REPUTATION IN REPEATED MORAL HAZARD GAMES[1]

Alp E. Atakan and Mehmet Ekmekci

We study an infinitely repeated game where two players with equal discount factors play a simultaneous-move stage game. Player one monitors the stage-game actions of player two imperfectly, while player two monitors the pure stage-game actions of player one perfectly. Player one's type is private information and he may be a "commitment type," drawn from a countable set of commitment types, who is locked into playing a particular strategy. Under a full-support assumption on the monitoring structure, we prove a reputation result for repeated moral hazard games: if there is positive probability that player one is a particular type whose commitment payoff is equal to player one's highest payoff, consistent with the players' individual rationality, then a patient player one secures this type's commitment payoff in any Bayes-Nash equilibrium of the repeated game.

Keywords: Repeated Games, Reputation, Equal Discount Factor, Long-run Players, Imperfect Monitoring, Complicated Types, Finite Automaton

JEL Classification Numbers: C73, D83.

Koç University, Department of Economics, Rumeli Feneri Yolu, Sarıyer 34450, Istanbul, Turkey. Email: aatakan@ku.edu.tr.

MEDS, Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208. Email: m-ekmekci@kellogg.northwestern.edu.

[1]We would like to thank Martin Cripps, Eddie Dekel, Eduardo Faingold, Johannes Horner, Christoph Kuzmics, and Larry Samuelson for useful discussions. This paper has also benefitted from the comments of seminar participant at Northwestern University, Princeton University, Yale University, University of Pennsylvania, Koç University. Part of this research was conducted while Mehmet Ekmekci was visiting the Cowles Foundation and Economics Department at Yale University.

Contents

1. Introduction

The desire to maintain one's reputation is a powerful incentive in a long-run relationship as a strong reputation can lend credibility to an individual's (or an institution's) commitments, threats, or promises. It can help a firm commit to fight competitors planning to enter its market, it can assist a government in committing to its monetary and fiscal policies, or it can facilitate trade based on trust when formal institutions are lacking. In fact, a patient player's reputation concerns are the dominant incentives that determine equilibrium payoffs in repeated games where a patient player faces a myopic opponent. And this is true regardless of the monitoring structure.[1]

Building a reputation when facing an equally patient opponent, however, is more difficult. A patient opponent might be willing to sacrifice short-term payoffs to test whether the player, who is trying to build a reputation, will go through with his threats or promises. This makes it prohibitively expensive to build a reputation in certain repeated simultaneous-move games played against a patient opponent if stage-game actions are *perfectly monitored* (Cripps and Thomas (1997)). In this paper, we instead focus on repeated simultaneous-move games played by equally patient players where the opponent's stage-game actions are *imperfectly monitored*. A leading example of significant economic interest is the repeated principal-agent game. We show that reputation effects are prominent under imperfect monitoring even in certain repeated games where reputation effects are absent under perfect monitoring.

Specifically, suppose that player one's type is private information and that he may be a "commitment type" who is locked into playing a particular strategy. We explore whether an uncommitted or "normal" player can exploit his opponent's uncertainty to establish a reputation for a particular behavior. We also address two related questions. First, we ask which behavior (strategy or strategic posture) would a "normal" player mimic in order to successfully build a beneficial reputation? In other words, which types, if available, facilitate reputation building for player one?[2] Second, we ask in which strategic situations (i.e., for which class of stage games) can player one successfully build a reputation?

Our central finding is a *reputation result* in repeated games where player one (he) observes only an imperfect public signal of his opponent's stage-game action while his opponent (she) perfectly monitors player one's actions. We show that a patient player one can guarantee his highest payoff compatible with the players' individual rationality (player one's highest IR

[1] See Fudenberg and Levine (1989) for the case of perfect monitoring, Fudenberg and Levine (1992) for imperfect public monitoring, and Gossner (2011) for imperfect private monitoring.

