of high effort, θ, is the seller's private information. I assume that θ is perfectly persistent over time. The buyers have a full support prior belief $\pi_0 \in \Delta(\Theta)$. Outcomes of all past interactions are perfectly observed. Let $y_t \in \{N,H,L\}$ be the stage-game outcome in period t. Let $h^t=\{y_s\}_{s=0}^{t-1} \in \mathcal{H}^t$ be a public history with $\mathcal{H} \equiv \bigcup_{t=0}^{+\infty} \mathcal{H}^t$ the set of public histories.[8] Let $A_1 \equiv \{H,L\}$

[8] In the simultaneous-move stage game, my results remain robust in the 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.

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 $\sigma_\theta : \mathcal{H} \to \Delta(A_1)$ is type θ's strategy that specifies her action choice at each public history after the buyer chooses T. A strategy σ_θ is *stationary* if it takes the same value for all $h^t \in \mathcal{H}$.

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

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

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 equilibrium behaviors of types that have low or even zero cost contrast with the stationary commitment strategies in the reputation literature that prescribe the same action at every history.

3.1 Highest Equilibrium Payoff

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. To ensure the robustness of my characterization against the choice of solution concepts and the ways of taking the limit, I introduce a *lower bound* and an *upper bound* of the seller's equilibrium payoff set, both of which are in the set inclusion sense. Formally, let $\underline{V}(\pi_0, \delta) \subset \mathbb{R}^m$ be the set of *sequential equilibrium* payoffs under parameter configuration (π_0, δ).[9] Let clo$(\cdot)$ be the closure of a set and let:[10]

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

Similarly, let $\overline{V}(\pi_0, \delta) \subset \mathbb{R}^m$ be the set of *Nash equilibrium* payoffs under (π_0, δ). Let

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

[9] I adopt the notion of sequential equilibrium introduced by Pęski (2014, page 658) for this infinitely repeated game.

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

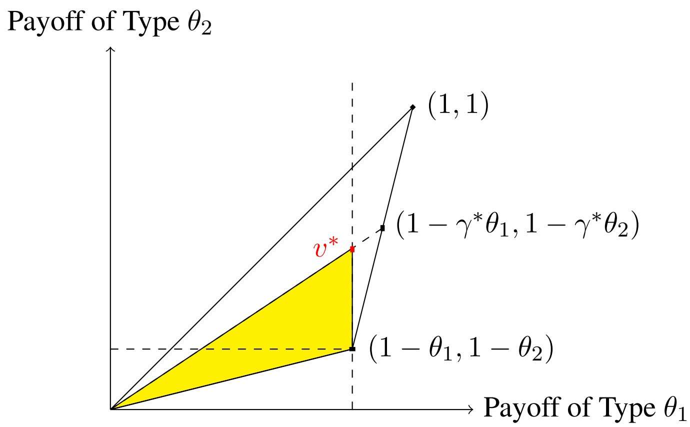

Figure 2: When $m = 2$, player 1's highest equilibrium payoff v^* in red and her equilibrium payoff set in yellow.

By definition, $\underline{V}(\pi_0) \subset \overline{V}(\pi_0)$. For every $j \in \{1, 2, ..., m\}$, let

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

and let $v^* \equiv (v_1^*, ..., v_m^*)$. Theorem 1 characterizes every type of the patient seller's highest equilibrium payoff, which unifies this incomplete information setting with the complete information benchmark:

Theorem 1. *If π_0 has full support, then*

- $v^* \in \underline{V}(\pi_0)$
- $v_j \leq v_j^*$ *for every* $v \equiv (v_1, v_2, ..., v_m) \in \overline{V}(\pi_0)$ *and* $j \in \{1, 2, ..., m\}$.

The proof is in Appendices A and B with an intuitive explanation in Section 4. I depict v^* in Figure 2. According to Theorem 1, v_j^* is type θ_j seller's highest equilibrium payoff when she is patient, and moreover, the highest equilibrium payoff for all types can be attained simultaneously in some sequential equilibria. Since other notions of the patient seller's limiting equilibrium payoff set contain $\underline{V}(\pi_0)$ and are contained in $\overline{V}(\pi_0)$, v^* provides a robust description of the seller's highest equilibrium payoff when she is patient.

Equation (3.3) provides a tractable formula for every type's highest equilibrium payoff, which is the product of her *Stackelberg payoff* and an *incomplete-information multiplier*. This multiplier is strictly below 1, common for all types, and is a sufficient statistic for the effects of incomplete information. Moreover, it depends only 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. Since the lowest-cost type is the best one available for the other types to imitate, the presence of other costs has no significant impact either.[11]

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 $1-\theta_1$. The latter is type θ_1's highest equilibrium payoff under complete information. To understand this observation, I explain the two linear constraints that pin down v^*. The first constraint is: the lowest-cost type (or type θ_1) cannot receive payoff strictly higher than $1-\theta_1$. Intuitively, recall the result in Fudenberg, Kreps, and Maskin (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, note that 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 $1-\theta$.[12] 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), who shows 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 play N 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 type 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 since a patient seller needs to extract information rent (that is, playing L for sure at histories where the buyer plays T with positive probability) in unbounded number of periods in order to get a payoff strictly greater than $1-\theta_j$. However, extracting information rent inevitably reveals information about the seller's type, which can undermine her ability to extract information rent in the future.

Second, the multiplier converges to 1 as θ_1 vanishes to 0. That is, every type's highest equilibrium payoff converges to her Stackelberg payoff. This suggests that although no type is immune to reneging temptations, every type of the long-run player can approximately attain her optimal commitment payoff when she is patient. This is stated as Corollary 1, the proof of which is implied by that of Proposition 4.1.

