

Equilibrium behaviors in repeated games [☆]

Yingkai Li ^a, Harry Pei ^{b,*}

^a Department of Computer Science, Northwestern University, United States of America

^b Department of Economics, Northwestern University, United States of America

Received 3 August 2020; final version received 7 February 2021; accepted 10 February 2021

Available online 12 February 2021

Abstract

We examine a patient player's behavior when he can build reputations in front of a sequence of myopic opponents. With positive probability, the patient player is a commitment type who plays his Stackelberg action in every period. We characterize the patient player's action frequencies in equilibrium. Our results clarify the extent to which reputations can refine the patient player's behavior and provide new insights to entry deterrence, business transactions, and capital taxation. Our proof makes a methodological contribution by establishing a new concentration inequality.

© 2021 Elsevier Inc. All rights reserved.

JEL classification: C72; C73; D82; D83

Keywords: Reputation; Action frequency; Behavior; Refinement; Concentration inequality

1. Introduction

Economists have long recognized that individuals, firms, and governments can benefit from good reputations. As shown in the seminal work of Fudenberg and Levine (1989), a patient player can guarantee himself a high payoff when his opponents believe that he might be committed to play a particular action. Their result can be viewed as a refinement, which selects the patient player's optimal equilibria in many games of interest.

[☆] We are grateful to Drew Fudenberg, Guillermo Ordoñez, Matthew Thomas, Linh To, and two anonymous referees for helpful comments. Pei acknowledges The National Science Foundation (Grant SES-1947021) for financial support.

^{*} Corresponding author.

E-mail addresses: yingkai.li@u.northwestern.edu (Y. Li), harrydp@northwestern.edu (H. Pei).

This paper studies the effects of reputations on the patient player's behavior instead of his payoffs, which have been underexplored in the reputation literature. Existing works on reputation-building behaviors restrict attention to particular equilibria or games with particular payoff functions. By contrast, we identify tight bounds on the patient player's action frequencies that apply to all equilibria under more general payoff functions. Our results clarify the extent to which reputations can refine the patient player's behavior and provide new insights to applications such as entry deterrence, business transactions, and capital taxation.

We analyze a repeated game between a patient player and a sequence of myopic opponents. The patient player is either a strategic type who maximizes his discounted average payoff, or a commitment type who plays his optimal pure commitment action (or *Stackelberg action*) in every period. The myopic players cannot observe the patient player's type, but can observe all the actions taken in the past.

We examine the extent to which the option to imitate the commitment type can motivate the patient player to play his Stackelberg action. Theorem 1 characterizes tight bounds on the discounted frequencies with which the strategic-type patient player plays his Stackelberg action in equilibrium. We show that the maximal frequency equals one and the minimal frequency equals the value of the following linear program: Choose a distribution over action profiles in order to minimize the probability of the Stackelberg action subject to two constraints. First, each action profile in the support of this distribution satisfies the myopic player's incentive constraint. Second, the patient player's expected payoff from this distribution is no less than his Stackelberg payoff. The first constraint is necessary since the myopic players best reply to the patient player's action in every period. The second constraint is necessary since the patient player can approximately attain his Stackelberg payoff by imitating the commitment type. In order to provide him an incentive not to play his Stackelberg action, his continuation value after separating from the commitment type must be at least his Stackelberg payoff.

The substantial part is to show that these constraints are not only necessary but also sufficient. Our proof is constructive and makes a methodological contribution by establishing a novel concentration inequality on the discounted sum of random variables that bounds the patient player's action frequencies (Lemma A.1).

Theorem 2 identifies a sufficient condition under which a distribution of the patient player's actions is his action frequency in some equilibria of the reputation game. In a number of leading applications such as the product choice game and the entry deterrence game, our sufficient condition is also necessary, in which case Theorem 2 fully characterizes of the set of action frequencies that can arise in equilibrium.

Our results provide new insights to classic applications of reputation models. For example, in the product choice game of Mailath and Samuelson (2006, Figure 15.1.1 on page 460),¹ our results imply that a policy maker can increase the frequency of high effort by subsidizing consumers for purchasing low-end products or by taxing consumers for purchasing high-end products. Intuitively, these policies increase the consumers' demand for high effort when they purchase the high-end product, which in turn increases the frequency of high effort in the worst

¹ In Mailath and Samuelson (2006)'s product choice game, a patient firm faces a sequence of consumers. In every period, the firm chooses between high effort and low effort, and a consumer chooses between buying a high-end product and a low-end product. The firm finds it costly to exert high effort and prefers the consumers to purchase high-end products. Each consumer has an incentive to buy the high-end product only when she believes that the firm will exert high effort with high enough probability. In this game, high effort is the firm's Stackelberg action but low effort is the dominant action in the stage game.

equilibrium. In the entry deterrence game of Kreps and Wilson (1982) and Milgrom and Roberts (1982a), our results imply that a small amount of subsidy to potential entrants for entering the market makes a reputation-building incumbent more aggressive in fighting entry, but a large amount of subsidy eliminates the incumbent's fighting incentives.

Our results contribute to the reputation literature by clarifying the role of reputations in refining the patient player's behavior. This is complementary to the result of Fudenberg and Levine (1989) that studies how reputations refine the patient player's payoff. Existing works on players' reputation-building behaviors restrict attention to particular equilibria or particular payoff functions. For example, Kreps and Wilson (1982) and Milgrom and Roberts (1982a) characterize sequential equilibria in entry deterrence games. Schmidt (1993) characterizes Markov equilibria in repeated bargaining games. Bar-Isaac (2003), Phelan (2006), Ekmekci (2011), Liu (2011), and Liu and Skrzypacz (2014) restrict attention to supermodular games or 2×2 games. By contrast, we characterize tight bounds on the patient player's action frequencies that apply to all equilibria. Our results are more general in terms of payoffs, which only require the patient player's optimal commitment payoff to be greater than his minmax value and that his optimal commitment outcome is not a stage-game Nash equilibrium.

Cripps et al. (2004) show that when the monitoring structure has full support, the myopic players eventually learn the patient player's type and the strategies converge to an equilibrium of the repeated complete information game. However, their results do not characterize the speed of convergence or players' behaviors in finite time, and hence do not imply what players' discounted action frequencies are. Ekmekci and Maestri (2019) study players' reputation-building behaviors in stopping games where a patient uninformed player chooses between continuing and irreversibly stopping the game in every period. By contrast, the uninformed players in our model are myopic and their action choices are reversible. Pei (2020a) provides sufficient conditions under which the patient player has a unique on-path behavior. Unlike our model that restricts attention to private value environments but allows for general stage-game payoffs, his result requires nontrivial interdependent values and monotone-supermodular stage-game payoffs.

Section 2 sets up the baseline model. Section 3 states our main results. Section 4 applies our results to several applied models of reputation formation and discusses the results' practical implications. Section 5 discusses our modeling assumptions as well as issues related to taking our predictions to the data. Section 6 concludes. The proofs of our results can be found in the appendix.

2. Model

Time is discrete, indexed by $t = 0, 1, 2, \dots$. A patient player 1 with discount factor $\delta \in (0, 1)$ interacts with an infinite sequence of myopic player 2s, arriving one in each period and each playing the game only once. In period t , a public randomization device $\xi_t \sim U[0, 1]$ is realized and is observed by both players, after which players simultaneously choose their actions. Player 1's action is denoted by $a_t \in A$. Player 2's action is denoted by $b_t \in B$. Their stage-game payoffs are $u_1(a_t, b_t)$ and $u_2(a_t, b_t)$. We assume A and B are finite, with $|A|, |B| \geq 2$.

Let $BR_1 : \Delta(B) \rightrightarrows 2^A \setminus \{\emptyset\}$ and $BR_2 : \Delta(A) \rightrightarrows 2^B \setminus \{\emptyset\}$ be player 1's and player 2's best reply correspondences in the stage-game. The set of player 1's (pure) Stackelberg actions is $\arg \max_{a \in A} \{\min_{b \in BR_2(a)} u_1(a, b)\}$.

Assumption 1. Player 1 has a unique Stackelberg action, denoted by a^* . Player 2 has a unique best reply to player 1's Stackelberg action, denoted by b^* .

Assumption 1 is satisfied when each player has a strict best reply to each of his opponent's pure actions and player 1 is not indifferent between any pair of pure action profiles, both of which are satisfied for generic (u_1, u_2) since A and B are finite sets. Player 1's *Stackelberg payoff* is $u_1(a^*, b^*)$. Let

$$\mathcal{B} \equiv \{\beta \in \Delta(B) | \exists \alpha \in \Delta(A) \text{ s.t. } \text{supp}(\beta) \subset \text{BR}_2(\alpha)\} \subset \Delta(B).$$

Since player 2s are myopic, they will never take actions that do not belong to $\mathcal{B}$. As a result, player 1's minmax value is $\underline{v}_1 \equiv \min_{\beta \in \mathcal{B}} \max_{a \in A} u_1(a, \beta)$.

Assumption 2. $a^* \notin \text{BR}_1(b^*)$ and $u_1(a^*, b^*) > \underline{v}_1$.

Assumptions 1 and 2 are satisfied in many leading applications of reputation models. For example,

1. In the product choice game of Mailath and Samuelson (2006), a firm benefits from committing to exert high effort since it can encourage consumers to purchase the high-end product or to purchase larger quantities. However, the firm can save costs by lowering its effort.
2. In the entry deterrence game of Kreps and Wilson (1982) and Milgrom and Roberts (1982a), and the limit pricing game of Milgrom and Roberts (1982b), an incumbent firm benefits from committing to set low prices and to fight potential entrants, but its stage-game payoff is higher when it accommodates entry.
3. In the fiscal policy game of Phelan (2006), the government benefits from committing to low tax rates in order to encourage investments, but it is tempted to expropriate the citizens after investment takes place.
4. In the monetary policy game of Barro (1986), the central bank can benefit from committing to low inflation rates. But given the households' expectations about inflation, the central bank is tempted to raise inflation in order to boost economic activities.

Assumption 2 rules out coordination games (such as the battle of sexes), common interest games, and chicken games, in which a^* best replies to b^* , and zero-sum games in which $u_1(a^*, b^*) \leq \underline{v}_1$. Section 5 discusses games that violate this assumption, and the role of Assumption 2 in our proofs is explained in Appendix A.

Player 1 has perfectly persistent private information about his type ω . Let $\omega \in \{\omega^s, \omega^c\}$, where ω^c stands for a *commitment type* who mechanically plays a^* in every period, and ω^s stands for a *strategic type* who can flexibly choose his actions in order to maximize his discounted average payoff $\sum_{t=0}^{+\infty} (1 - \delta)\delta^t u_1(a_t, b_t)$. Player 2's prior belief attaches probability $\pi \in (0, 1)$ to the commitment type.

Players' past actions are perfectly monitored. A typical public history is denoted by $h^t \equiv \{a_s, b_s, \xi_s\}_{s=0}^{t-1}$. Let $\mathcal{H}^t$ be the set of h^t and let $\mathcal{H} \equiv \cup_{t \in \mathbb{N}} \mathcal{H}^t$. Strategic-type player 1's strategy is $\sigma_1 : \mathcal{H} \rightarrow \Delta(A)$. Player 2's strategy is $\sigma_2 : \mathcal{H} \rightarrow \Delta(B)$. Let Σ_1 and Σ_2 be the set of player 1's and player 2's strategies, respectively.

The solution concept is (Bayes) Nash equilibrium. Let $\text{NE}(\delta, \pi) \subset \Sigma_1 \times \Sigma_2$ be the set of equilibria. Since the stage game is finite and payoffs are discounted, an equilibrium exists (Fudenberg and Levine, 1983).