[2] We say that a certain type is available if player two believes that player one is this type with positive probability.

payoff) in any Bayesian-Nash equilibrium of the repeated game. For our reputation result, we assume that a certain commitment type, which satisfies two properties, is available. The first property, which we call *no shortfall*, requires that the type's commitment payoff is equal to player one's highest IR payoff.[3] The second requires that the per period cost of not best responding to this type is positive, even for an arbitrarily patient player two. If this type is available, then player one guarantees this type's commitment payoff simply by mimicking its strategy, even if player two believes that player one is another commitment type with arbitrarily higher probability. In other words, this commitment type with no shortfall facilitates reputation building.

For our reputation result, we also assume that the stage game has *locally nonconflicting interests* (LNCI).[4] There are LNCI in a game if player two's payoff, in the payoff profile where player one receives his highest IR payoff, strictly exceeds her pure minimax payoff. This restriction on the stage game ensure the existence of a commitment type that satisfies the aforementioned two properties.

One key assumption, which we have not yet discussed at length, is that player one does not observe player two's intended action, but only sees an imperfect signal of it, as in a model of moral hazard. We also assume that the support of the distribution of signals is independent of how player two plays; we call this the *full-support imperfect-monitoring* assumption. This assumption is indispensable and, intuitively, ensures that every reward and punishment in player one's strategy will occasionally be triggered, so that player two will learn how player one responds to all sequences of public outcomes.

We obtain our reputation result by calculating a lower bound, which holds across all equilibria, on player one's payoff when he mimics a commitment type that plays a *pure strategy* (as in Fudenberg and Levine (1989)). In this context, our assumption that player one's stage-game actions are *perfectly monitored* greatly aids our analysis. This is because the perfect-monitoring assumption simplifies the dynamics of how player one's reputation evolves. In particular, because player two perfectly monitors player one's stage-game actions and because the commitment type plays a pure strategy, player one's reputation level weakly increases - but only as long as player two observes him play the same stage-game action as the action the commitment type would have played; otherwise, his reputation level collapses to zero.[5] If we relax the assumption that player one's actions are perfectly monitored, then a

[3]The commitment payoff of a type is the payoff that player one can guarantee by publicly committing to play the repeated-game strategy that this type plays. A type's (or strategy's) *shortfall* is the difference between player one's highest IR payoff and the type's commitment payoff.

[4]We also assume that the stage-game satisfies a certain technical genericity property. Specifically, we assume that the payoff profile in which player one obtains his highest IR payoff is unique. We term this genericity property *no gap*.

[5]We use these dynamics in proving both reputation results and our non reputation results.

technically challenging statistical learning problem arises. Whether an appropriate statistical learning technique can be developed or applied for this framework remains an open question beyond the scope of this paper.[6,7]

Lastly, the reputation results in games with asymmetric discounting (Fudenberg and Levine (1989, 1992) or Celentani et al. (1996)) are robust to the introduction of two-sided uncertainty, while the reputation result that we present in this paper is not. In order to obtain our one-sided reputation result, we allow for only one-sided uncertainty. In other words, we replace asymmetric discount factors as in Fudenberg and Levine (1989, 1992) or Celentani et al. (1996) with one-sided asymmetric information.

1.1. Related literature and our contribution. This paper is most closely related to work on reputation effects in repeated simultaneous-move games with equally patient agents (see Cripps and Thomas (1997), Cripps et al. (2005), and Chan (2000)).[8] We make three main contributions to this literature. First, we provide the first reputation result for all games with a strong Stackelberg action and games LNCI.[9] Previous reputation results are for only a strict subset of stage games with a strong Stackelberg action: those with strictly conflicting interests (Cripps et al. (2005)) or strictly-dominant-action stage games (Chan (2000)).[10] Second, we are the first to explore reputation effects under imperfect monitoring. Previous work assumed perfect monitoring. Finally, our work highlights the role that full-support imperfect monitoring plays for a reputation effect in repeated games with LNCI. Without full-support imperfect monitoring, our reputation result may fail to obtain for repeated games with LNCI (Cripps and Thomas (1997) and Chan (2000)).