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 $j \in \{1, ..., m\}$.*

[11] This is similar to the Coasian bargaining model of Gul et al. (1986) that the informed player's information rent depends only on the best type she can imitate. Nevertheless, both insights rely on a private value assumption that θ does not affect the buyers' payoffs.

[12] This argument breaks down in repeated *incomplete information* games. This is because the short-run player's incentive to trust at h^t does not imply that any particular type of the long-run player exerts high effort with positive probability at h^t. As a result, some types of the long-run player may obtain strictly higher payoffs.

Third, 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 called 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 2. *For every $N \in \mathbb{N}$ and $v = (v_1, ..., v_m)$ 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 with nontrivial incomplete information, that is, $m \geq 2$. I derive properties of the patient seller's behavior that are true for *all* Nash equilibria in which she approximately attains her highest equilibrium payoff v^*.

As a first step, I show that in the repeated game with incomplete information, no type of the patient player uses stationary strategies or has completely mixed best replies, no matter how low her cost is. This contrasts with 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^*.*[13] *Moreover, no type has a completely mixed equilibrium best reply.*[14]

The proof, which is in Online Appendix C, exploits the interplay between supermodularity and incomplete information. For an informal illustration, 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. Since low-cost types enjoy comparative advantages in playing H and vice versa, every type with cost higher than θ_j plays L with probability 1 at every on-path history, and every type with cost lower than θ_j plays H with probability 1 at every on-path history.[15] I will argue next that 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 result in Fudenberg and Levine (1989), if the short-run players face type θ_k, then they will eventually believe that

[13]Strategy $\sigma_\theta : \mathcal{H} \to \Delta(A_1)$ is *completely mixed*, if $\sigma_\theta(h^t)$ is a completely mixed action for every h^t that (1) occurs with positive probability under the probability measure induced by $(\sigma_\theta, \sigma_2)$, and (2) T is played with positive probability at h^t according to σ_2.

[14]The conclusion in Theorem 2 fails when $m = 1$. See Online Appendix E.1 for a counterexample.

[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 (2018). 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 their result on repeated signalling games (Pei 2018).

L is played with probability greater than $1-\gamma^*$ in all future periods, after which they will have a strict incentive to play N. This leaves type θ_k with a discounted average payoff close to 0 when δ is close to 1.

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 (Proposition B.1), in which case her discounted average payoff cannot exceed $(1-\delta)+\delta(1-\theta_j)$. The latter converges to $1-\theta_j$ as $\delta \to 1$. Since $1-\theta_j < v_j^*$, it contradicts the hypothesis that type θ_j's equilibrium payoff is close to v_j^*.

The above argument also suggests that the conclusion of Theorem 2 remains valid for a type whose cost of high effort is zero. This is somewhat surprising since the seller's best reply cannot be completely mixed although she is indifferent between high and low effort in the stage game. It also implies that every type (including the zero cost one) faces *nontrivial intertemporal incentives* in all high-payoff equilibria. In another word, some of the seller's stage-game action choices have a nontrivial impact on the frequencies of trust in the future.

To further explore the properties of these nonstationary behaviors, I establish bounds on the seller'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 for every pure-strategy best reply $\widehat{\sigma}_\theta$ of type θ against σ_2,*

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

2. *For every $\theta \neq \theta_1$, and for every pure-strategy best reply $\widehat{\sigma}_\theta$ of type θ against σ_2,*[16]

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

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 replies cannot be lower than $\gamma^*/(1-\gamma^*)$. Conversely, when her cost is not the lowest one in the

[16] Although playing H at every history can never be type θ_1's *equilibrium strategy* in any high-payoff equilibrium, it can be her *equilibrium best reply* in some high-payoff equilibria. Examples of this include the constructed equilibria in the proof of Theorem 1.

support, this relative frequency under each of her pure-strategy best replies cannot be greater than $\gamma^*/(1-\gamma^*)$. For every $j \in \{2,3,...,m-1\}$, type θ_j long-run player's payoff is approximately v_j^* in every high-payoff equilibrium, which together with the above bounds on the relative frequencies between (T,H) and (T,L), *pin down* the frequencies of each stage-game outcome:

$$\mathbb{E}^{(\widehat{\sigma}_{\theta_j},\sigma_2)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t \mathbf{1}\{y_t=(T,H)\}\Big] \approx \gamma^* \frac{1-\theta_1}{1-\gamma^*\theta_1}, \tag{3.6}$$

$$\mathbb{E}^{(\widehat{\sigma}_{\theta_j},\sigma_2)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t \mathbf{1}\{y_t=(T,L)\}\Big] \approx (1-\gamma^*) \frac{1-\theta_1}{1-\gamma^*\theta_1} \tag{3.7}$$

and

$$\mathbb{E}^{(\widehat{\sigma}_{\theta_j},\sigma_2)}\Big[\sum_{t=0}^{\infty}(1-\delta)\delta^t \mathbf{1}\{y_t=N\}\Big] \approx \frac{(1-\gamma^*)\theta_1}{1-\gamma^*\theta_1}. \tag{3.8}$$

Next, I explain the differences between the bounds in Theorem 3 and the following bound implied by Fudenberg and Levine (1992)'s result, which is, under every type's *equilibrium strategy*, the relative discounted average frequency between outcome (T,H) and outcome (T,L) is no less than $\gamma^*/(1-\gamma^*)$. In contrast, the lower and upper bounds in Theorem 3 apply to *all pure strategies* in the support of any given type's equilibrium strategy, 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 (such as playing L at every history), the frequency of L exceeds the bound in (3.4), and for others in its support (such as 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. For example, it sheds light on how different types of the long-run player cherry-pick actions based on the history of play, which provides guidance for computing those high-payoff equilibria. For the lowest-cost type, she has a strict incentive to play H when she has already played L with occupation measure greater than the RHS of (3.7). 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 frequencies of actions in the long run. The only difference between these best replies is in the timing of actions. As a result, these types have strict incentives to play H when the occupation measure of L exceeds the RHS of (3.7) and have strict incentives to play L when the occupation measure of H exceeds the RHS of (3.6).