Existing result on equilibrium payoffs Fudenberg and Levine (1989) show that for every $\pi \in (0, 1)$ and $\varepsilon > 0$, there exists $\underline{\delta} \in (0, 1)$ such that

$$\inf_{(\sigma_1, \sigma_2) \in NE(\delta, \pi)} \mathbb{E}^{(\sigma_1, \sigma_2)} \left[\sum_{t=0}^{+\infty} (1-\delta)\delta^t u_1(a_t, b_t) \right] \geq u_1(a^*, b^*) - \varepsilon \text{ for every } \delta > \underline{\delta}, \quad (2.1)$$

where $\mathbb{E}^{(\sigma_1, \sigma_2)}[\cdot]$ is the expectation when player 1's strategy is σ_1 and player 2's strategy is σ_2 .

Inequality (2.1) unveils the effects of reputations on the patient player's payoff. Fudenberg and Levine (1989) view this result as a refinement, which selects among the plethora of equilibria in repeated complete information games. According to the folk theorem of Fudenberg et al. (1990), the patient player can attain any payoff between $\underline{v}_1$ and $\bar{v}_1 \equiv \max_{\{(\alpha, \beta) | \text{supp}(\beta) \subset \text{BR}_2(\alpha)\}} \min_{a \in \text{supp}(\alpha)} u_1(a, \beta)$ in a repeated complete information game without any commitment type. By definition, $\bar{v}_1 \geq u_1(a^*, b^*)$, which implies that introducing a commitment type selects equilibria in which player 1's payoff is between $u_1(a^*, b^*)$ and $\bar{v}_1$. In the entry deterrence game, product choice game, and fiscal and monetary policy games, $\bar{v}_1$ equals $u_1(a^*, b^*)$, in which case the reputation model selects equilibria where the patient player receives his highest equilibrium payoff.

3. Results

Our results examine the *discounted frequencies* of the patient player's actions. Formally, the discounted frequency of action $a \in A$ under (σ_1, σ_2) is

$$G^{(\sigma_1, \sigma_2)}(a) \equiv \mathbb{E}^{(\sigma_1, \sigma_2)} \left[\sum_{t=0}^{\infty} (1-\delta)\delta^t \mathbf{1}\{a_t = a\} \right]. \quad (3.1)$$

Our first result characterizes the discounted frequencies with which the patient player plays his Stackelberg action a^* . Let

$$\Gamma \equiv \left\{ (\alpha, b) \in \Delta(A) \times B \mid b \in \text{BR}_2(\alpha) \right\} \quad (3.2)$$

be the set of *incentive compatible action profiles*. Let

$$F^*(u_1, u_2) \equiv \min_{(\alpha_1, \alpha_2, b_1, b_2, q) \in \Delta(A) \times \Delta(A) \times B \times B \times [0, 1]} \left\{ q\alpha_1(a^*) + (1-q)\alpha_2(a^*) \right\}, \quad (3.3)$$

subject to

$$(\alpha_1, b_1) \in \Gamma, \quad (\alpha_2, b_2) \in \Gamma, \quad (3.4)$$

and

$$qu_1(\alpha_1, b_1) + (1-q)u_1(\alpha_2, b_2) \geq u_1(a^*, b^*), \quad (3.5)$$

where $\alpha_i(a)$ stands for the probability of action $a \in A$ in $\alpha_i \in \Delta(A)$.

Theorem 1. Suppose (u_1, u_2) satisfies Assumptions 1 and 2.

1. For every $f \in [F^*(u_1, u_2), 1]$ and $\varepsilon > 0$, there exists $\underline{\delta} \in (0, 1)$ such that for every $\delta > \underline{\delta}$, there exists $(\sigma_1, \sigma_2) \in NE(\delta, \pi)$ such that $G^{(\sigma_1, \sigma_2)}(a^*) \in (f - \varepsilon, f + \varepsilon)$.
2. For every $\hat{f} < F^*(u_1, u_2)$, there exist $\underline{\delta} \in (0, 1)$ and $\eta > 0$ such that $G^{(\sigma_1, \sigma_2)}(a^*) > \hat{f} + \eta$ for every $\delta > \underline{\delta}$ and $(\sigma_1, \sigma_2) \in NE(\delta, \pi)$.

Theorem 1 implies that when player 1 is patient, the discounted frequency with which he plays a^* can take any value between $F^*(u_1, u_2)$ and 1, but it cannot be strictly lower than $F^*(u_1, u_2)$. Therefore, $[F^*(u_1, u_2), 1]$ is the set of frequencies with which a^* can arise in equilibrium. Our result applies to every prior belief $\pi \in (0, 1)$, which includes but not limited to situations where the probability of the commitment type is small. Since $F^*(u_1, u_2) < 1$ under Assumption 2, Theorem 1 implies that an arbitrarily patient player can play his Stackelberg action with frequency bounded away from one despite having the option to build a reputation.

The upper bound on the frequency of a^* is 1 since there exists an equilibrium where player 1 plays a^* and player 2s play b^* . Once player 1 plays any action other than a^* , future player 2s can observe this deviation after which they can punish player 1 by driving his continuation value to his minmax payoff $\underline{v}_1$. Such a punishment is feasible since player 1 separates from the commitment type after any deviation from his equilibrium strategy, and according to Fudenberg et al. (1990), there exists an equilibrium of the repeated complete information game in which player 1's payoff is $\underline{v}_1$. Since Assumption 2 requires that $u_1(a^*, b^*) > \underline{v}_1$, this punishment provides player 1 an incentive to play a^* when his discount factor δ is large enough.

For some intuition on the linear program that defines the lower bound $F^*(u_1, u_2)$, consider a static planning problem in which a planner commits to a mixed action $\alpha \in \Delta(A)$ on behalf of player 1 after which player 2 best replies to α . Suppose the planner faces a constraint that player 1's expected payoff is no less than $u_1(a^*, b^*)$, then by definition, $F^*(u_1, u_2)$ is the lowest probability with which a^* needs to be played.²

We map the two constraints in the planning problem to the reputation game studied by Theorem 1. First, since player 2s are myopic, they play a best reply to α after they learn that the patient player will play α . This explains the necessity of constraint (3.4). Second, the presence of commitment type implies that the patient player can guarantee payoff approximately $u_1(a^*, b^*)$ by playing a^* in every period. Therefore, the patient player has an incentive to play α_1 with probability q and α_2 with probability $1 - q$ only when his expected payoff from doing so is at least $u_1(a^*, b^*)$. This explains the necessity of constraint (3.5). The substantial part of our result is to show that constraints (3.4) and (3.5) are not only necessary but are also sufficient.

Our second result examines the set of discounted action frequencies that can arise in equilibrium. Let

$$\mathcal{A} \equiv \left\{ \alpha^* \in \Delta(A) \mid \exists q \in \Delta(\Gamma) \text{ such that } \alpha^* = \int_{\alpha} \alpha dq \text{ and } \int_{(\alpha, b)} u_1(\alpha, b) dq = u_1(a^*, b^*) \right\}, \quad (3.6)$$

which is the set of marginal distributions of player 1's actions such that one can find a distribution of incentive compatible action profiles $q \in \Delta(\Gamma)$ from which player 1's expected payoff equals his Stackelberg payoff.

Theorem 2. Suppose (u_1, u_2) satisfies Assumptions 1 and 2.

² The planner in the planning problem can randomize between any number of commitment actions, while in the linear program that defines $F^*(u_1, u_2)$, he can randomize between at most two commitment actions. Lemmas C.2 and C.3 show that this is without loss and the value of $F^*(u_1, u_2)$ remains the same even when the planner can randomize between any arbitrary number of commitment actions.

1. For every $\alpha^* \in \mathcal{A}$ and $\varepsilon > 0$, there exists $\underline{\delta} \in (0, 1)$ such that for every $\delta > \underline{\delta}$, there exists $(\sigma_1, \sigma_2) \in NE(\delta, \pi)$ such that $\left| G^{(\sigma_1, \sigma_2)}(a) - \alpha^*(a) \right| < \varepsilon$ for every $a \in A$.³
2. In games where $u_1(a^*, b^*) = \bar{v}_1$. For every $\hat{\alpha} \notin \mathcal{A}$, there exist $\eta > 0$ and $\underline{\delta} \in (0, 1)$ such that for every $\delta > \underline{\delta}$ and $(\sigma_1, \sigma_2) \in NE(\delta, \pi)$, $\left| G^{(\sigma_1, \sigma_2)}(a) - \hat{\alpha}(a) \right| > \eta$ for some $a \in A$.

According to Theorem 2, every action distribution that belongs to $\mathcal{A}$ is arbitrarily close to the patient player's action frequency in some equilibria of the reputation game. In fact, the first statement of Theorem 2 is a generalization of Statement 1 of Theorem 1 since it is without loss of generality to focus on distributions of incentive compatible action profiles such that constraint (3.5) is binding (Lemma B.1) and it is without loss of generality to focus on distributions supported on Γ that have at most two elements in their support when the objective is to minimize the discounted frequency of a^* (Lemma C.2 and Lemma C.3).

In games where $u_1(a^*, b^*) = \bar{v}_1$, such as the product choice game and the entry deterrence game, an action distribution is the patient player's discounted action frequency in some equilibria if and only if it belongs to $\mathcal{A}$. In this class of games, any action frequency that satisfies player 2's incentive constraints and yields player 1 his Stackelberg payoff can be attained in some equilibria of the repeated game.

4. Economic applications

We apply our results to *monotone-supermodular games* that include the leading applications of reputation models, such as the product choice game, the entry deterrence game, and the fiscal policy game.

Definition 1. (u_1, u_2) is monotone-supermodular if there exist a complete order on A and a complete order on B such that $u_1(a, b)$ is strictly decreasing in a , and $u_2(a, b)$ has strictly increasing differences.⁴

In order to facilitate the application of Theorem 1, we simplify the linear program that defines $F^*(u_1, u_2)$. Let $\underline{a}$ be the lowest element of A and let $\underline{b} \in B$ be player 2's best reply to $\underline{a}$. If player 2 has multiple best replies to $\underline{a}$, then let $\underline{b}$ be the one that maximizes player 1's payoff. Let

$$\Gamma^* \equiv \left\{ (\alpha, b) \in \Gamma \mid |\text{BR}_2(\alpha)| \geq 2 \text{ and } b \in \arg \max_{b' \in \text{BR}_2(\alpha)} u_1(\alpha, b') \right\}. \quad (4.1)$$

Intuitively, Γ^* is a subset of Γ that consists of incentive compatible action profiles where player 2 has at least two best replies, and for every α that player 2 has multiple best replies, b is the one that maximizes player 1's payoff. Under generic stage-game payoff functions, Γ^* is a finite set. Proposition 1 implies that in games with monotone-supermodular payoffs, it is without loss of generality to choose incentive compatible action profiles from the finite set $\Gamma^* \cup \{\underline{a}, \underline{b}\}$ instead of the infinite set Γ .

³ We can also show that if δ is large enough and $\mathcal{A}$ satisfies a full dimensionality assumption, then every α^* that belongs to the interior of $\mathcal{A}$ can be exactly attained as the discounted action frequency of some equilibria.

⁴ This definition resembles the one in Liu and Pei (2020) and Pei (2020a) except that there is no state that affects players' payoffs. We also do not require $u_1(a, b)$ to be strictly increasing in b .

Proposition 1. *If (u_1, u_2) is monotone-supermodular, then*

$$F^*(u_1, u_2) = \min_{(\alpha_1, \alpha_2, b_1, b_2, q) \in \Delta(A) \times \Delta(A) \times B \times B \times [0, 1]} \left\{ q\alpha_1(a^*) + (1 - q)\alpha_2(a^*) \right\},$$

subject to $(\alpha_1, b_1), (\alpha_2, b_2) \in \Gamma^ \cup \{(\underline{a}, \underline{b})\}$, and $qu_1(\alpha_1, b_1) + (1 - q)u_1(\alpha_2, b_2) \geq u_1(a^*, b^*)$.*

The proof is in Appendix D. For the rest of this section, we apply our theorems as well as Proposition 1 to study product choice games, entry deterrence games, and capital taxation games.