This paper also relates to work on reputation effects in repeated games where a patient player one faces a nonmyopic, but arbitrarily less patient, opponent (Schmidt (1993), Celentani et al. (1996), Aoyagi (1996), Cripps et al. (1996), Evans and Thomas (1997)). In repeated games where a patient player faces a less patient opponent, Celentani et al. (1996) and Aoyagi (1996) establish reputation results under full-support imperfect monitoring. How-

[6]Fudenberg and Levine (1992)'s learning result (Theorem 4.1) does not help in our framework with equally patient agents.

[7]Note that we place no restriction on player one's other commitment types. In fact, we allow player one's other commitment types to be any countable set of finite automata including those which play mixed strategies. For example, if there is a strong Stackelberg action in the stage-game, and the set of player one's types is any set of finite automata that includes the simple type that plays the pure strong Stackelberg action in each period, then player one guarantees his highest IR payoff.

[8]By *equal patience*, we mean that the players share the same discount factor. There is also a literature on reputation effects in repeated games without discounting. See, for example, Cripps and Thomas (1995).

[9]Atakan and Ekmekci (2011) also present a reputation result for repeated games with LNCI and equally patient players. However, in that paper the stage game is an extensive-form game of perfect information as opposed to the simultaneous-move game that we assume here.

[10]For a precise definition of a strictly-dominant-action stage game, see Mailath and Samuelson (2006), Page 540.

ever, as in the case with equal discounting, under perfect monitoring a reputation result is obtained only in games with conflicting interests (see Schmidt (1993) and Cripps et al. (1996) for a generalization).

Although the results in repeated games with a less patient opponent are similar in spirit to the results we establish here, we should point out three important differences. First, against a less patient opponent, player one can build a reputation by mimicking a commitment type with positive shortfall, i.e., player one can guarantee a compromise payoff (Celentani et al. (1996) and Cripps et al. (1996)). In contrast, this is not possible when player one faces an equally patient opponent. Second, with equally patient agents, the limitation on the types that facilitate reputation building to those with no shortfall implies a restriction on the class of stage games (i.e., those with a strong Stackelberg action or LNCI). Again, this contrasts with the case where player one faces a less patient opponent, as in Celentani et al. (1996). Because player one can guarantee a compromise payoff against a less patient opponent, Celentani et al. (1996) are able to establish a reputation result which applies to all stage games when there is full-support imperfect monitoring. Third, the arguments for reputation results in repeated games where player one faces a less patient opponent rely on the learning result (Theorem 4.1) in Fudenberg and Levine (1992). In our framework with equally patient players, this learning result has no traction. We instead introduce a dynamic-programming methodology where the state variable is player two's beliefs.[11]

This paper is also closely related to Atakan and Ekmekci (2011), which proves a reputation result for repeated extensive-form games of perfect information with equally patient players. The three main differences between the two papers are as follows: First, in this paper we study the Bayesian equilibria of repeated simultaneous-move games whereas the focus of Atakan and Ekmekci (2011) is on the perfect Bayesian equilibria of a repeated game where the two players never move simultaneously. In particular, the reputation result of Atakan and Ekmekci (2011) leverages the particular form of sequential rationality, implied by perfect Bayesian equilibrium for games where the two players move sequentially, in a way that one cannot if the two players move simultaneously or if the focus is on Bayesian equilibria. Two, this paper assumes imperfect monitoring whereas Atakan and Ekmekci (2011) assumes that both players' moves are perfectly monitored. Three, here we assume that the other commitment types (i.e., the commitment types other than the type that player one mimics) are finite automata but we place no restriction on player two's prior. In contrast, the reputation result in Atakan and Ekmekci (2011) depends on the set of other commitment

[11] Also, see Cripps and Thomas (2003) for an asymptotic contrast of the equilibrium payoff sets of incomplete-information repeated games where the players share the same discount factor with those games where the informed player is arbitrarily more patient than his opponent.

types having sufficiently low prior probability.