Moreover, the bounds that apply to all equilibrium best replies have stronger testable implications compared to the ones that only apply to all equilibrium strategies. This distinction is economically important since in many markets, researchers can only observe a few *realized paths* of the seller's dynamic behavior instead of an entire distribution of her action paths. For example, since the RHS of (3.8) is strictly increasing in θ_1, a researcher can identify the position of the seller's cost in the cost distribution as well as bounds on θ_1 as long as he knows the value of γ^* and can observe the frequencies of outcomes under a realized path of the long-run player's actions.

4 Ideas Behind the Proof of Theorem 1

In subsection 4.1, I explain the necessity of the two constraints that characterize v^*. In subsection 4.2, I construct a class of equilibria that approximately attain v^* when the long-run player becomes patient. In subsection 4.3, I compare the reputation dynamics in those equilibria to the ones in other models of reputation building.

4.1 Necessity of Constraints

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 convex weight attached to (T, H) and the convex weight attached to (T, L) is no less than $\gamma^*/(1-\gamma^*)$ in the long-run player's equilibrium payoff.

To understand the necessity of the first constraint, it is instructive to focus on an environment with two types. Let $\sigma \equiv (\sigma_{\theta_1}, \sigma_{\theta_2}, \sigma_2)$ be a generic Nash equilibrium. Let $\mathcal{H}(\sigma)$ be the set of histories that occur with positive probability under σ. 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 long-run player's *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 $h^t \in \mathcal{H}(\sigma)$ only when $\overline{\sigma}_1(h^t) = H$.

I show that type θ_1's equilibrium payoff is no more than $1-\theta_1$. First, suppose $\overline{\sigma}_1$ is type θ_1's best reply to σ_2. Since type θ_1's payoff in every period cannot exceed $1-\theta_1$ when she plays according to $\overline{\sigma}_1$, her equilibrium payoff is at most $1-\theta_1$. Next, suppose following $\overline{\sigma}_1$ is optimal for type θ_1 until h^t. Then we know that first, $\theta_2 \in \Theta(h^t)$, and second, type θ_2's continuation payoff at h^t is no more than $1-\theta_2$. Since the difference between type θ_1's and type θ_2's stage-game payoffs is at most $\theta_2-\theta_1$, type θ_1's continuation payoff at h^t is no more than $1-\theta_1$. Moreover, by following $\overline{\sigma}_1$ until h^t, which is optimal for type θ_1, her stage-game payoffs from h^0 to h^t cannot exceed $1-\theta_1$. Therefore, her continuation payoff at h^0 cannot exceed $1-\theta_1$.

For the second constraint, I provide an alternative explanation based on the *relative rate* of reputation building versus reputation milking. The motivation is to offer a new perspective which is complementary to the one that is based on the canonical reputation logic (see subsection 3.1). For this purpose, I introduce an alternative version of the highest action path based on type θ_2's equilibrium strategy:

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

By construction, $\overline{\sigma}_{\theta_2}$ is type θ_2's best reply to σ_2. If type θ_2 plays according to $\overline{\sigma}_{\theta_2}$, then her stage-game payoff exceeds $1-\theta_2$ only at histories where 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 a 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, that is, $\eta(h^t, L) < \eta(h^t)$. But nevertheless, she may rebuild her reputation in periods where $\overline{\sigma}_{\theta_2}$ prescribes H, that is, $\eta(h^t, H) > \eta(h^t)$.

Next comes the key question: suppose play proceeds according to $(\overline{\sigma}_{\theta_2}, \sigma_2)$, what is the *maximal relative frequency* between L and H under which player 2 has an incentive to trust whenever σ_2 prescribes him to do so? The answer to this question depends on the speed with which player 1 rebuilds her reputation (by playing H), relative 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, namely,

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

The above inequality suggests that the long-run player needs to play H and L with relative frequency at least $\gamma^*/(1-\gamma^*)$ in order to ensure that (1) the short-run players' have incentives to trust, and (2) the long-run player's long-term reputation does not decline compared to the prior.

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:

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 $v(\gamma)$.*

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

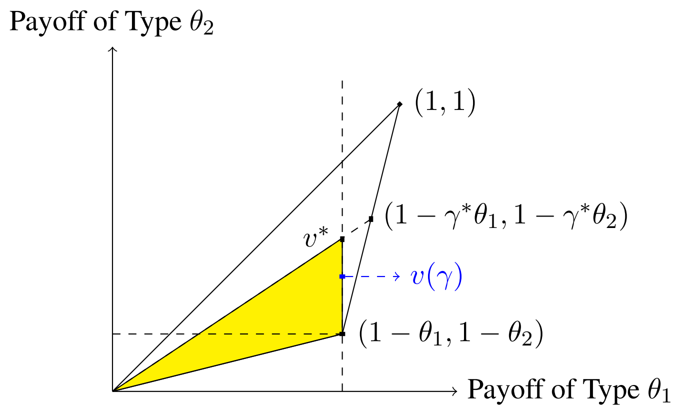

Figure 3: The set of attainable payoffs in yellow and $v(\gamma)$ in blue for some $\gamma \in (\gamma^*, 1)$.

overcomes the aforementioned challenges, such as (1) how to ensure that the promised continuation payoffs can be delivered in the continuation game, (2) how to provide incentives to the rational reputational type, and (3) how to let the high-cost types extract information rent while preserving their informational advantage.