Product choice game Player 1 is a firm that chooses between high (action H) and low effort (action L). Player 2s are consumers, each chooses between purchasing a high-end product (action h) and a low-end product (action l). Players' payoffs are:

–	h	l
H	$1 - c_h, 2 - \gamma^*$	$-c_l, 1$
L	$1, -\gamma^*$	$0, 0$

where $c_h, c_l \in (0, 1)$ are the costs of effort when the consumer buys the high-end product and the low-end product, respectively, and consumers are willing to choose h only when they believe that the firm exerts high effort with probability more than $\gamma^* \in (0, 1)$.

This game has monotone-supermodular payoffs once we rank the firm's actions according to $H \succ L$ and the consumers' actions according to $h \succ l$. The firm's Stackelberg action is H . According to (4.1), Γ^* is a singleton set $\{(\gamma^*H + (1 - \gamma^*)L, h)\}$. Proposition 1 implies that

$$F^*(u_1, u_2) = \min_{q \in [0, 1]} q\gamma^*, \quad \text{subject to} \quad q\gamma^*u_1(H, h) + q(1 - \gamma^*)u_1(L, h) + (1 - q)u_1(L, l) \geq u_1(H, h), \quad (4.2)$$

from which we obtain

$$F^*(u_1, u_2) = \frac{\gamma^*(1 - c_h)}{1 - \gamma^*c_h}. \quad (4.3)$$

Claim 1. *The lowest discounted frequency with which the firm exerts high effort strictly increases in γ^* , strictly decreases in c_h , and is independent of c_l .*

In terms of practical implications, consider a policy maker who wants to increase the frequency with which the firm exerts high effort but does not know which equilibrium players coordinate on. The policy maker is ambiguity averse and evaluates the effectiveness of each policy according to the frequency of high effort in the *worst equilibrium*. That is, his objective is to increase $F^*(u_1, u_2)$.

Claim 1 implies that the policy maker can increase $F^*(u_1, u_2)$ by subsidizing consumers for purchasing the low-end product or by taxing consumers for purchasing the high-end product. Intuitively, these policies increase the consumers' demand for high effort when they purchase the high-end product. This leads to an increase in the equilibrium frequency of high effort since the firm needs to induce consumers to purchase the high-end product with high enough probability in order to obtain its Stackelberg payoff.

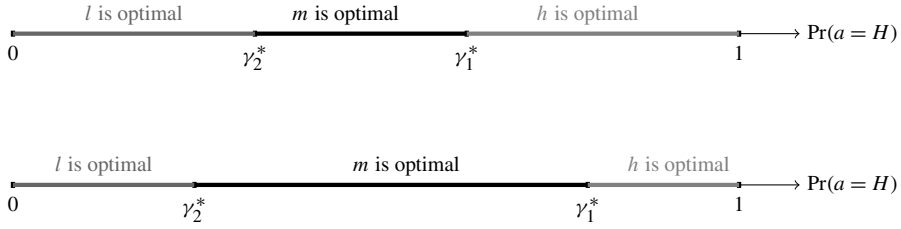


Fig. 1. Product choice game with three options: Consumer's best response before (upper panel) and after they receive a subsidy for purchasing the intermediate product (lower panel).

Next, we consider a variant of the product choice game in which every consumer chooses whether to buy a high-end product, an intermediate product, or a low-end product. The firm's payoffs are:

–	h	m	l
H	$1 - c$	$p - c$	$-c$
L	1	p	0

where its cost of effort is $c \in (0, 1)$, its benefit from selling the high-end product is 1, its benefit from selling the intermediate product is $p \in (0, 1)$, and its benefit from selling the low-end product is 0.

The value of $F^*(u_1, u_2)$ depends on consumers' payoffs only through two sufficient statistics γ_1^* and γ_2^* with $0 < \gamma_2^* < \gamma_1^* < 1$, such that a consumer has an incentive to choose h when the firm exerts high effort with probability more than γ_1^* , has an incentive to choose m when the firm exerts high effort with probability between γ_2^* and γ_1^* , and has an incentive to choose l when the firm exerts high effort with probability less than γ_2^* (see Fig. 1).

This game has monotone-supermodular payoffs once the firm's actions are ranked according to $H > L$ and consumers' actions are ranked according to $h > m > l$. Applying Proposition 1 to this game, we have:

$$F^*(u_1, u_2) = \begin{cases} \frac{\gamma_1^*(1-c)}{1-\gamma_1^*c} & \text{if } p \leq \frac{\gamma_2^*}{\gamma_1^*} \\ \frac{\gamma_2^*(1-c)}{p-\gamma_2^*c} & \text{if } p > \frac{\gamma_2^*}{\gamma_1^*} \text{ and } c \geq \frac{1-p}{1-\gamma_2^*} \\ \frac{\gamma_1^*(1-p)-c(\gamma_1^*-\gamma_2^*)}{(1-p)-c(\gamma_1^*-\gamma_2^*)} & \text{if } p > \frac{\gamma_2^*}{\gamma_1^*} \text{ and } c < \frac{1-p}{1-\gamma_2^*}. \end{cases} \quad (4.4)$$

Similar to the game with two purchasing options, we examine the effects of a small amount of sales taxes and subsidies for each product on $F^*(u_1, u_2)$.

1. A tax on consumers for purchasing the high-end product (i.e., an increase in γ_1^*) has no effect on $F^*(u_1, u_2)$ when $p > \frac{\gamma_2^*}{\gamma_1^*}$ and $c \geq \frac{1-p}{1-\gamma_2^*}$, and increases $F^*(u_1, u_2)$ otherwise. A subsidy on consumers for purchasing the low-end product (i.e., an increase in γ_2^*) has no effect on $F^*(u_1, u_2)$ when $p \leq \frac{\gamma_2^*}{\gamma_1^*}$, and increases $F^*(u_1, u_2)$ otherwise.
2. A subsidy on consumers for purchasing the intermediate product (i.e., a decrease in γ_2^* and an increase in γ_1^*) leads to an increase in $F^*(u_1, u_2)$ when $p \leq \frac{\gamma_2^*}{\gamma_1^*}$, leads to a decrease in

$F^*(u_1, u_2)$ when $p > \frac{\gamma_2^*}{\gamma_1^*}$ and $c \geq \frac{1-p}{1-\gamma_2^*}$, and has an ambiguous effect on $F^*(u_1, u_2)$ when $p > \frac{\gamma_2^*}{\gamma_1^*}$ and $c < \frac{1-p}{1-\gamma_2^*}$.

We obtain two additional insights compared to the case with two products. First, the effectiveness of subsidizing low-end products depends on the firm's benefit from selling intermediate products (i.e., the comparison between p and $\frac{\gamma_2^*}{\gamma_1^*}$). This is because a small subsidy for purchasing the low-end product only increases the demand for effort when the consumer decides whether to purchase the intermediate product instead of the low-end product, but does not affect consumers' demand for effort when deciding whether to purchase the high-end product instead of the intermediate product. When selling the intermediate product is unprofitable (i.e., $p \leq \frac{\gamma_2^*}{\gamma_1^*}$), an increase in the demand for effort when consumers decide between l and m does not affect the firm's equilibrium action frequencies. Similarly, the effectiveness of taxing high-end products also depends on the profitability of selling the intermediate product, that is, the comparison between c and $\frac{1-p}{1-\gamma_2^*}$.

Second, subsidizing consumers for purchasing intermediate products encourages the firm to exert effort more frequently when the firm's profit from selling the intermediate product is low (i.e., $p \leq \frac{\gamma_2^*}{\gamma_1^*}$), but encourages the firm to shirk more frequently otherwise. Intuitively, subsidizing the intermediate product has an effect similar to that of subsidizing the high-end product in the two-product setting when selling the intermediate product is attractive for the firm, and has an effect similar to that of subsidizing the low-end product when selling the intermediate product is unattractive.

Entry deterrence game Player 1 is an incumbent firm that chooses between fight (action F) and accommodate (action A). Player 2s are potential entrants. Each of them chooses between staying out (action O) and entering the market (action I). Players' payoffs are:

	O	I
F	$1 - c_o, 0$	$-c_i, -(1 - \gamma^*)$
A	$1, 0$	$0, \gamma^*$

where $c_o \in (0, 1)$ is the incumbent's cost of setting low prices when the potential entrant stays out, and $c_i > 0$ is its cost of setting low prices when the potential entrant enters. Each potential entrant prefers to stay out only when the incumbent fights with probability more than $\gamma^* \in (0, 1)$.

These payoffs are monotone-supermodular once we rank the incumbent's actions according to $F \succ A$, and the entrant's actions according to $O \succ I$. The incumbent's Stackelberg action is F . Proposition 1 implies that:

$$F^*(u_1, u_2) = \frac{(1 - c_o)\gamma^*}{1 - c_o\gamma^*}.$$

Claim 2. *The lowest discounted frequency with which the incumbent fights potential entrants strictly increases in γ^* , strictly decreases in c_o , and is independent of c_i .*

In terms of practical implications, consider a policy maker who can subsidize potential entrants for entering the market. This is modeled as an increase in every entrant's payoff from

action I by $s > 0$. Claim 2 implies that the frequency with which the incumbent fights entry is non-monotone with respect to the amount of subsidy. In particular,

1. When the subsidy to potential entrants is close to but strictly less than $1 - \gamma^*$, the strategic-type incumbent fights with frequency close to 1 in *all* equilibria. More generally, our formula implies that when $s < 1 - \gamma^*$, a marginal increase in the amount of subsidy increases $F^*(u_1, u_2)$.
2. When the subsidy is more than $1 - \gamma^*$, each entrant has a strict incentive to enter the market regardless of the incumbent's action, so the incumbent plays A in every period. Therefore, the frequency with which the incumbent fights is zero in *all* equilibria.

Fiscal policy game: Player 1 is a government that chooses between a normal tax rate and a high tax rate (i.e., expropriation) and player 2s are citizens who decide whether to invest. Players' payoffs are:

–	Invest	Not Invest
Normal Tax Rate	$\tau, 1 - \tau - c$	0, 0
Expropriate	1, $-c$	0, 0

where the low tax rate is $\tau \in (0, 1)$ and the cost of investment is $c \in (0, 1 - \tau)$. These payoffs are monotone-supermodular. The government's Stackelberg action is “normal tax rate” and its Stackelberg payoff is τ . According to Proposition 1, the highest frequency with which the government expropriates is:

$$1 - F^*(u_1, u_2) = 1 - \frac{\tau}{1 - \tau} \cdot \frac{c}{1 - c},$$

which is a decreasing function of both τ and c . This conclusion implies that in the worst case scenario, the frequency of government expropriation is lower when the government's revenue is higher under a normal tax rate (i.e., τ is larger), or when it is more costly for the citizens to invest (i.e., c is larger).

5. Discussions of modeling assumptions and results

The role of Assumption 2 Assumption 2 rules out games in which the optimal commitment outcome (a^*, b^*) is a stage-game Nash equilibrium (such as coordination games and chicken games), as well as games where player 1's optimal commitment payoff is no more than his minmax payoff (such as matching pennies).

Our formula for the lowest discounted frequency of the Stackelberg action fails when $u_1(a^*, b^*) \leq \underline{v}_1$. For example, consider the following variant of the matching penny game that satisfies Assumption 1 and the first part of Assumption 2 but violates the second part of Assumption 2:

–	h	t
H	$1 + \varepsilon, -1$	$-1 + \varepsilon, 1$
T	$-1, 1$	$1, -1$

where $\varepsilon > 0$. Player 1's unique Stackelberg action is H , his Stackelberg payoff is $-1 + \varepsilon$, and his minmax payoff is close to 0 when ε is small enough. Therefore, $F^*(u_1, u_2)$ is close to 0 when ε is close to 0.

However, if both π and ε are small, then the discounted frequency of action H is close to $1/2$ in every equilibrium. This means that neither the lower bound $F^*(u_1, u_2)$ nor the upper bound 1 can be approximately attained in any equilibrium of the reputation game.