2. The model

We consider an infinitely repeated game in which a finite, two-player, simultaneous-move stage game Γ is played in periods $t \in \{0, 1, 2, ...\}$. The players discount payoffs using a common discount factor $\delta \in [0, 1)$. For any set X, $\Delta(X)$ denotes the set of all probability distribution functions over X. The set of pure actions for player i in the stage game is A_i, and the set of mixed stage-game actions is $\Delta(A_i)$. After each period, player two's stage-game action is imperfectly observed through a public signal while player one's pure stage-game action is perfectly observed.[12] Let Y denote the set of public signals generated by player two's actions. Thus, after each period, a public signal $(a_1, y) \in A_1 \times Y$ is observed. The probability of signal y if player two chooses action $a_2 \in A_2$ is $\pi_y(a_2)$. For any mixed action $\alpha_2 \in \Delta(A_2)$, $\pi_y(\alpha_2) := \sum_{a_2 \in A_2} \alpha_2(a_2)\pi_y(a_2)$. We maintain the following full-support imperfect-monitoring assumption throughout the paper:

Assumption (**FS**) *Define* $\underline{\pi} := \min_{(a_2,y) \in A_2 \times Y} \pi_y(a_2)$*. We assume that* $\underline{\pi} > 0$.

If the stage game satisfies FS, then player one is never exactly sure about player two's action. The assumption does not, however, put any limits on the degree of imperfect monitoring.[13]

In the stage game, the payoff for any player i is given by the function $r_i : A_1 \times Y \to \mathbb{R}$ and depends only on publicly observed outcomes a_1 and y. Let $M = \max\{|r_i(a_1, y)| : i \in \{1, 2\}, a_1 \in A_1, y \in Y\}$. The payoff function for player i is $g_i(a_1, a_2) := \sum_{y \in Y} r_i(a_1, y)\pi_y(a_2)$ for $(a_1, a_2) \in A_1 \times A_2$. The mixed minimax payoff for player i is $\hat{g}_i$, and the pure minimax payoff for player i is $\hat{g}_i^p$. Let $a_1^p \in A_1$ be such that $g_2(a_1^p, a_2) \leq \hat{g}_2^p$ for all $a_2 \in A_2$. The set of feasible payoffs F is the convex hull of the set $\{g_1(a_1, a_2), g_2(a_1, a_2) : (a_1, a_2) \in A_1 \times A_2\}$; and the set of feasible and individually-rational payoffs is $G = F \cap \{(g_1, g_2) : g_1 \geq \hat{g}_1, g_2 \geq \hat{g}_2\}$. Let $\bar{g}_1 = \max\{g_1 : (g_1, g_2) \in G\}$; hence, $\bar{g}_1$ is player one's highest payoff compatible with the players' individual rationality (player one's highest IR payoff).

In the repeated game Γ^∞, the players have perfect recall and can observe past outcomes. The set of period t public histories is $H^t = A_1^t \times Y^t$, a typical element is $h^t = (a_1^0, y^0, a_1^1, y^1, ..., a_1^{t-1}, y^{t-1})$ for $t > 0$, and $h^0 = \varnothing$. The set of all public histories is $H =$

[12] If player one plays a mixed action, then only the pure action that he eventually chooses is observed publicly. The mixed action he uses is not observed.

[13] In extensive-form stage games, where player one's pure action is a full contingent plan, the perfect monitoring assumption that we impose is stringent. This is because it requires that player one's whole contingent plan be observed at the end of the period. We can relax this assumption by requiring that player one's moves are observed perfectly while player two's moves are observed with full-support noise. The results we present in this paper go through with this weaker assumption, and we discuss this further in section **??**.

$\bigcup_{t=0}^{\infty} H^t$. The set of period t private histories for player two is $H_2^t = A_1^t \times A_2^t \times Y^t$, a typical element is $h_2^t = (a_1^0, a_2^0, y^0, ..., a_1^{t-1}, a_2^{t-1}, y^{t-1})$, and $H_2 = \bigcup_{t=0}^{\infty} H_2^t$ is the set of all private histories for player two. The set of private histories of player one coincides with the public histories, i.e., $H_1^t = H^t$.

2.1. Types and strategies. A behavior strategy for player i is a function $\sigma_i : H_i \to \Delta(A_i)$, and Σ_i is the set of all behavior strategies for player i. A behavior strategy chooses a mixed stage-game action given player i's period t private history. A behavior strategy for player i is a function $\sigma_i : H_i \to \Delta(A_i)$ and Σ_i is the set of all behavior strategies for player i.[14] We use σ to denote a strategy profile $(\sigma_1(N), \sigma_2)$ and the set of all such strategy profiles is $\Sigma = \Sigma_1 \times \Sigma_2$.