Part I: Equilibrium Strategies The equilibrium consists of three phases. Play starts from an *active learning phase*, and transits to one of the two *absorbing phases* in finite time. Phase transitions happen either when the long-run player has played H too many times or has played L too many times. The high-cost types extract 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. 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 the following state variables: (1) the probability with which player 2's posterior attaches to type θ_1. I refer this as player 1's *reputation*, denoted by $\eta(h^t)$; and (2) the remaining weights of the three stage-game outcomes, 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 $\lambda > 0$.

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. When λ is small enough, it is sufficient to guarantee that player 1's long-term reputation is no less than her initial reputation, given that first, player 2s' beliefs are updated according to (4.5) and (4.6), and second, the relative frequencies with which player 1 plays H and L is at least $\gamma^*/(1-\gamma^*)$. This is because for every $\gamma > \gamma^*$, there exists $\overline{\lambda} > 0$, such that for every $\lambda \in (0, \overline{\lambda})$,

$$\big(1 - \lambda\gamma^*\big)^{1-\gamma}\big(1 + \lambda(1-\gamma^*)\big)^{\gamma} > 1. \tag{4.7}$$

The other three state variables evolve according to the realized stage-game outcomes:

$$p^y(h^t, y_t) \equiv \begin{cases} p^y(h^t) & \text{if } y_t \neq y \\ p^y(h^t) - (1-\delta)\delta^t & \text{if } y_t = y, \end{cases} \tag{4.8}$$

for every $y, y_t \in \{N, H, L\}$. Intuitively, player 1 has three accounts, one for each stage-game outcome. If outcome y is realized in period t, then her account for outcome y is deducted by $(1-\delta)\delta^t$. Therefore, the interpretation of $p^y(h^t)$ is the remaining credit in the account for outcome y.

Play transits to the first absorbing phase when $\eta(h^t)$ reaches 1, after which the continuation value of type θ_j is $v_1(h^t)\frac{1-\theta_j}{1-\theta_1}$ for every $j \in \{1, 2, ..., m\}$, with

$$v_1(h^t) \equiv \frac{p^H(h^t)(1-\theta_1) + p^L(h^t)}{p^H(h^t) + p^L(h^t) + p^N(h^t)}. \tag{4.9}$$

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 the payoff that has been promised to her by the end of the active learning phase. However, if another type that is not type θ_1 reaches this absorbing phase, then her continuation payoff is strictly less than what she has been promised before. This discourages her from doing so. Such an inequality is driven by the supermodularity of stage-game payoffs, namely, a lower cost type has a stronger preference for (T, H), and reaching the first absorbing phase leads to a higher frequency of trust at the expense of a higher relative frequency of (T, H) to (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. In the simple situation where $p^L(h^t) = 0$, type θ_j's continuation payoff is:

$$v_j(h^t) \equiv \frac{p^H(h^t)(1-\theta_j) + p^L(h^t)}{p^H(h^t) + p^L(h^t) + p^N(h^t)} \tag{4.10}$$

for every $j \in \{1, 2, ..., m\}$. The resulting payoff vector $\big(v_j(h^t)\big)_{j=1}^m$ can be delivered by a convex combination of outcomes N and (T, H). The more complicated case in which $p^L(h^t) \in \big(0, (1-\delta)\delta^t\big)$ requires the design of a *reshuffling stage*, the description of which is relegated to Appendix A.

Part II: Summary of Main Ideas To start with, I introduce *reputation cycles* so that the long-run player can rebuild her reputation after extracting information rent. This idea is incorporated in the design of the active learning phase. In particular, the long-run player can flexibly choose when to build and milk reputations, and her equilibrium reputation depends only on the number of times she has played each action in the past. Since extracting information rent requires learning, the high-cost and low-cost types need to mix with different probabilities. Therefore, one needs to provide them the incentive to mix at each of those histories.

These incentive issues are addressed by designing transition rules to the absorbing phases. First, if the long-run player shirks too frequently by front-loading 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 rents in the future. Second, in order to deliver the high-cost types their promised continuation payoffs, one also needs to deter them from front-loading the play of H. This is because type θ's continuation value *increases* whenever she plays H at histories where her continuation value is strictly greater than $1 - \theta$. If her equilibrium payoff exceeds $1 - \theta$ and she plays H for many consecutive periods, her continuation payoff will become so high that it cannot be delivered in any equilibrium of the continuation game.

The first absorbing phase solves this problem by introducing an upper bound on the extent to which the high-cost types can front-load the play of H. It screens out those high-cost types by promising a higher absolute frequency of trust as well as a higher relative frequency between H and L compared to the active learning phase. It ensures that at every history of the active learning phase, if the relative weight of L and H in the high-cost types' continuation payoffs exceeds its prior value (which is $\frac{1-\gamma}{\gamma}$), then the short-run player's posterior belief attaches higher probability to the lowest-cost type compared to his prior (Lemma A.1). Since the constraint on front-loading the play of H is tighter when the long-run player's initial reputation is higher, one can use the self-generation argument similar to the one in Abreu, Pearce and Stacchetti (1990) to conclude that there exists an equilibrium in the continuation game that can deliver the continuation payoff promised to the long-run player.

Given the above building blocks, what remains to be designed is the speed of learning in the active learning phase. 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. To understand why, 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. 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 can increase the patient player's payoff.