In games where $u_1(a^*, b^*) > \underline{v}_1$, but (a^*, b^*) is a stage-game Nash equilibrium, our formula for the lowest discounted frequency for a^* applies to the battle of sexes game and the chicken game,

Battle of Sexes	o	f	Chicken Game	h	d
O	2, 1	0, 0	H	0, 0	7, 2
F	0, 0	1, 2	D	2, 7	6, 6

or more generally, when $u_1(a^*, b^*)$ is player 1's highest feasible payoff and $u_1(a^*, b^*) > u_1(a, b)$ for every $(a, b) \neq (a^*, b^*)$. In those games, $F^*(u_1, u_2) = 1$. This is because player 1's payoff is close to $u_1(a^*, b^*)$ in every equilibrium of the reputation game, so a^* must be played with discounted frequency close to 1.

Next, we present a counterexample that satisfies Assumption 1 and the second part of Assumption 2 but violates the first part of Assumption 2. Suppose players' payoffs are:

$-$	T	N
H	1, 1	0, 0
M	0, 3	3, 0
L	0, 0	0, 3

Player 1's Stackelberg action is H . Since N is player 2's best reply to player 1's mixed action $\frac{1}{2}M + \frac{1}{2}L$, from which player 1's expected payoff is $3/2$, the value of $F^*(u_1, u_2)$ is 0.

When the prior probability of commitment type π is strictly greater than $3/4$, the discounted frequency with which player 1 plays H is 1 in every equilibrium of the reputation game. This is because in every period where player 2 has not observed player 1 playing actions other than H , she has a strict incentive to play T , so player 1's payoff is 1 by playing H in every period. When player 1 deviates to M or L , his stage-game payoff is 0, and his continuation value is no more than 1 according to the folk theorem result of Fudenberg et al. (1990). This implies that player 1 plays H at every on-path history in every equilibrium.

Mixed-strategy commitment types Our model excludes commitment types that play mixed strategies. In order to understand the new challenges brought by mixed-strategy commitment types, consider the product choice game in Section 4 where with positive probability, player 1 is a type who mechanically plays $(\gamma^* + \varepsilon)H + (1 - \gamma^* - \varepsilon)L$ in every period, where $\varepsilon > 0$ is small.

A new complication arises since the strategic type can never be separated from the mixed-strategy commitment type. As a result, the continuation game always has nontrivial incomplete information regardless of the strategies being played. This stands in contrast to games where all commitment types play pure strategies, in which the strategic type is separated from a commitment type as soon as he stops imitating that type.

Analyzing repeated games with persistent private information and short-lived uninformed players is a well-known challenge in the repeated games literature, and to the best of our knowledge, there is no existing result that characterizes the informed player's equilibrium behaviors or his equilibrium action frequencies.⁵

Rich set of commitment types Our baseline model focuses on settings where there is only one commitment type. Our theorems extend to environments with any finite number of commitment types, as long as all of them play pure strategies, and there exists a commitment type who plays a^* in every period.

Our proof for the discounted frequency of action a^* being no less than $F^*(u_1, u_2)$ remains the same. On the construction of equilibria that approximately attain a given frequency in $\mathcal{A}$, for every type space that satisfies the above requirements, there exists $T \in \mathbb{N}$ such that for every $\delta \in (0, 1)$ and in every equilibrium under δ , player 2's posterior belief in period T assigns positive probability to at most one commitment type. Construct the continuation equilibrium starting from period T according to our proof in Appendix A, the discounted frequency of player 1's action is close to $\alpha^* \in \mathcal{A}$ when δ is close to 1.

Testable predictions Generally speaking, there are three challenges to test the predictions of reputation models.⁶ First, econometricians do not know which equilibrium players coordinate on. Second, econometricians usually observe players' behaviors rather than their payoffs, while most of the existing reputation results that apply to all equilibria (such as those in Fudenberg and Levine, 1989) are stated in terms of the patient player's payoff but not his behaviors. Third, many interesting equilibria in reputation games are in mixed strategies, but econometricians usually cannot observe these mixed strategies and can only observe the realized pure strategy.

Our results overcome the first and the second challenge by delivering predictions on the patient player's action frequencies that apply to all equilibria. Take the product choice game example in Section 4. The expression for $F^*(u_1, u_2)$ depends only on two terms:

1. γ^* : the minimal probability of high effort above which player 2 is willing to play h ;
2. c_h : the ratio between the cost of effort and the firm's benefit when a consumer buys the high-end product.

The values of γ^* and c_h can be computed without knowing all the details of players' stage-game payoff functions. Therefore, testing our predictions on the patient player's action frequencies has less demanding data requirements compared to testing the predictions on payoffs in canonical reputation models.

In context of the product choice game between a firm and a sequence of consumers, one way to address the third challenge is to use the distribution of the firm's actions across different markets

⁵ Very few results are obtained in repeated games between an informed patient player and a sequence of uninformed myopic players. Pei (2020b) characterizes the set of equilibrium payoffs between an informed seller and a sequence of uninformed buyers when the seller has persistent private information about his cost. His result relies on the assumption that all types of the seller have the same ordinal preference over stage-game outcomes, and does not apply when there are mixed-strategy commitment types.

⁶ Despite the large literature that takes repeated game predictions to the lab, see Dal Bó and Fréchet (2018), we are unaware of experimental results on repeated games with incomplete information between a patient player and a sequence of myopic players.

as a proxy for its mixed actions. This idea is applicable when the firm is a chain store that operates in many independent and geographically separated markets, and moreover, the consumers in each market can only observe the firm's actions in their own market but cannot observe the firm's actions in other markets. This is usually the case in developing countries where there is a lack-of record-keeping institutions, so that most consumers rely on word-of-mouth communication to learn about the firm's past behaviors. In these situations, it is reasonable to assume that consumers in one market cannot observe the firm's past behaviors in other markets. Using this idea, suppose an econometrician can observe the firm's behavior in every period and in every market, then he can compute the frequency of the firm's behaviors using his observations. He can then apply Theorems 1 and 2 to examine whether his observations are consistent with the predictions of reputation models.

The above discussion also unveils a limitation of our results, that they only characterize the set of action frequencies that can arise in equilibrium, but do not deliver predictions on the action frequencies that apply to *every path of equilibrium play*. Therefore, an econometrician cannot test our predictions after observing a realized path of equilibrium play. He can do that after observing the firm's mixed actions, e.g., observing the firm's behaviors across many markets and use the empirical distribution as a proxy for the firm's mixed action.

6. Conclusion

We examine the effects of reputation on the frequencies with which a patient player plays each of his actions. Our results characterize tight bounds that apply to all equilibria in a broad class of games. Our research question stands in contrast to the reputation literature that focuses on the patient player's equilibrium payoff. Our results stand in contrast to those that study the patient player's behavior in some particular equilibria.

Our results imply that in games where the optimal commitment outcome is not a stage-game Nash equilibrium, the patient player may play his optimal commitment action with frequency bounded away from one no matter how patient he is. When the patient player's optimal commitment payoff coincides with his highest equilibrium payoff in the repeated complete information game, reputation effects cannot further refine the patient player's behavior beyond that fact that his equilibrium payoff is at least his optimal commitment payoff.

In terms of applications, our results imply that a policy maker can increase the frequency with which a firm exerts high effort by subsidizing consumers for purchasing low-end products or by taxing consumers for purchasing high-end products. They also imply that a small amount of subsidy to potential entrants for entering the market makes an incumbent more aggressive in fighting entrants, but a large amount of subsidy encourages the incumbent to accommodate entry.

Appendix A. Overview of proofs

Our proof consists of two parts. Part 1 constructs a class of equilibria in which player 1's discounted action frequency is close to $\alpha^* \in \mathcal{A}$ when δ is close to 1. Part 2 shows that $G^{(\sigma_1, \sigma_2)}(a^*)$ cannot be strictly lower than $F^*(u_1, u_2)$ in any equilibrium when δ is large enough, and in games where $\bar{v}_1 = u_1(a^*, b^*)$, any action distribution that does not belong to $\mathcal{A}$ cannot be player 1's action frequency in any equilibrium.

The first part of our proof makes a methodological contribution, where we establish a discounted version of the Wald's inequality to bound the discounted frequency of each action. We provide an overview of our equilibrium construction and explain our methodological contribution

in this section, with details relegated to Appendix B. The second part of our proof is standard, which we relegate to Appendix C.

Equilibrium construction The first part of Assumption 2 implies the existence of $a' \neq a^*$ such that $u_1(a', b^*) > u_1(a^*, b^*)$. Since a^* is player 1's unique Stackelberg action, there exists $b' \neq b^*$ that best replies to a' such that $u_1(a', b') < u_1(a^*, b^*)$. Let $\alpha' \in \Delta\{a^*, a'\}$ be such that $\{b^*\} = \text{BR}_2(\alpha')$ and $u_1(\alpha', b^*) > u_1(a^*, b^*)$.

We construct a three-phase equilibrium in which the discounted frequency of player 1's actions is close to $\alpha^* \in \mathcal{A}$. Let $q \in \Delta(\Gamma)$ be a distribution of incentive compatible action profiles such that $\alpha^* = \int_{\alpha} \alpha dq$ and $\int_{(\alpha, b) \in \Gamma} u_1(\alpha, b) dq = u_1(a^*, b^*)$. The equilibrium play starts from a *preparation phase*, gradually reaches a *normal phase*, and reaches a *punishment phase* if and only if player 1 has made an off-path deviation.

1. Play belongs to the preparation phase when $t = 0$, or when $t \geq 1$ and (a^*, b^*) was played from period 0 to $t - 1$. In this phase, the strategic-type player 1 plays α' and player 2 plays b^* .
2. Play belongs to the normal phase when there exists $s \leq t - 1$ such that $(a_s, b_s) \neq (a^*, b^*)$. The normal phase consists of a number of *blocks*, and players' strategies in each block will be specified later on.
3. Player 1's continuation value when play first reaches the punishment phase is $\underline{v}_1$. This is feasible since player 2's belief attaches zero probability to the commitment type at every off-path history.

In every block of the normal phase, (α', b^*) is played for the first $T_1 \in \mathbb{N}$ periods, where T_1 is a constant that is independent of δ . A *review* happens by the end of these T_1 periods:

1. If (a', b^*) was not played in all T_1 periods, then play enters a *compensation subphase*, where (a', b') is played until period $T \in \mathbb{N}$ such that $(1 - \delta) \sum_{t=0}^T \delta^t u_1(a_t, b_t) = (1 - \delta^{T+1}) u_1(a^*, b^*)$. The current block ends in period T and the next block starts in period $T + 1$. If there is no such integer T , then use the public randomization device in the last period that satisfies $(1 - \delta) \sum_{t=0}^T \delta^t u_1(a_t, b_t) > (1 - \delta^{T+1}) u_1(a^*, b^*)$.
2. If (a', b^*) was played in all T_1 periods, then play enters an *absorbing subphase*, in which (α', b) is played with probability $\varepsilon_1 > 0$ and $q \in \Delta(\Gamma)$ is played with complementary probability, dictated by the realization of public randomization in the beginning of each period. The absorbing subphase ends in period T where T is the smallest integer that satisfies either

$$\sum_{t=0}^T \delta^t u_1(a_t, b_t) < (1 - \delta^{T+1}) u_1(a^*, b^*) + c(1 - \delta),$$

or

$$\sum_{t=0}^T \delta^t u_1(a_t, b_t) > (1 - \delta^{T+1}) \left(\varepsilon_1 u_1(\alpha', b^*) + (1 - \varepsilon_1) \mathbb{E}_{(\alpha, b) \sim q} [u_1(\alpha, b)] \right) - c(1 - \delta),$$

where $c > 0$ is a constant that is independent of δ . Once the absorbing subphase ends, play enters the compensation subphase described in the first bullet point, and the current block ends when $(1 - \delta) \sum_{t=0}^T \delta^t u_1(a_t, b_t) = (1 - \delta^{T+1}) u_1(a^*, b^*)$.