For any strategy $\sigma_1 \in \Sigma_1$, $H(\sigma_1)$ denotes the set of public histories that are compatible with σ_1. More precisely, $h^T = (y^0, a_1^0, ..., y^{T-1}, a_1^{T-1}) \in H(\sigma_1)$ if and only if $a_1^k \in supp(\sigma_1(h^k))$ for all $k \leq T-1$, where h^k is any history that is identical to the first k periods of h^T. For any period t public history h^t and for any $\sigma_i \in \Sigma_i$, the expression $\sigma_i|_{h^t}$ denotes the continuation strategy induced by h^t. The probability measure over the set of (infinite) histories induced by $(\sigma_1, \sigma_2) \in \Sigma_1 \times \Sigma_2$ is $\Pr_{(\sigma_1, \sigma_2)}$.

Before time 0, nature selects player one as a normal type N or a commitment type ω, from an *at most countable* set of types $\Omega \subset \Sigma_1 \cup \{N\}$ according to a prior μ that is common knowledge. Each type $\omega \in \Omega \setminus \{N\}$ is committed to playing the repeated-game strategy $\omega \in \Sigma_1$. Player two is known to be a normal type with certainty and she maximizes her expected discounted payoffs. Player two's belief over player one's types, $\mu : H \to \Delta(\Omega)$, is a probability measure over Ω after each period t public history.

A *finite automaton* $\omega = (\Theta, \theta_0, o, \tau)$ consists of a finite set of states Θ, an initial state $\theta_0 \in \Theta$, an output function $o : \Theta \to \Delta(A_1)$ that assigns a (possibly mixed) stage-game action to each state, and a transition function $\tau : Y \times A_1 \times \Theta \to \Theta$ that determines the transitions across states as a function of the outcomes of the stage game. Abusing notation, we denote the strategy that an automaton induces by the automaton itself. For any finite automaton ω and any history $h^t \in H(\omega)$, $\theta(h^t)$ denotes the unique state θ which is the automaton's state at history h^t. A pure-strategy finite automaton is a finite automaton $\omega = (\Theta, \theta_0, o, \tau)$, where the output function o is deterministic. For a finite automaton ω, a state $\theta \in \Theta$ is *recurrent* if θ is visited infinitely often under the probability measure $\Pr_{(\omega, \sigma_2)}$ for any $\sigma_2 \in \Sigma_2$. A finite automaton is *irreducible* if all of its states are recurrent (see Definition A.1 in the appendix).

For any particular commitment type $\omega \in \Omega$, let $w(h^t) = \{\omega' : \omega'|_{h^t} = \omega|_{h^t}\}$; in words, $w(h^t)$ denotes the set of types that play the same repeated-game strategy as type ω plays

[14] For player one, any behavior strategy is also a public behavior strategy because $H_1 = H$.

after history h^t. Consequently, $\Sigma_1 \setminus \{\omega\}$ is the set of commitment types other than ω, and $\Sigma_1 \setminus \omega(h^t)$ is the set of commitment types that play a strategy that is not identical to the strategy of ω, given that history h^t has been reached.

Given automaton $\omega = (\Theta, \theta_0, o, \tau)$, we say that player two's strategy σ_2 is stationary with respect to ω if $\sigma_2(h^t) = \sigma_2(h^k)$ for any two histories h^t and h^k such that $\theta(h^k) = \theta(h^t) \in \Theta$, where $\theta(h^k) = \tau(a_1^{k-1}, y^{k-1}, \theta(h^{k-1}))$ and $\theta(h^0) = \theta_0$. Abusing notation slightly, we will denote a stationary strategy by a function $\sigma_2 : \Theta \to \Delta(A_2)$, i.e., player two plays mixed action $\sigma_2(\theta)$ whenever the state of ω is θ.