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 with gradual learning (Benabou and Laroque 1992, Ekmekci 2011, Wiseman 2005,2012) and the capital-theoretic models of reputations (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* in the reputation model of Jehiel and Samuelson (2012). In their setting, there are multiple commitment types of the long-run player who are playing stationary mixed strategies, and one strategic type who can flexibly choose her actions. The short-run 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* in which she plays H 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 short-run 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 short-run 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 short-run players' ability to correctly predict the long-run player's average action in *every period*. This constraint is absent when short-run players use analogy-based reasoning since they can only correctly predict the average action *across all periods*. Second, the short-run players learn the true state with positive probability in every high-payoff equilibrium of my model, while in Jehiel and Samuelson (2012), the probability with which they learn the true state is zero. This is because analogy-based short-run players' posterior beliefs depend only on the empirical frequencies of

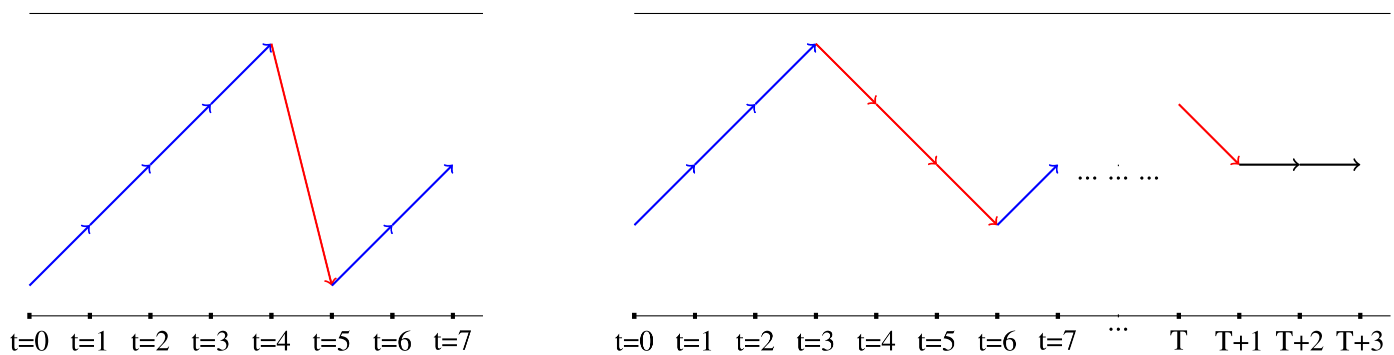

Figure 4: The horizontal axis represents the timeline and the vertical axis measures the informed player's reputation, that is, the probability of the commitment type or lowest-cost type. Left: Reputation cycles in Phelan (2006). Right: A sample path of the reputation cycle in my model.

the observed actions. That is to say, their beliefs are not responsive enough to each individual observation.

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. As a result, one misbehavior reveals the long-run player's rationality.[17] In my model, the comparison between good and bad types is not that stark since 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. These differences are depicted in Figure 4.

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 have obtained in the past. In particular, one recent negative rating neither

[17] Similarly in Ghosh and Ray (1996), players are matched to play repeated prisoner's dilemma, and each player has private information about whether she is patient or myopic. In their renegotiation-proof equilibria, cheating immediately reveals a player's myopia.

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 reputation games with commitment types and the long-run player's actions are imperfectly monitored. In their equilibria, learning also happens gradually since 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 since the reputational type cheats with positive probability. The different driving forces behind gradual learning also lead to different long-run 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 has a strict incentive to betray. This together with the short-run players' incentive constraints lead to an upper bound on the reputational type's equilibrium payoff. Another difference is that the short-run players never fully learn the state in the models of Benabou and Laroque (1992) and Ekmekci (2011), while in mine, the lowest-cost type fully reveals her private information at unbounded number of on-path histories in every high-payoff equilibrium.

Capital-Theoretic Models of Reputation: Similar reputation cycles also occur in the Poisson good news models of Board and Meyer-ter-Vehn (2013) and Dilmé (2018).[18] 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 depends only 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 sources 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 arrival.

5 Extensions

I examine two variants of the trust game in which the insights from my baseline model remain robust. I start from simultaneous-move stage games and then move on to games with imperfect monitoring.

Simultaneous-Move Stage Game: Consider a simultaneous-move trust game with stage-game payoffs given by:

[18] 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. These differ from my model in which the high-cost types cheat for sure when reputation is close to being perfect.

-	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 both 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.

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 to the simultaneous-move stage game. In particular, no type of the long-run player uses a stationary strategy or 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 $y_t = (T, H)$ and $y_t = (T, L)$ 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 2018). Their result implies that in every Nash equilibrium of the repeated game, for every $i > j$, and compare any equilibrium best reply of type θ_i with any equilibrium best reply of type θ_j, the discounted average frequency of H under the former must be weakly lower compared to the discounted average frequency of H under the latter.

Stage Game with Imperfect Monitoring: Player 1 is an agent, for example a worker, a supplier or a private contractor. In every period, a principal (player 2, for example an employer or a final good producer) is randomly matched with the agent. The principal then decides whether to incur a fixed cost and interact with the agent or to skip the interaction.[19] The agent chooses her effort from a closed interval unbeknownst to the principal.