One can verify that player 2s' incentive constraints are satisfied. The strategic type's discounted average payoff is $u_1(a^*, b^*)$ from each of his on-path strategies, and his continuation value at every on-path history is bounded away from $\underline{v}_1$. The second part of Assumption 2 requires that $u_1(a^*, b^*) > \underline{v}_1$, which implies that player 1 has no incentive to make any off-path deviations when δ is large enough. The two together verify player 1's incentive constraints.

The challenging step is to compute player 1's discounted action frequencies when he uses this history-dependent mixed strategy. To the best of our knowledge, the existing centrality results in probability theory either cannot handle geometric discounting (such as the Chernoff-Hoeffding's inequality) or do not provide tight bounds on the probability of concentration (such as the Lindeberg-Feller central limit theorem), making those results inapplicable in our context. We establish a novel concentration inequality that can overcome both challenges, which is also applicable to future studies of players' behaviors in dynamic games.

Lemma A.1. *For every $\delta \in (0, 1)$, $c \geq 0$, and sequence of i.i.d. random variables Z_t with finite support and mean $\mu < 0$, and Z_t takes positive value with positive probability, we have:*

$$\Pr \left[\bigcup_{n=1}^{\infty} \left\{ \sum_{t=1}^n \delta^t Z_t \geq c \right\} \right] \leq \exp(-r^* \cdot c)$$

where $r^* > 0$ is the smallest positive real number such that $\mathbf{E}_{z \sim Z_1} [\exp(r^* z)] = 1$.

Intuitively, Lemma A.1 bounds the probability with which the discounted sum of a sequence of random variables deviates significantly from its expectation.

Proof of Lemma A.1. Let $\gamma_{Z,t}(r) = \ln \mathbf{E}_{z \sim Z_t} [\exp(rz\delta^t)]$, and let

$$q_{Z,r,t}(z) = p_Z(z) \exp(rz\delta^t - \gamma_{Z,t}(r)),$$

where $p_Z(z)$ is the probability mass function of random variable Z . One can verify that q is a well-defined probability measure. For a sequence of random variables $Z^n \equiv \{Z_1, \dots, Z_n\}$, we have

$$q_{Z^n,r}(z_1, \dots, z_n) = p_{Z^n}(z_1, \dots, z_n) \exp \left(\sum_{t=1}^n rz_t \delta^t - \sum_{t=1}^n \gamma_{Z_t,t}(r) \right).$$

Let $s_n = \sum_{t=1}^n z_t \delta^t$, we have

$$q_{S^n,r}(s_n) = p_{S^n}(s_n) \exp \left(rs_n - \sum_{t=1}^n \gamma_{Z_t,t}(r) \right).$$

Since $q_{S^n,r}$ is a probability measure, we have

$$\mathbf{E} \left[\exp \left(rs_n - \sum_{t=1}^n \gamma_{Z_t,t}(r) \right) \right] = 1. \quad (\text{A.1})$$

Let $\gamma(r) \equiv \mathbf{E}_{z \sim Z_1} [\exp(rz)]$, we have $\gamma(0) = 1$ and $\gamma'(0) = \mathbf{E}_{z \sim Z_1} [z] < 0$. Since $r^* > 0$ is the smallest positive real number such that $\mathbf{E}_{z \sim Z_1} [\exp(r^* z)] = 1$, we have $\gamma(r) \leq 1$ for any $0 \leq r \leq r^*$. Since random variables Z_t are i.i.d., we have

$$\gamma_{Z_t,t}(r^*) = \ln \mathbf{E}_{z \sim Z_t} [\exp(r^* z \delta^t)] = \ln \mathbf{E}_{z \sim Z_1} [\exp(r^* z \delta^t)] \leq 0$$

for every $t \geq 1$. By substituting $r = r^*$ in inequality (A.1), we have $\mathbf{E}[\exp(r^* s_n)] \leq 1$.⁷ Let J be the stopping time that the sum s_J first exceeds the threshold c , we have

$$\Pr[s_J \geq c] \cdot \mathbf{E}[\exp(r^* s_J) | s_J \geq c] \leq 1,$$

which implies that

$$\Pr\left[\bigcup_{n=1}^{\infty} \left\{ \sum_{t=1}^n \delta^t Z_t \geq c \right\}\right] = \Pr[s_J \geq c] \leq \exp(-r^* \cdot c). \quad \square$$

Back to the illustration of our constructive proof. Let Z be a random variable that equals 0 with probability $\varepsilon_1 \alpha'(a^*)$, equals $u_1(a^*, b^*) - u_1(a', b^*)$ with probability $\varepsilon_1 \alpha'(a')$, and equals $u_1(a^*, b^*) - u_1(\alpha, b)$ with probability $1 - \varepsilon_1$ where $(\alpha, b) \in \Gamma$ is drawn according to distribution q . Intuitively, Z_t measures the difference between the stage-game payoff player 1 receives in the absorbing subphase and his target payoff $u_1(a^*, b^*)$.

Since the support of Z is a finite set and the expectation of Z is negative, we can apply Lemma A.1 to a sequence of random variables with distribution Z . Our lemma implies that once play enters the absorbing subphase, the event that:

- $\sum_{t=0}^T \delta^t u_1(a_t, b_t)$ is between $(1 - \delta^{T+1})u_1(a^*, b^*) + c(1 - \delta)$ and $(1 - \delta^{T+1})\left(\varepsilon_1 u_1(\alpha', b^*) + (1 - \varepsilon_1)\mathbb{E}_{(\alpha, b) \sim q}[u_1(\alpha, b)]\right) - c(1 - \delta)$ for all $T \in \mathbb{N}$,

occurs with probability bounded away from 0. Since all other phases end in finite time in expectation, the discounted frequency of player 1's action is close to his discounted action frequency in the absorbing subphase, which is at most ε_1 away from α^* .

Remark on public randomization device The public randomization device is introduced to ease the exposition. It can be dispensed in our constructive proof for a reason similar to that in Fudenberg and Maskin (1991). In what follows, we provide an intuitive explanation based on the constructive proof of Theorem 1 in Appendix A. The details of the construction without public randomization are available upon request.

Recall $(\alpha_1, \alpha_2, b_1, b_2, q)$ which solves the constrained minimization problem that defines $F^*(u_1, u_2)$. Intuitively, the public randomization device plays two roles. First, it is used to implement particular interior action frequencies, i.e., those in which a^* is played with frequency strictly between 0 and 1. For this purpose, it is sufficient to choose a sequence of pure actions under which the discounted frequency approximates that of the implemented mixed action. Second, the public randomization device delivers the exact continuation payoff that makes player 1 indifferent by mixing between pure action profiles $(a, b) \in \Gamma$. As shown in Fudenberg and Maskin (1991), any payoff v can be decomposed as the discounted average payoff of an infinite sequence of deterministic pure action profiles $(a, b) \in \Gamma$ when player 1 is sufficiently patient. Therefore, our constructed equilibrium can be sustained in absence of public randomization. Finally, for any $\varepsilon > 0$, let T be the time period such that $\delta^T = \varepsilon$. When players have access to a public randomization device, we use the public randomization device by the end of each block to set the

⁷ Note that when $\delta = 1$, the inequality holds with equality, which is the Wald's identity established in Wald (1944).

discounted average payoff exactly to $u_1(a^*, b^*)$. In environments without the public randomization device, we can immediately start the next block if that block ends before period T . For any block after period T , we replace the public randomization device with an infinite sequence of deterministic pure action profiles that exactly implements the desired discounted payoff. Note that this does not affect the incentives of player 1 for using mixed strategies because the payoff differences in earlier blocks will be rectified by the compensation phase in later blocks. The public randomization device can be dispensed since the discounted frequency of any action affected by replacing the public randomization device after time T is at most ε .

Appendix B. Proofs of statement 1 of Theorems 1 and 2

We start from showing that Statement 1 of Theorem 1 is implied by Statement 1 of Theorem 2 by showing that it is without loss of generality to focus on $\{\alpha_1, \alpha_2, b_1, b_2, q\}$ such that (3.5) is binding in the constrained optimization problem that defines $F^*(u_1, u_2)$. Let

$$F^{**}(u_1, u_2) \equiv \min_{(\alpha_1, \alpha_2, b_1, b_2, q) \in \Delta(A) \times \Delta(A) \times B \times B \times [0, 1]} \{q\alpha_1(a^*) + (1 - q)\alpha_2(a^*)\}, \quad (\text{B.1})$$

subject to

$$(\alpha_1, b_1) \in \Gamma, \quad (\alpha_2, b_2) \in \Gamma, \quad (\text{B.2})$$

and

$$qu_1(\alpha_1, b_1) + (1 - q)u_1(\alpha_2, b_2) = u_1(a^*, b^*). \quad (\text{B.3})$$

Compared to $F^*(u_1, u_2)$, the objective function and the first constraint remains the same, but the inequality constraint (3.5) is replaced by the equality constraint (B.3).

Lemma B.1. Suppose (u_1, u_2) satisfies Assumptions 1 and 2, then $F^{**}(u_1, u_2) = F^*(u_1, u_2)$.

Proof. The part in which $F^{**}(u_1, u_2) \geq F^*(u_1, u_2)$ is straightforward. Next, we show $F^{**}(u_1, u_2) \leq F^*(u_1, u_2)$. Suppose the constrained minimum in (3.3) is attained by $\{\alpha_1, \alpha_2, b_1, b_2, q\}$ where $qu_1(\alpha_1, b_1) + (1 - q)u_1(\alpha_2, b_2) > u_1(a^*, b^*)$. Since a^* is player 1's unique Stackelberg action, for every $a' \neq a^*$, there exists $b' \in \text{BR}_2(a')$ such that $u_1(a', b') < u_1(a^*, b^*)$. Let $r \in [0, 1]$ be defined via:

$$ru_1(a', b') + (1 - r)(qu_1(\alpha_1, b_1) + (1 - q)u_1(\alpha_2, b_2)) = u_1(a^*, b^*).$$

Consider an alternative distribution $q' \in \Delta(\Gamma)$ that attaches probability r to (a', b') , probability $(1 - r)q$ to (α_1, b_1) , and probability rq to (α_2, b_2) . The probability of a^* is weakly lower under q' compared to that under q , and constraint (3.5) is binding. Later on, we show in Lemma C.3 that there exists a distribution over incentive compatible action profiles supported on two elements under which constraint (3.4) is satisfied, constraint (3.5) is binding, and attains the constrained minimum. Therefore, $F^{**}(u_1, u_2) \leq F^*(u_1, u_2)$. $\square$

In the remainder of this appendix, we show that the equilibrium constructed in Appendix A achieves the desired (discounted) action frequencies. We first define the parameters used in the construction of the equilibrium. Let $\varepsilon_1 > 0$ be a small positive real number, and let $Z_1 = u_1(a^*, b^*) - u_1(a, b)$ be a random variable that

- equals $u_1(a^*, b^*) - u_1(a^*, b^*)$ with probability $\varepsilon_1 \alpha'(a^*)$,
- equals $u_1(a^*, b^*) - u_1(a', b^*)$ with probability $\varepsilon_1 \alpha'(a')$,
- with probability $1 - \varepsilon_1$, equals $u_1(a^*, b^*) - u_1(a, b)$ where (a, b) is distributed according to q .

One can verify that Z_1 has finite support and $\mathbf{E}[Z_1] < 0$. Let $r_1^* > 0$ be the smallest real number such that $\mathbf{E}_{z \sim Z_1}[\exp(r_1^* \cdot z)] = 1$.⁸ Similarly, let $Z_2 = u_1(a, b) - \varepsilon_1$ be the random variable that:

- equals $u_1(a^*, b^*) - \varepsilon_1$ with probability $\varepsilon_1 \alpha'(a^*)$,
- equals $u_1(a', b^*) - \varepsilon_1$ with probability $\varepsilon_1 \alpha'(a')$,
- with probability $1 - \varepsilon_1$, equals $u_1(a, b) - \varepsilon_1$ where (a, b) is distributed according to q .