2.2. Payoffs. A player's repeated-game payoff is the normalized discounted sum of the stage-game payoffs. For any infinite public history h, define $u_i(h, \delta) = (1-\delta)\sum_{k=0}^{\infty} \delta^k r_i(a_1^k, y^k)$, and $u_i(h^{-t}, \delta) = (1-\delta)\sum_{k=t}^{\infty} \delta^{k-t} r_i(a_1^k, y^k)$, where $h^{-t} = (a_1^t, y^t, a_1^{t+1}, y^{t+1}, ...)$. Player one and player two's expected continuation payoffs, following a period t public history h^t and under strategy profile $\sigma = (\{\omega\}_{\omega \in \Omega \setminus \{N\}}, \sigma_1(N), \sigma_2)$, are given by the following two equations, respectively:

$$U_1(\sigma, \delta | h^t) = U_1(\sigma_1(N), \sigma_2, \delta | h^t),$$
$$U_2(\sigma, \delta, \mu | h^t) = \sum\nolimits_{\omega \in \Omega \setminus \{N\}} \mu(\omega | h^t) U_2(\omega, \sigma_2, \delta | h^t) + \mu(N | h^t) U_2(\sigma_1(N), \sigma_2, \delta | h^t),$$

where $U_i(\omega, \sigma_2, \delta | h^t) = \mathbb{E}_{(\omega, \sigma_2)}[u_i(h^{-t}, \delta) | h^t]$ is the expectation over continuation histories h^{-t} with respect to $\Pr_{(\omega|_{h^t}, \sigma_2|_{h^t})}$. Also, $U_1(\sigma, \delta) = U_1(\sigma, \delta | h^0)$ and $U_2(\sigma, \delta, \mu) = U_2(\sigma, \delta, \mu | h^0)$.

2.3. Repeated game and equilibrium. The repeated game of complete information, that is, the repeated game without any commitment types, with discount factor equal to $\delta \in [0, 1)$, is denoted as $\Gamma^\infty(\delta)$. The repeated game of incomplete information, with the prior over the set of commitment types given by $\mu \in \Delta(\Omega)$ and the discount factor equal to $\delta \in [0, 1)$, is denoted as $\Gamma^\infty(\mu, \delta)$.

The analysis in this paper focuses on Bayesian Nash equilibria (**NE**) of the game of incomplete information $\Gamma^\infty(\mu, \delta)$. In particular, a pair of strategies $(\sigma_1(N), \sigma_2) \in \Sigma_1 \times \Sigma_2$ is a NE of $\Gamma^\infty(\mu, \delta)$ if $\sigma_1(N) \in \arg\max_{\sigma_1 \in \Sigma_1} U_1(\sigma_1, \sigma_2, \delta)$ and $\sigma_2 \in \arg\max_{\sigma_2 \in \Sigma_2} U_2(\sigma_1(N), \sigma_2, \delta, \mu)$. Let

$$U_1^{NE}(\delta, \mu) = \inf\{U_1(\sigma, \delta) : \sigma \in NE(\Gamma^\infty(\delta, \mu))\},$$

where $NE(\Gamma^\infty(\delta, \mu))$ denotes the set of all NE of the repeated game $\Gamma^\infty(\delta, \mu)$. In words, $U_1^{NE}(\delta, \mu)$ is player one's the worst NE payoff. Also, let $U_1^{NE}(\mu) = \liminf_{\delta \to 1} U_1^{NE}(\delta, \mu)$. Again in words, $U_1^{NE}(\mu)$ is the worst NE payoff for a patient player one.

Remark 1 *Suppose σ is a NE strategy profile of $\Gamma^\infty(\mu, \delta)$.*

(i). FS implies that if $h^t \in H(N)$, then $\Pr_\sigma(h^t) > 0$, that is, if h^t is compatible with player one's strategy, then it has positive probability under σ. This is because, under FS, any finite sequence of signals has positive probability regardless of which strategy player two uses.

(ii). For any history $h^t \in H$, if h^t has positive probability under σ, that is, if $\Pr_\sigma(h^t) > 0$, then $(\sigma_1|_{h^t}, \sigma_2|_{h^t})$ is a NE profile of $\Gamma^\infty(\mu(h^t), \delta)$, where $\mu(h^t)$ is the posterior belief over player one's types given history h^t.