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

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.[20]

Players move sequentially in the stage game. Different from the baseline model, after player 2 chooses to trust, player 1 chooses among a continuum of effort levels $e \in [0, 1]$. The quality of the output being produced is denoted by $z \in \{G, B\}$, which is *good* (or $z = G$) with probability e and is *bad* (or $z = B$) with complementary probability. The cost of effort for type θ_i is $\theta_i e$. Player 1's benefit from her opponent's trust is normalized to 1. Therefore, her stage-game payoff under outcome N is 0 and that under outcome (T, e) is $1 - \theta_i e$. Player 2's payoff is 0 if he chooses N. His benefit from good output is b while his loss from bad output is c, with $b, c > 0$. Therefore, player 2 is willing to trust only when player 1's expected effort 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 output quality. In another word, player 1's effort choice is her private information. Formally, let $a_{1,t}$, $a_{2,t}$ and z_t be player 1's action, player 2's action and the realized output quality in period t, respectively. Let $h^t = \{a_{2,s}, z_s\}_{s=0}^{t-1} \in \mathcal{H}^t$ be a public history with $\mathcal{H} \equiv \bigcup_{t=0}^{+\infty} \mathcal{H}^t$ the set of public histories. Let $h_1^t = \{a_{1,s}, a_{2,s}, z_s\}_{s=0}^{t-1} \in \mathcal{H}_1^t$ 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 $y_t = (T, H)$ and $y_t = (T, L)$ in (3.4) and (3.5) with $(a_{2,t}, z_t) = (T, G)$ and $(a_{2,t}, z_t) = (T, B)$, respectively.

6 Conclusion

This paper introduces a new approach to studying reputations, in which all types of the long-run player are rational and have reasonable stage-game payoffs. Motivated by various applications of reputation effects, I require all types to share the same ordinal preferences over stage-game outcomes, that is, all types are facing strict lack-of-commitment problems. This captures, for example, the fact that most consumers are likely to understand that firms profit from their purchases, and that providing high-quality products and making timely

[20]Chassang (2010) studies a game in which players face similar incentives. The main difference is that the agent's cost of effort is common knowledge but the set of actions that are available in each period 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.

deliveries are costly for the firms. In contrast, a firm's cost of providing high-quality products and making timely deliveries tends to be its private information.

I characterize the highest equilibrium payoff for every type of the patient player in this repeated game with incomplete information. My formula identifies a sufficient statistic for the effects of incomplete information and helps to compute every type'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 her behavior that applies not only to all of her equilibrium strategies, but also to all of her equilibrium best replies.

The implications of my results on the behavior of the lowest-cost type can be used to choose among the many commitment strategies in the process of setting up reputation models. 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 theoretical studies of repeated games with incomplete information.

A Proof of Theorem 1: Sufficiency

In this Appendix, I provide a constructive proof to Proposition 4.1 Recall that the target payoff for type θ_j is:

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

and $v(\gamma) \equiv \big(v_j(\gamma)\big)_{1\leq j\leq m}$. 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 result that verifies the attainability of continuation payoffs, 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 $\pi_0(\theta_1)$.

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

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

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

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

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 stage-game outcomes (T, L), (T, 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. To simplify the exposition, the construction is performed when players have access to a public randomization device in the beginning of each period. Since actions can be perfectly monitored, the public randomization device can be dispensed according to the arguments in Fudenberg and Maskin (1991).

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 $p^y(h^t)$ for $y \in \{N, H, L\}$ such that

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

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

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

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 action 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 nontrivially 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 $h^T = (y_0, ..., y_{T-1})$, if play remains in the active learning phase for every $h^t \preceq h^T$, then*

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

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: we need 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 establish the necessity of the two linear constraints that characterize the highest equilibrium payoff vector 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 $d(\theta_i) \geq 0$ for every $\theta_i \in \Theta$ and $|\theta_i - \theta_j| \geq |d(\theta_i) - d(\theta_j)|$ for every $i < j$. In another word, the stage-game payoffs are monotone-supermodular according to Liu and Pei (2018). 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 show that type θ_1's payoff in any Bayesian Nash equilibrium is no more than $1 - \theta_1$. This holds not only when player 2 is myopic, but also applies whenever her discount factor δ_2 is strictly less than $\frac{c}{b+c}$. I establish the general version of this result by allowing for player 2 to be forward-looking, as well as player 2's prior belief attaches zero probability to some types of player 1.

Given strategy profile $\sigma \equiv \Big((\sigma_\theta)_{\theta\in\Theta}, \sigma_2\Big)$, let $\mathcal{H}^\sigma$ be the set of histories that occur with positive probability under σ. For every $h^t \in \mathcal{H}^\sigma$, let $\Theta^\sigma(h^t)$ be the set of types in support of player 2's posterior belief at h^t, which is nonempty and can be derived from Bayes Rule. I show the following result:

Proposition B.1. *Suppose $\delta_2 \in [0, \frac{c}{b+c})$. For every Bayesian Nash equilibrium σ and every $h^t \in \mathcal{H}^\sigma$, if θ_i is the smallest type in $\Theta^\sigma(h^t)$, then type θ_i's continuation payoff at h^t is no more than $1 - \theta_i$.*

Proof. It is sufficient to show that for *every* prior belief $\pi \in \Delta(\Theta)$, not necessarily have full support, the lowest type in the support of π, denoted by θ, receives payoff at most $1 - \theta$ at h^0. I show this by induction on $|\Theta^\sigma(h^0)|$, the number of types in the support of player 2's prior belief. The proof consists of three steps:

1. I show that when $\delta_2 < \frac{c}{b+c}$, if player 2 plays T with positive probability at an on-path history, then there exists some type in the support of player 2's posterior belief (at that history) who plays H with strictly positive probability at that history.

2. I show the result when $|\Theta^\sigma(h^0)| = 1$.[21]

[21] Notice that $|\Theta^\sigma(h^0)| = 1$ is *not* equivalent to the game has complete information. Given that the solution concept is Bayesian Nash Equilibrium, player 2's posterior belief at off-path histories can attach positive probability to multiple types.

3. Suppose the result holds when $|\Theta^\sigma(h^0)| \leq n$, I establish the conclusion for $|\Theta^\sigma(h^0)| = n+1$.