Let $r_2^* > 0$ be the smallest real number such that $\mathbf{E}_{z \sim Z_2}[\exp(r_2^* \cdot z)] = 1$. Let $\bar{M} \equiv \max_{(a,b) \in A \times B} u_1(a, b)$ and let $T_1 = \lceil \frac{\bar{M} + c}{u_1(a', b^*) - u_1(a^*, b^*)} \rceil$ where $c \in \mathbb{R}_+$ is such that $\exp(-\min\{r_1^*, r_2^*\} \cdot c) \leq \varepsilon_1$. Next we introduce several minor changes in the construction of the equilibrium in Appendix A to simplify the exposition.

- We impose a universal upper bound on the length of each absorbing subphase as $\bar{T}_2 \equiv \lceil \frac{\ln(1-\varepsilon_1)}{\ln \delta} \rceil$, and let $T_2 \leq \bar{T}_2$ be the stopping time of the absorbing subphase.⁹
- Letting T_0 be the starting time of the absorbing subphase, $T_2 \leq \bar{T}_2$ is the smallest integer that satisfies

$$\sum_{t=0}^{T_2} \delta^t u_1(a_{t+T_0}, b_{t+T_0}) < (1 - \delta^{T_2+1}) u_1(a^*, b^*) - c(1 - \delta)$$

or

$$\sum_{t=0}^{T_2} \delta^t u_1(a_{t+T_0}, b_{t+T_0}) > (1 - \delta^{T_2+1}) \left(\varepsilon_1 u_1(a', b^*) + (1 - \varepsilon_1) \mathbf{E}_{(\alpha, b) \sim q} [u_1(\alpha, b)] \right) + c(1 - \delta).$$

The second bulletin point defines the stopping criterion based on the discounted average payoff within the absorbing subphase. Moreover, the first inequality in the second bulletin is consistent with the constraint that the discounted average payoff from time 0 to T is above $u_1(a^*, b^*)$ because the accumulated payoff in the first T_1 periods of the current block is sufficiently high when we start the absorbing subphase.

Next we prove Statement 1 of Theorem 2 with the above parameters constructed in the equilibrium when $\delta > \bar{\delta}$ with

$$\bar{\delta} = \max \left\{ \frac{\ln(1 - \varepsilon_1^3)}{\ln T_1}, 1 - \varepsilon_1^2 \right\}. \quad (\text{B.4})$$

In the equilibrium constructed in Appendix A, the discounted payoff for player 1 in each block equals $(1 - \delta^T) u_1(a^*, b^*)$, in which $T \in \mathbb{N}$ is the number of time periods in the block.

⁸ Here we consider the case that the random variable Z_1 takes positive value with positive probability. As will become clearer in the analysis, the case when Z_1 only has non-positive support is trivial. We made the same assumption for Z_2 as well.

⁹ T_2 is the number of period in the current absorbing subphase, not the time horizon.

This implies that the strategic type has an incentive to play the mixed action in the beginning of the game to separate from the commitment type. In addition, one can verify that player 1 has no incentive to make any off-path deviations, since his expected continuation value at every on-path history is strictly greater than $\underline{v}_1$ when δ is sufficiently close to 1.

Let $\mathcal{E}_1$ be the event that player 1's discounted payoff in the absorbing subphase is less than $(1 - \delta^T)u_1(a^*, b^*) - c(1 - \delta)$. Let $\mathcal{E}_2$ be the event that player 1's discounted payoff in the absorbing subphase is more than $(1 - \delta^T)(\varepsilon_1 u_1(a', b^*) + (1 - \varepsilon_1)\mathbf{E}_{(\alpha, b) \sim q}[u_1(\alpha, b)] + \varepsilon_1) + c(1 - \delta)$. First, the probability that event $\mathcal{E}_1$ happens is bounded from above by the probability that $\sum_{t=1}^n \delta^t z_{1;t}$ is greater than c for some $n \geq 1$ where $z_{1;t} \sim Z_1$ for all t . According to Lemma A.1, the latter probability is bounded from above by $\exp(-r_1^* \cdot c) \leq \varepsilon_1$, which implies that $\Pr[\mathcal{E}_1] \leq \varepsilon_1$. Similarly, we have $\Pr[\mathcal{E}_2] \leq \varepsilon_1$. Let $\mathcal{E}_3$ be the event that action profile (a', b^*) is observed for T_1 periods, and by definition we have $\Pr[\mathcal{E}_3] = p^{T_1}$.

We first show that $G^{\sigma_1, \sigma_2}(a) \leq \alpha^*(a) + \varepsilon$ for every $a \in A$. Let G denote the discounted number of times action a is chosen from the beginning of each block. By construction, we have

$$\begin{aligned} G &\leq (1 - \delta^{T_1}) + (1 - p^{T_1} \cdot (1 - 2\varepsilon_1)) \cdot \delta^{T_1} G + (1 - 2\varepsilon_1) \cdot p^{T_1} \delta^{T_1 + \bar{T}_2} G \\ &\quad + p^{T_1} \delta^{T_1} (1 - \delta^{\bar{T}_2})(\varepsilon_1 + (1 - \varepsilon_1)\alpha^*(a)) \\ \Rightarrow G &\leq \frac{1 - \delta^{T_1} + p^{T_1} \delta^{T_1} (1 - \delta^{\bar{T}_2})(\varepsilon_1 + (1 - \varepsilon_1)\alpha^*(a))}{(1 - 2\varepsilon_1)(1 - \delta^{\bar{T}_2})\delta^{T_1} p^{T_1} + (1 - \delta^{T_1})} \leq \frac{\alpha^*(a) + \varepsilon_1}{1 - 2\varepsilon_1}. \end{aligned}$$

The first term in the first inequality is the upper bound on the discounted number of times action a is chosen from period 1 to T_1 ; the second term is the upper bound on the discounted number of times action a is chosen in future blocks conditional on event $(\mathcal{E}_1 \cup \mathcal{E}_2)$ happens; the third term is the upper bound on the discounted number of times action a is chosen in future blocks conditional on event $\neg(\mathcal{E}_1 \cup \mathcal{E}_2)$, and the last term is the upper bound on the discounted number of times action a is chosen in the absorbing subphase. The second inequality holds by rearranging terms. By setting $\varepsilon_1 \ll p^{T_1}$, the last inequality holds since $1 - \delta^{T_1} \leq \varepsilon_1^3$ and $1 - \delta^{\bar{T}_2} \approx \varepsilon_1$. Therefore,

$$\begin{aligned} \mathbb{E}^{(\sigma_1, \sigma_2)} \left[\sum_{t=0}^{\infty} (1 - \delta) \delta^t \mathbf{1}\{a_t = a\} \right] &\leq \sum_{t=0}^{\infty} p(1 - p)^t (1 - \delta^t + \delta^t G) \\ &= \frac{(1 - p)(1 - \delta)}{1 - (1 - p)\delta} + \frac{\alpha^*(a) + \varepsilon_1}{(1 - (1 - p)\delta)(1 - 2\varepsilon_1)} \leq \alpha^*(a) + \varepsilon, \end{aligned}$$

where the last inequality holds for sufficiently small $0 < \varepsilon_1 \ll \varepsilon$.

Next we show that $G^{\sigma_1, \sigma_2}(a) \geq \alpha^*(a) - \varepsilon$ for every $a \in A$. First, we provide an upper bound for the stopping time T . Conditional on event $\mathcal{E}_2 \cap \mathcal{E}_3$, the stopping time T satisfies

$$\begin{aligned} (1 - \delta^{T_1 + T_2})\bar{M} + \delta^{T_1 + T_2} (1 - \delta^{T - T_1 - T_2})u_1(a', b') &\geq (1 - \delta^T)u_1(a^*, b^*) \\ \Rightarrow \delta^T &\geq \frac{u_1(a^*, b^*) - \delta^{T_1 + T_2}(u_1(a', b') - (1 - \delta^{T_1 + T_2})\bar{M})}{u_1(a^*, b^*) - u_1(a', b')} \\ &\geq \delta^{T_1 + T_2} - \frac{(1 - \delta^{T_1 + T_2})\bar{M}}{u_1(a^*, b^*) - u_1(a', b')} \geq \delta^{T_1 + \bar{T}_2} - \frac{(1 - \delta^{T_1 + \bar{T}_2})\bar{M}}{u_1(a^*, b^*) - u_1(a', b')} \end{aligned} \quad (\text{B.5})$$

Conditional on event $(\neg \mathcal{E}_2) \cap \mathcal{E}_3$, the stopping time T satisfies

$$\begin{aligned}
& (1 - \delta^{T_1})\bar{M} + \delta^{T_1}(1 - \delta^{T_2})(\varepsilon_1 u_1(\alpha', b^*) + (1 - \varepsilon_1)\mathbf{E}_{(\alpha, b) \sim q}[u_1(\alpha, b)] + \varepsilon_1) \\
& \quad + c(1 - \delta) + \delta^{T_1+T_2}(1 - \delta^{T-T_1-T_2})u_1(a', b') \geq (1 - \delta^T)u_1(a^*, b^*) \\
\Rightarrow \delta^T & \geq \frac{\delta^{T_1+T_2}(u_1(a^*, b^*) - (u_1(\alpha', b'))) - (1 - \delta^{T_1})\bar{M} - c(1 - \delta) - \delta^{T_1}(1 - \delta^{T_2})(\varepsilon_1 u_1(\alpha', b^*) + \varepsilon_1)}{u_1(a^*, b^*) - u_1(\alpha', b')} \\
& \geq \delta^{T_1+T_2} - \frac{\varepsilon_1^2(1 - \delta^{T_1+T_2})(2\bar{M} + c)}{u_1(a^*, b^*) - u_1(\alpha', b')} \geq \delta^{T_1+\bar{T}_2} - \frac{\varepsilon_1^2(1 - \delta^{T_1+\bar{T}_2})(2\bar{M} + c)}{u_1(a^*, b^*) - u_1(\alpha', b')}
\end{aligned} \tag{B.6}$$

Conditional on event $\neg \mathcal{E}_3$, the stopping time T satisfies

$$\begin{aligned}
& (1 - \delta^{T_1})\bar{M} + \delta^{T_1}(1 - \delta^{T-T_1})u_1(a', b') \geq (1 - \delta^T)u_1(a^*, b^*) \\
\Rightarrow \delta^T & \geq \frac{u_1(a^*, b^*) - \delta^{T_1}u_1(\alpha', b') - (1 - \delta^{T_1})\bar{M}}{u_1(a^*, b^*) - u_1(\alpha', b')} \\
& \geq \delta^{T_1} - \frac{(1 - \delta^{T_1})\bar{M}}{u_1(a^*, b^*) - u_1(\alpha', b')}
\end{aligned} \tag{B.7}$$

Let G denote the discounted number of times action a is chosen in each block. By construction, we have

$$\begin{aligned}
G & \geq (1 - p^{T_1})(\delta^{T_1} - \frac{(1 - \delta^{T_1})\bar{M}}{u_1(a^*, b^*) - u_1(\alpha', b')})G + p^{T_1}(1 - \varepsilon_1)(\delta^{T_1+\bar{T}_2} \\
& \quad - \frac{\varepsilon_1^2(1 - \delta^{T_1+\bar{T}_2})(2\bar{M} + c)}{u_1(a^*, b^*) - u_1(\alpha', b')})G \\
& \quad + p^{T_1}\varepsilon_1(\delta^{T_1+\bar{T}_2} - \frac{(1 - \delta^{T_1+\bar{T}_2})\bar{M}}{u_1(a^*, b^*) - u_1(\alpha', b')})G + p^{T_1}\delta^{T_1}(1 - \delta^{\bar{T}_2})(1 - \varepsilon_1)\alpha^*(a) \\
\Rightarrow G & \geq \frac{p^{T_1}\delta^{T_1}(1 - \delta^{\bar{T}_2})(1 - \varepsilon_1)\alpha^*(a)}{p^{T_1}\delta^{T_1}(1 - \delta^{\bar{T}_2}) + O(\varepsilon_1^2)} \geq \frac{\alpha^*(a)(1 - \varepsilon_1)}{1 + O(\varepsilon_1)}.
\end{aligned}$$