(iii). Consequently, if $h^t \in H(N)$, then $(\sigma_1(N)|_{h^t}, \sigma_2|_{h^t})$ is a NE profile of $\Gamma^\infty(\mu(h^t), \delta)$, i.e., $(\sigma_1(N)|_{h^t}, \sigma_2|_{h^t})$ is a NE profile of the continuation game.

2.4. Commitment payoff and shortfall of a strategy. The commitment payoff of a repeated-game strategy σ is the payoff that a patient player one can guarantee through public commitment to this strategy. The formal definition is as follows:

Definition (**Commitment Payoff**) *For any repeated-game strategy σ_1, define*

$$U_1^C(\sigma_1, \delta|h^t) = \min\{U_1(\sigma_1, \sigma_2, \delta|h^t) : \sigma_2 \in BR(\sigma_1, \delta)\},$$

where $BR(\sigma_1, \delta)$ denotes the set of best responses of player two to σ_1 in the repeated game of complete information $\Gamma^\infty(\delta)$. The commitment payoff of a repeated-game strategy σ_1 after history h^t is defined as $U_1^C(\sigma_1|h^t) = \liminf_{\delta\to 1} U_1^C(\sigma_1, \delta|h^t)$.[15]

The shortfall of a repeated-game strategy σ is the difference between the commitment payoff of the strategy and player one's highest IR payoff. The shortfall of a commitment type is an important concept in our analysis because, as we show, only those types with no shortfall can facilitate successful reputation building for player one. The formal definition is as follows:

Definition (**Shortfall**) *The shortfall of a repeated-game strategy σ_1 is defined as follows:*

$$d(\sigma_1) = \bar{g}_1 - \sup_{h^t \in H(\sigma_1)} U_1^C(\sigma_1|h^t).$$

A type ω has no shortfall if $d(\omega) = 0$, i.e., if the best commitment payoff among all histories for type ω is equal to player one's highest IR payoff.

[15] Although we define the commitment payoff using $\liminf_{\delta\to 1} U_1^C(\sigma_1, \delta|h^t)$, in the context of this paper the limit $\lim_{\delta\to 1} U_1^C(\sigma_1, \delta|h^t)$ exists.

If the shortfall of the commitment type ω is positive, then there is typically a range of feasible and individually-rational payoffs for player two, given that player one receives $U_1^C(\omega)$ (see figure 1).

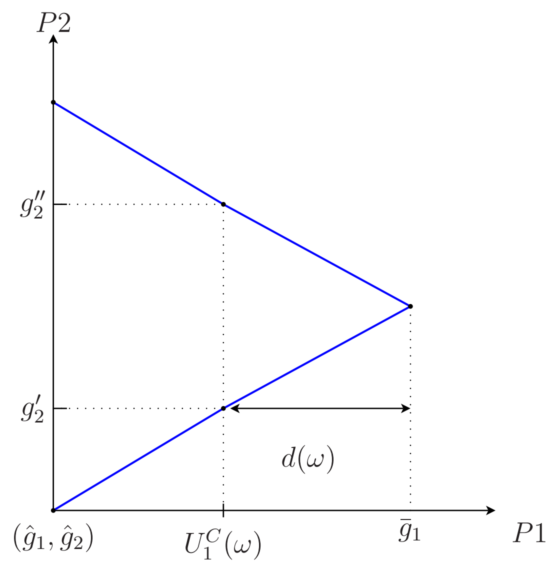


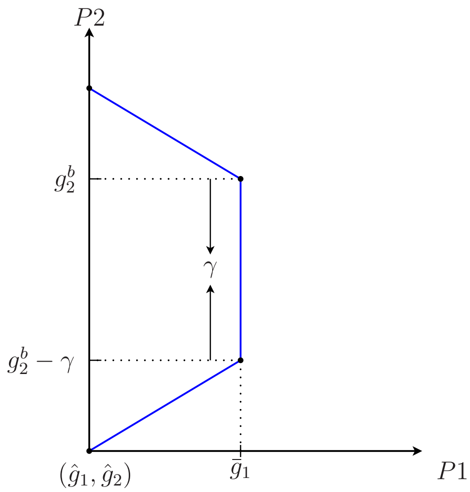
Figure 1: Shortfall of a strategy. Player two can receive any payoff between g_2' and g_2'' while player one receives $U_1^C(\omega)$.