Step 1: Rank player 1's actions according to $H \succ L$. Given strategy profile σ and history $h^t \in \mathcal{H}^\sigma$, let

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

be the highest action played by player 1 with positive probability at h^t. I show the following lemma, which makes use of the upper bound on player 2's discount factor.

Lemma B.1. *Suppose $\delta_2 \in [0, \frac{c}{b+c})$. For every Bayesian Nash equilibrium σ and $h^t \in \mathcal{H}^\sigma$, $\sigma_2(h^t)$ assigns positive probability to T only if $\overline{a}_1^\sigma(h^t) = H$.*

Proof of Lemma: Suppose toward a contradiction that $\sigma_2(h^t)$ assigns positive probability to T but $\overline{a}_1^\sigma(h^t) = L$. Player 2's maximal payoff by playing T at h^t is:

$$\underbrace{(1-\delta_2)(-c)}_{\text{P2's stage-game payoff under } (T,L)} + \underbrace{\delta_2 b}_{\text{P2's maximal continuation payoff}},$$

which is strictly lower than 0, that she can obtain by playing N in all subsequent histories given that $\delta_2 < \frac{c}{b+c}$. This leads to a contradiction. □

Step 2: I show that when $|\Theta^\sigma(h^0)| = 1$, the only type in the support of player 2's prior belief, denoted by θ_i, receives payoff no more than $1-\theta_i$. This also implies that for every equilibrium σ and for every $h^t \in \mathcal{H}^\sigma$, if $|\Theta^\sigma(h^t)| = \{\theta_i\}$, then type θ_i's continuation payoff at h^t cannot exceed $1-\theta_i$.

This is because $\Theta^\sigma(h^0) = \{\theta_i\}$ implies that $|\Theta^\sigma(h^t)| = \{\theta_i\}$ for every $h^t \in \mathcal{H}^\sigma$. Therefore, $\overline{a}_1^\sigma(h^t)$ is played by type θ_i with positive probability at every $h^t \in \mathcal{H}^\sigma$. Given type θ_i's equilibrium strategy σ_{θ_i}, the following strategy $\widetilde{\sigma}_{\theta_i} : \mathcal{H} \to \Delta(A_1)$, defined as:

$$\widetilde{\sigma}_{\theta_i}(h^t) \equiv \begin{cases} \overline{a}_1^\sigma(h^t) & \text{if } h^t \in \mathcal{H}^\sigma \\ \sigma_{\theta_i}(h^t) & \text{otherwise.} \end{cases} \tag{B.2}$$

also best replies against player 2's equilibrium strategy σ_2, from which type θ_i receives his equilibrium payoff. If type θ_i plays according to $\widetilde{\sigma}_{\theta_i}$ against σ_2, then Lemma 1 implies that the outcome at every history in $\mathcal{H}^\sigma$ is either (T,H) or (N,H) or (N,L). Therefore, type θ_i's stage-game payoff at every history in $\mathcal{H}^\sigma$ cannot exceed $1-\theta_i$, so his discounted average payoff cannot exceed $1-\theta_i$.

Step 3: I show that if the result holds when $|\Theta^\sigma(h^0)| \leq n$, then it also holds when $|\Theta^\sigma(h^0)| = n+1$. I define $\overline{\mathcal{H}}_t^\sigma$ for every $t \in \mathbb{N}$ recursively. Let $\overline{\mathcal{H}}_0^\sigma \equiv \{h^0\}$. Given the definition of $\overline{\mathcal{H}}_t^\sigma$, let

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

and let $\overline{\mathcal{H}}^\sigma \equiv \bigcup_{t=0}^\infty \overline{\mathcal{H}}_t^\sigma$. Recall that θ_i is the smallest type in the support of player 2's belief. Given type θ_i's equilibrium strategy σ_{θ_i}, let $\widehat{\sigma}_{\theta_i} : \mathcal{H} \to \Delta(A_1)$ be defined as:

$$\widehat{\sigma}_{\theta_i}(h^t) \equiv \begin{cases} \overline{a}_1^\sigma(h^t) & \text{if } h^t \in \mathcal{H}^\sigma \text{ and } \overline{a}_1^\sigma(h^t) \in \text{supp}\big(\sigma_{\theta_i}(h^t)\big) \\ \sigma_{\theta_i}(h^t) & \text{otherwise.} \end{cases} \tag{B.3}$$

By construction, $\widehat{\sigma}_{\theta_i}$ is type θ_i's best reply against σ_2. Let $\mathcal{H}^{(\widehat{\sigma}_{\theta_i}, \sigma_2)}$ be the set of histories that occur with positive probability under $(\widehat{\sigma}_{\theta_i}, \sigma_2)$. Let

$$\overline{\mathcal{H}}^{\sigma,\theta_i} \equiv \Big\{ h^t \in \overline{\mathcal{H}}^\sigma \Big| \theta_i \in \Theta^\sigma(h^t) \text{ and } \overline{a}_1^\sigma(h^t) \notin \text{supp}\big(\sigma_{\theta_i}(h^t)\big)\Big\}.$$

Intuitively, $h^t \in \overline{\mathcal{H}}^{\sigma,\theta_i}$ if and only if first, at every $h^s \prec h^t$, type θ_i plays $\overline{a}_1^{\sigma}(h^s)$ with positive probability; and second, at h^t, type θ_i plays $\overline{a}_1^{\sigma}(h^t)$ with zero probability.