The first term in the first inequality is the lower bound on the discounted number of times action a is chosen in future blocks conditional on event $\neg \mathcal{E}_3$; the second term is the lower bound on the discounted number of times action a is chosen in future blocks conditional on event $\mathcal{E}_3 \cap (\neg \mathcal{E}_2)$; the third term is the lower bound on the discounted number of times action a is chosen in future blocks conditional on event $\mathcal{E}_3 \cap \mathcal{E}_2$; and the last term is the lower bound on the discounted number of times action a is chosen in absorbing subphase. Finally, we have

$$\begin{aligned}
\mathbb{E}^{(\sigma_1, \sigma_2)} \left[\sum_{t=0}^{\infty} (1 - \delta)\delta^t \mathbf{1}\{a_t = a\} \right] & \geq \sum_{t=0}^{\infty} p(1 - p)^t \delta^t G \\
& = \frac{\alpha^*(a)(1 - \varepsilon_1)}{(1 - (1 - p)\delta)(1 + O(\varepsilon_1))} \geq \alpha^*(a) - \varepsilon,
\end{aligned}$$

where the last inequality holds when ε_1 is sufficiently small compared to ε . Combining these bounds, we have

$$\left| \mathbb{E}^{(\sigma_1, \sigma_2)} \left[\sum_{t=0}^{\infty} (1 - \delta)\delta^t \mathbf{1}\{a_t = a\} \right] - \alpha^*(a) \right| \leq \varepsilon \quad \text{for every } a \in A.$$

Appendix C. Proofs of statement 2 of Theorems 1 and 2

First, we establish Statement 2 of Theorem 1. Let $\Delta(\Gamma)$ be the set of probability distributions on Γ whose support has countable number of elements. Let $F(u_1, u_2, \varepsilon)$ be the value of the following constrained optimization problem:

$$F(u_1, u_2, \varepsilon) \equiv \inf_{p \in \Delta(\Gamma)} \int \alpha(a^*) dp(\alpha, b), \quad (C.1)$$

subject to

$$\int u_1(\alpha, b) dp(\alpha, b) \geq u_1(a^*, b^*) - \varepsilon. \quad (C.2)$$

Our proof of the necessity part of Theorem 1 consists of three lemmas.

Lemma C.1. *For every $\pi > 0$ and $\varepsilon > 0$, there exists $\underline{\delta} \in (0, 1)$ such that for every $\delta > \underline{\delta}$,*

$$G^{(\sigma_1, \sigma_2)}(a^*) \geq F(u_1, u_2, \varepsilon) - (1 - \underline{\delta}) \quad \text{for every } (\sigma_1, \sigma_2) \in NE(\delta, \pi). \quad (C.3)$$

Lemma C.2. *For every u_1 and u_2 that satisfy Assumptions 1 and 2, $\lim_{\varepsilon \downarrow 0} F(u_1, u_2, \varepsilon) = F(u_1, u_2, 0)$.*

Lemma C.3. *For every u_1 and u_2 that satisfy Assumptions 1 and 2, $F^*(u_1, u_2) = F(u_1, u_2, 0)$.*

Proof of Lemma C.1. The reputation result in Fudenberg and Levine (1989) implies that for every $\pi > 0$ and $\varepsilon > 0$, there exists $\underline{\delta} \in (0, 1)$ such that for every $\delta > \underline{\delta}$,

$$\mathbb{E}^{(\sigma_1, \sigma_2)} \left[\sum_{t=0}^{+\infty} (1 - \delta) \delta^t u_1(a_t, b_t) \right] \geq u_1(a^*, b^*) - \varepsilon/2 \quad \text{for every } (\sigma_1, \sigma_2) \in NE(\delta, \pi). \quad (C.4)$$

For given $(\sigma_1, \sigma_2) \in NE(\delta, \pi)$, let $\mathcal{H}^*$ be a set of on-path histories such that $h^t \in \mathcal{H}^*$ if and only if

- a^* was played from period 0 to $t - 1$, and $\sigma_1(h^t)$ assigns positive probability to actions other than a^* .

By construction, for every $h^t \in \mathcal{H}^*$, player 2's posterior belief at h^t assigns probability at least π to the commitment type, and therefore, player 1's continuation value at h^t is at least $u_1(a^*, b^*) - \varepsilon/2$. Let $\overline{M} \equiv \max_{(a,b) \in A \times B} u_1(a, b)$. For every $a \in \text{supp}(\sigma_1(h^t)) \setminus \{a^*\}$ and $b \in \text{supp}(\sigma_2(h^t))$, player 1's continuation value at (h^t, a, b) , denoted by $v(h^t, a, b)$, satisfies:

$$v(h^t, a, b) \geq \frac{1}{\delta} \left(u_1(a^*, b^*) - \frac{\varepsilon}{2} - (1 - \delta) \overline{M} \right).$$

The right-hand-side is strictly greater than $u_1(a^*, b^*) - \varepsilon$ when δ is close enough to 1. For every on-path history h^s such that $h^s \succeq (h^t, a, b)$, player 2 attaches probability 1 to the rational type at h^s , and therefore, $\sigma_2(h^s)$ best replies against $\sigma_1(h^s)$. Therefore, $(\sigma_1(h^s), b) \in \Gamma$ for every $b \in \text{supp}(\sigma_2(h^s))$. Let $p_{(h^t, a, b)} \in \Delta(\Gamma)$ be a probability measure on Γ such that for every $(\alpha, b) \in \Gamma$,

$$p_{(h^t, a, b)}(\alpha, b) \equiv \mathbb{E}^{(\sigma_1, \sigma_2)} \left[\sum_{s=t+1}^{\infty} (1-\delta)\delta^{s-t-1} \mathbf{1}\{\sigma_1(h^s) = \alpha\} \sigma_2(b) \middle| (h^t, a, b) \right]. \quad (\text{C.5})$$

By construction, $p_{(h^t, a, b)}$ has a countable number of elements in its support, and player 1's continuation value at (h^t, a, b) , denoted by $v(h^t, a, b)$, satisfies

$$v(h^t, a, b) = \int u_1(\alpha, b) dp_{(h^t, a, b)}(\alpha, b) \geq u_1(a^*, b^*) - \varepsilon. \quad (\text{C.6})$$

The definition of $F(u_1, u_2, \varepsilon)$ in (C.1) and (C.2) suggests that:

$$G^{(h^t, a, b)}(a^*) \equiv \mathbb{E}^{(\sigma_1, \sigma_2)} \left[\sum_{s=t+1}^{\infty} (1-\delta)\delta^{s-t-1} \mathbf{1}\{a_s = a^*\} \middle| (h^t, a, b) \right] \geq F(u_1, u_2, \varepsilon). \quad (\text{C.7})$$

Next, we compute a lower bound on $G^{(\sigma_1, \sigma_2)}(a^*)$. Let $\widehat{\mathcal{H}}$ be the set of on-path histories $h^t \equiv (h^{t-1}, a_{t-1}, b_{t-1})$ such that $t \geq 1$, $h^{t-1} \in \mathcal{H}^*$, and $a_{t-1} \neq a^*$. Let $p^{(\sigma_1, \sigma_2)}(h^t)$ be the ex ante probability of history h^t under the probability measure induced by (σ_1, σ_2) . By definition, $1 - \sum_{h^t \in \widehat{\mathcal{H}}} p^{(\sigma_1, \sigma_2)}(h^t)$ is the ex ante probability with which player 1 plays a^* in every period conditional on him being the rational type. Therefore,

$$\begin{aligned} G^{(\sigma_1, \sigma_2)}(a^*) &= \left(1 - \sum_{h^t \in \widehat{\mathcal{H}}} p^{(\sigma_1, \sigma_2)}(h^t)\right) + \sum_{h^t \in \widehat{\mathcal{H}}} p^{(\sigma_1, \sigma_2)}(h^t) \left((1 - \delta^{t-1}) + \delta^t X^{(h^t)}(a^*)\right) \\ &\geq -(1 - \delta) + \left(1 - \sum_{h^t \in \widehat{\mathcal{H}}} p^{(\sigma_1, \sigma_2)}(h^t)\right) \\ &\quad + \sum_{h^t \in \widehat{\mathcal{H}}} p^{(\sigma_1, \sigma_2)}(h^t) \left((1 - \delta^t) + \delta^t X^{(h^t)}(a^*)\right) \\ &\geq F(u_1, u_2, \varepsilon) - (1 - \delta) \geq F(u_1, u_2, \varepsilon) - (1 - \underline{\delta}) \quad \square \end{aligned} \quad (\text{C.8})$$

Proof of Lemma C.2. By definition, the value of $F(u_1, u_2, \varepsilon)$ is a decreasing function of ε and is bounded by $[0, 1]$. Therefore, $\lim_{\varepsilon \downarrow 0} F(u_1, u_2, \varepsilon)$ exists and moreover, $\lim_{\varepsilon \downarrow 0} F(u_1, u_2, \varepsilon) \leq F(u_1, u_2, 0)$.

Next, we show that $\lim_{\varepsilon \downarrow 0} F(u_1, u_2, \varepsilon) \geq F(u_1, u_2, 0)$. The optimization problem that defines $F(u_1, u_2, \varepsilon)$ implies that for every $\varepsilon > 0$, there exists $p_\varepsilon \in \Delta(\Gamma)$ that has countable number of elements in its support such that $\int \alpha(a^*) dp_\varepsilon(\alpha, b) \leq F(u_1, u_2, \varepsilon) + \varepsilon$ and $\int u_1(\alpha, b) dp_\varepsilon(\alpha, b) \geq u_1(a^*, b^*) - \varepsilon$.

According to Assumption 2, there exists $a' \in A$ such that $u_1(a', b^*) > u_1(a^*, b^*)$. According to Assumption 1, b^* is player 2's strict best reply against a^* . This implies the existence of $\alpha^* \in \Delta(A)$ such that $\alpha^*(a^*) \neq 1$, $b^* \in \text{BR}_2(\alpha^*)$, and $u_1(\alpha^*, b^*) > u_1(a^*, b^*)$. Let $\rho \equiv u_1(\alpha^*, b^*) - u_1(a^*, b^*)$. Since the support of p_ε is countable, there exists $\alpha_\varepsilon^* \in \Delta(A)$ such that $\alpha_\varepsilon^*(a^*) \neq 1$, $b^* \in \text{BR}_2(\alpha_\varepsilon^*)$, $u_1(\alpha_\varepsilon^*, b^*) - u_1(a^*, b^*) \geq \frac{\rho}{2}$, and $(\alpha_\varepsilon^*, b^*)$ does not belong to the support of p_ε . We construct probability measure $p'_\varepsilon \in \Delta(\Gamma)$ according to:

- $p'_\varepsilon(\alpha_\varepsilon^*, b^*) \equiv \frac{2\varepsilon}{\rho + 2\varepsilon}$.
- $p'_\varepsilon(\alpha, b) \equiv \frac{\rho}{\rho + 2\varepsilon} p_\varepsilon(\alpha, b)$ for every (α, b) that belongs to the support of p_ε .