Consider type θ_i's payoff if he plays $\widehat{\sigma}_{\theta_i}$ and player 2 plays σ_2. For any given $h^t \in \mathcal{H}^{(\widehat{\sigma}_{\theta_i},\sigma_2)}$,

1. If there *does not exist* $h^s \preceq h^t$ such that $h^s \in \overline{\mathcal{H}}^{\sigma,\theta_i}$, then Lemma 1 implies that type θ_i's payoff at h^t and at all on-path histories preceding h^t is no more than $1-\theta_i$.

2. If there *exists* $h^s \preceq h^t$ such that $h^s \in \overline{\mathcal{H}}^{\sigma,\theta_i}$, then I show below that type θ_i's continuation payoff at h^s is no more than $1-\theta_i$.

   First, since $h^s \in \overline{\mathcal{H}}^{\sigma,\theta_i}$, after player 2 observes $\overline{a}_1^{\sigma}(h^s)$ at h^s, θ_i is no longer in the support of player 2's posterior belief. Therefore, for every $h^{s+1} \succ h^s$ with $\overline{a}_1^{\sigma}(h^s)$ being played at h^s, there exist at most n types in the support of player 2's posterior belief at h^{s+1}. Let θ_j be the lowest type in the support of this posterior belief. Since θ_i is the lowest type in the support of player 2's prior belief, we have $\theta_j > \theta_i$.

   By induction hypothesis, type θ_j's continuation payoff *after* playing $\overline{a}_1^{\sigma}(h^s)$ at h^s is no more than $1-\theta_j$. Moreover, according to Lemma 1, type θ_j's stage-game payoff by playing $\overline{a}_1^{\sigma}(h^s)$ at h^s is also no more than $1-\theta_j$. This implies that his continuation payoff at h^s is at most $1-\theta_j$.

   Therefore, type θ_j's continuation payoff by deviating to $\widehat{\sigma}_{\theta_i}$ starting from h^s is no more than $1-\theta_j$. Since $\theta_i < \theta_j$, and the maximal difference between type θ_i and θ_j's stage-game payoff is $\theta_j - \theta_i$, we know that type θ_i's continuation payoff at h^s under $\widehat{\sigma}_{\theta_i}$ is no more than $1-\theta_i$.

The two parts together imply that type θ_i's equilibrium payoff is no more than $1-\theta_i$ when $|\Theta^{\sigma}(h^0)| = n+1$.

□

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). Since $v_j > v_j(\gamma^*)$, we have:

$$\frac{X^{(\sigma_{\theta_j},\sigma_2)}}{Y^{(\sigma_{\theta_j},\sigma_2)}} < \frac{\gamma^*}{1-\gamma^*}. \tag{B.4}$$

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:

$$\mathbb{E}^{(\sigma_{\theta_j},\sigma_2)}\Big[\sum_{\tau=0}^{+\infty} d(\sigma_{\theta_j}(h^\tau)||\alpha_1(\cdot|h^\tau))\Big] \leq -\log \pi_0(\theta_j). \tag{B.5}$$

This implies that for every $\epsilon > 0$, the expected number of periods such that $d(\sigma_{\theta_j}(h^\tau)||\alpha_1(\cdot|h^\tau)) > \epsilon$ is no more than

$$T(\epsilon) \equiv \Big\lceil \frac{-\log \pi_0(\theta_j)}{\epsilon} \Big\rceil. \tag{B.6}$$

Let

$$\epsilon \equiv d\Big(\frac{\gamma + 2\gamma^*}{3} H + (1 - \frac{\gamma+2\gamma^*}{3})L \Big\| \gamma^* H + (1-\gamma^*)L\Big), \tag{B.7}$$

and let δ be large enough such that:

$$\frac{X^{(\sigma_{\theta_j},\sigma_2)}}{Y^{(\sigma_{\theta_j},\sigma_2)}-(1-\delta^{T(\epsilon)})}<\frac{2\gamma+\gamma^*}{3-2\gamma-\gamma^*}. \tag{B.8}$$

According to (B.5) and (B.6), 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.7), 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:

$$\underbrace{(1-\delta^{T(\epsilon)})}_{\text{periods in which player 2's prediction is wrong}}+\underbrace{\Big(X^{(\sigma_{\theta_j},\sigma_2)}+Y^{(\sigma_{\theta_j},\sigma_2)}-(1-\delta^{T(\epsilon)})\Big)\frac{2\gamma+\gamma^*}{\gamma+2\gamma^*}}_{\text{maximal frequency with which player 2 trusts after he learns}}, \tag{B.9}$$

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

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

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 against 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:

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

The long-run player's ex ante incentive constraints, namely, first, type θ_i prefers $\widehat{\sigma}_{\theta_i}$ to $\widehat{\sigma}_{\theta_m}$, and second, 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 further implies that according to type θ_m's equilibrium strategy σ_{θ_m},

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

or equivalently,

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

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 $\delta > \overline{\delta}$,

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

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

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 $\widehat{\sigma}_\theta$,

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

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

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

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 $\widehat{\sigma}_\theta$, she receives her equilibrium payoff, which is $p(1-\gamma\theta)$. 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 since $\gamma > \gamma^*$ 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] Bar-Isaac, Heski and Steven Tadelis (2008) "Seller Reputations," *Foundations and Trends in Micro*.

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

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

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

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

[9] 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.

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

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

[12] 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.

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

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

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

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

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

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

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

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

[25] 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.

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

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

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

[29] Liu, Shuo and Harry Pei (2018) "Monotone Equilibria in Signalling Games," Working Paper.

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

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

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

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

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

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

[36] 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.

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

[38] Wiseman, Thomas (2005) "A Partial Folk Theorem for Games with Unknown Payoff Distributions," *Econometrica*, 73(2), 629-645.

[39] Wiseman, Thomas (2012) "A Partial Folk Theorem for Games with Private Learning," *Theoretical Economics*, 7, 217-239.