By construction, $\int u_1(\alpha, b) dp'_\varepsilon(\alpha, b) \geq u_1(a^*, b^*)$, and therefore,

$$\frac{2\varepsilon}{\rho + 2\varepsilon} + \frac{\rho}{\rho + 2\varepsilon} \left(F(u_1, u_2, \varepsilon) + \varepsilon \right) \geq \int \alpha(a^*) dp'_\varepsilon(\alpha, b) \geq F(u_1, u_2, 0). \quad (\text{C.9})$$

This implies that

$$\lim_{\varepsilon \downarrow 0} \left\{ \frac{2\varepsilon}{\rho + 2\varepsilon} + \frac{\rho}{\rho + 2\varepsilon} \left(F(u_1, u_2, \varepsilon) + \varepsilon \right) \right\} = \lim_{\varepsilon \downarrow 0} F(u_1, u_2, \varepsilon) \geq F(u_1, u_2, 0). \quad \square$$

Proof of Lemma C.3. The inequality that $F^*(u_1, u_2) \geq F(u_1, u_2, 0)$ is implied by the definitions of $F^*(u_1, u_2)$ and $F(u_1, u_2, 0)$. In what follows, we show that $F^*(u_1, u_2) \leq F(u_1, u_2, 0)$. For every $\eta > 0$, there exists $p_\eta \in \Delta(\Gamma)$ that has countable number of elements in its support such that $\int \alpha(a^*) dp_\eta(\alpha, b) \leq F(u_1, u_2, 0) + \eta$ and $\int u_1(\alpha, b) dp_\eta(\alpha, b) \geq u_1(a^*, b^*)$. Let Γ_η be a countable subset of Γ that contains the support of p_η . Consider the following minimization problem:

$$F_\eta \equiv \min_{p \in \Delta(\Gamma_\eta)} \sum_{(\alpha, b) \in \Gamma_\eta} p(\alpha, b) \alpha(a^*), \quad (\text{C.10})$$

subject to

$$\sum_{(\alpha, b) \in \Gamma_\eta} p(\alpha, b) u_1(\alpha, b) \geq u_1(a^*, b^*). \quad (\text{C.11})$$

By construction, $F_\eta \leq \int \alpha(a^*) dp_\eta(\alpha, b) \leq F(u_1, u_2, 0) + \eta$. We show that F_η can be attained via a distribution that contains at most two elements in its support. The Lagrangian of the minimization problem is:

$$\sum_{(\alpha, b) \in \Gamma_\eta} p(\alpha, b) \alpha(a^*) + \lambda \left(\sum_{(\alpha, b) \in \Gamma_\eta} p(\alpha, b) u_1(\alpha, b) - u_1(a^*, b^*) \right), \quad (\text{C.12})$$

where λ is the Lagrange multiplier. If constraint (C.11) is not binding, then the minimum is zero and is attained by a degenerate distribution. If constraint (C.11) is binding, then for every pair of elements (α, b) and (α', b') in the support of the minimand $p_\eta^* \in \Delta(\Gamma_\eta)$,

$$\alpha(a^*) + \lambda u_1(\alpha, b) = \alpha'(a^*) + \lambda u_1(\alpha', b'). \quad (\text{C.13})$$

Label the elements in the support of p_η^* as $\{(\alpha_i, b_i)\}_{i=1}^{+\infty}$. Equation (C.13) implies that for every $\alpha_i(a^*) \neq \alpha_j(a^*)$,

$$\frac{u_1(\alpha_i, b) - u_1(\alpha_j, b)}{\alpha_i(a^*) - \alpha_j(a^*)} = -\frac{1}{\lambda}. \quad (\text{C.14})$$

Let

$$\begin{aligned} \bar{u}_1 &\equiv \sup_{(\alpha, b) \in \{(\alpha_i, b_i)\}_{i=1}^{+\infty}} u_1(\alpha, b), & \underline{u}_1 &\equiv \inf_{(\alpha, b) \in \{(\alpha_i, b_i)\}_{i=1}^{+\infty}} u_1(\alpha, b), \\ \bar{q} &\equiv \sup_{(\alpha, b) \in \{(\alpha_i, b_i)\}_{i=1}^{+\infty}} \alpha(a^*), & \text{and } \underline{q} &\equiv \inf_{(\alpha, b) \in \{(\alpha_i, b_i)\}_{i=1}^{+\infty}} \alpha(a^*). \end{aligned}$$

Equation (C.14) implies that

$$\frac{\bar{u} - \underline{u}}{\bar{q} - \underline{q}} = -\frac{1}{\lambda}.$$

Let $\gamma \in (0, 1)$ be such that $\gamma \bar{u}_1 + (1 - \gamma) \underline{u}_1 = u_1(a^*, b^*)$. According to (C.14), we have $\gamma \bar{q} + (1 - \gamma) \underline{q} = F_\eta$.

Since $\Delta(A) \times B$ is compact, there exist $(\bar{\alpha}, \bar{b})$ and $(\underline{\alpha}, \underline{b})$ which are limit points of set $\{(\alpha_i, b_i)\}_{i=1}^{+\infty}$ such that $u_1(\bar{\alpha}, \bar{b}) = \bar{u}_1$, $\bar{\alpha}(a^*) = \bar{q}$, $u_1(\underline{\alpha}, \underline{b}) = \underline{u}_1$, and $\underline{\alpha}(a^*) = \underline{q}$. Since player 2's best reply correspondence is upper-hemi-continuous, $(\bar{\alpha}, \bar{b}), (\underline{\alpha}, \underline{b}) \in \Gamma$. Our analysis above suggests that there exists a distribution on $\Gamma_\eta \cup \{(\bar{\alpha}, \bar{b}), (\underline{\alpha}, \underline{b})\}$ with at most two elements in its support that satisfies constraint (C.2) and the value of the objective function (C.1) is at most $F(u_1, u_2, 0) + \eta$.

Take a decreasing sequence of positive real numbers $\{\eta_n\}_{n \in \mathbb{N}}$ such that $\lim_{n \rightarrow \infty} \eta_n = 0$. For every $n \in \mathbb{N}$, there exists $p_n \in \Delta(\Gamma)$ with at most two elements in its support that satisfies constraint (C.2) and the value of the objective function is at most $F(u_1, u_2, 0) + \eta_n$. Since $(\Delta(A_1) \times B)^2$ is compact, there exists a converging subsequence $\{p_{k_n}\}_{n \in \mathbb{N}}$ such that its limit p^* has at most two elements in its support, satisfies constraint (C.2), and the value of the objective function is at most $F(u_1, u_2, 0)$. This implies that $F^*(u_1, u_2) \leq F(u_1, u_2, 0)$. $\square$

In the last step, we modify the above proof in order to establish Statement 2 of Theorem 2. Since $\bar{v}_1 = u_1(a^*, b^*)$, for every $\varepsilon > 0$, there exists $\underline{\delta} \in (0, 1)$ such that player 1's payoff in every equilibrium where $\delta > \underline{\delta}$ is no more than $u_1(a^*, b^*) + \varepsilon$. Let

$$\mathcal{A}^\varepsilon = \left\{ \alpha^* \in \Delta(A) \mid \exists q \in \Delta(\Gamma) \text{ such that } \alpha^* = \int_{\alpha} \alpha dq \text{ and } \left| \int_{(\alpha, b)} u_1(\alpha, b) dq - u_1(a^*, b^*) \right| \leq \varepsilon \right\}. \quad (\text{C.15})$$

Lemma C.1 implies that for every $\alpha' \notin \mathcal{A}^\varepsilon$, there exist $\eta > 0$ and $\underline{\delta} \in (0, 1)$ such that for every $\delta > \underline{\delta}$ and every $(\sigma_1, \sigma_2) \in \text{NE}(\delta, \pi)$, we have:

$$\left| G^{(\sigma_1, \sigma_2)}(a) - \alpha^*(a) \right| > \eta \text{ for some } a \in A. \quad (\text{C.16})$$

The conclusion of Theorem 2 is obtained since $\lim_{\varepsilon \rightarrow 0} \mathcal{A}^\varepsilon = \mathcal{A}$.

Appendix D. Proof of Proposition 1

First, suppose toward a contradiction that $\{\alpha_1, \alpha_2, b_1, b_2, q\}$ solves (3.3), player 2 has a strict incentive to play b_1 against α_1 , α_1 does not attach probability 1 to player 1's lowest action, and $q \neq 0$. One can increase the probability of $\underline{a}$ in α_1 and decrease the probability of other actions, after which player 1's expected payoff strictly increases and the probability of action a^* strictly decreases. This contradicts the presumption that $\{\alpha_1, \alpha_2, b_1, b_2, q\}$ solves the constrained minimization problem.

Next, suppose $q > 0$ and α_1 is such that $|\text{BR}_2(\alpha_1)| \geq 2$, and there exists $b \in \text{BR}(\alpha_1)$ such that $u_1(\alpha_1, b) > u_1(\alpha_1, b_1)$. Then replace b_1 by b in the constrained minimization problem, the value of F^* remains unchanged but constraint (3.5) becomes slack. This contradicts Lemma B.1 that it is without loss of generality to focus on $\{\alpha_1, \alpha_2, b_1, b_2, q\}$ such that constraint (3.5) binds.

References

- Bar-Isaac, Heski, 2003. Reputation and survival: learning in a dynamic signalling model. *The Review of Economic Studies* 70 (2), 231–251.
- Barro, Robert, 1986. Reputation in a model of monetary policy with incomplete information. *J. Monet. Econ.* 17 (1), 3–20.

- Cripps, Martin, Mailath, George, Samuelson, Larry, 2004. Imperfect monitoring and impermanent reputations. *Econometrica* 72 (2), 407–432.
- Dal Bó, Pedro, Fréchette, Guillaume R., 2018. On the determinants of cooperation in infinitely repeated games: a survey. *J. Econ. Lit.* 56 (1), 60–114.
- Ekmekci, Mehmet, 2011. Sustainable reputations with rating systems. *J. Econ. Theory* 146 (2), 479–503.
- Ekmekci, Mehmet, Maestri, Lucas, 2019. Reputation and screening in a noisy environment with irreversible actions. Working Paper.
- Fudenberg, Drew, Levine, David, 1983. Subgame-perfect equilibria of finite and infinite horizon games. *J. Econ. Theory* 31 (2), 251–268.
- Fudenberg, Drew, Levine, David, 1989. Reputation and equilibrium selection in games with a patient player. *Econometrica* 57 (4), 759–778.
- Fudenberg, Drew, Maskin, Eric, 1991. On the dispensability of public randomization in discounted repeated games. *J. Econ. Theory* 53 (2), 428–438.
- Fudenberg, Drew, Kreps, David, Maskin, Eric, 1990. Repeated games with long-run and short-run players. *The Review of Economic Studies* 57 (4), 555–573.
- Kreps, David, Wilson, Robert, 1982. Reputation and imperfect information. *J. Econ. Theory* 27 (2), 253–279.
- Liu, Qingmin, 2011. Information acquisition and reputation dynamics. *The Review of Economic Studies* 78 (4), 1400–1425.
- Liu, Qingmin, Skrzypacz, Andrzej, 2014. Limited records and reputation bubbles. *J. Econ. Theory* 151, 2–29.
- Liu, Shuo, Pei, Harry, 2020. Monotone equilibria in signaling games. *Eur. Econ. Rev.*, 103408.
- Mailath, George J., Samuelson, Larry, 2006. *Repeated Games and Reputations: Long-Run Relationships*. Oxford University Press.
- Milgrom, Paul, Roberts, John, 1982a. Predation, reputation, and entry deterrence. *J. Econ. Theory* 27 (2), 280–312.
- Milgrom, Paul, Roberts, John, 1982b. Limit pricing and entry under incomplete information: an equilibrium analysis. *Econometrica*, 443–459.
- Pei, Harry, 2020a. Reputation effects under interdependent values. *Econometrica* 88 (5), 2175–2202.
- Pei, Harry, 2020b. Trust and betrayals: reputational payoffs and behaviors without commitment. *Theor. Econ.* Forthcoming.
- Phelan, Christopher, 2006. Public trust and government betrayal. *J. Econ. Theory* 130 (1), 27–43.
- Schmidt, Klaus, 1993. Commitment through incomplete information in a simple repeated bargaining game. *J. Econ. Theory* 60 (1), 114–139.
- Wald, Abraham, 1944. On cumulative sums of random variables. *Ann. Math. Stat.* 15 (3), 283–296.