2.5. Class of stage games. Below we define the various restrictions on the set of stage games that we will utilize in the remainder of the paper. We say a game has no gap if the payoff profile where player one receives his highest IR payoff is unique. The formal definition is as follows:

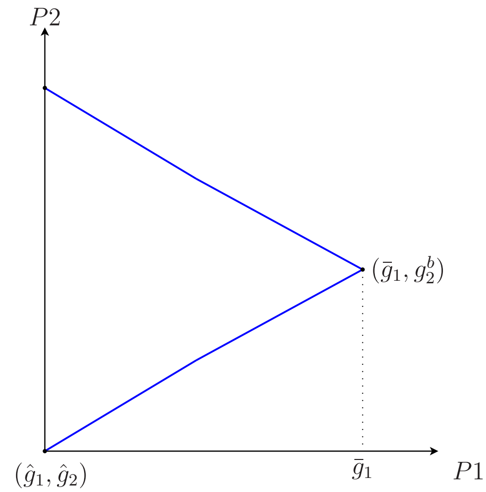
Definition (**No gap**) *Let* $g_2^b = \max\{g_2 : (\bar{g}_1, g_2) \in G\}$. *A stage game has no gap if* $(\bar{g}_1, g_2) \in G$ *implies that* $g_2 = g_2^b$. *Otherwise, we say that the stage game has a positive gap.*

For our reputation result we assume that the stage game has no gap, an assumption that is generically satisfied. The implication of the assumption is as follows: If the stage game has no gap, then there are *linear bounds* on the feasible payoffs for player two that pass through the point $(\bar{g}_1, g_2^b)$; hence, player two's payoffs are in a narrow range if player one's payoff is close to $\bar{g}_1$. In contrast, if the stage game has a positive gap, then there is a range of payoffs that are feasible and individually rational for player two if player one's payoff is equal to $\bar{g}_1$ (see figure 2).

Our main reputation result focuses on stage games that have either a strong Stackelberg action or LNCI, and we denote the set of such stage games by $\mathcal{G}$. A stage game has LNCI



(a) A game with a positive gap: there is a range of feasible payoffs that player two can receive while player one receives $\bar{g}_1$.



(b) A game with no gap: The set F is bounded by the lines that go through $(\bar{g}_1, g_2^b)$. Hence, player two's feasible payoffs are in a narrow range if player one's payoff is close to $\bar{g}_1$.

Figure 2: The gap of a game.

if player two's payoff is strictly higher than her pure strategy minimax in the payoff profile where player one receives his highest IR payoff. The formal definitions of a strong Stackelberg action and LNCI are as follows:

Definition (**LNCI**) *For any $g \in G$, if $g_1 = \bar{g}_1$, then $g_2 > \hat{g}_2^p$.*

Definition (**Strong Stackelberg action**) *There exists $a_1^s \in A_1$ such that any best response to a_1^s yields player one a payoff equal to $\bar{g}_1$.*

If player one has a strong Stackelberg action in Γ, then there is a pure strategy Nash equilibrium of Γ where player one plays the strong Stackelberg action a_1^s, player two best responds to a_1^s, and player one's payoff is equal to $\bar{g}_1$. The battle-of-the-sexes (figure 3d), the common-interest game (figure 3a), and the chain-store game (figure 3c) all have strong Stackelberg actions. The stage-game actions U, F, and A are strong Stackelberg actions for the battle-of-the-sexes, the common-interest game, and the chain-store game, respectively. In contrast, the principal-agent game (figure 3b) has LNCI but does not have a strong Stackelberg action. In this game, player one gets his highest IR payoff in the action profile (U, W). However, W is not a best response to U because player two would rather play S.

We will establish our main reputation result for stage games in $\mathcal{G}$ with no gap. The two

	L	R
U	$1, 0$	$1/2, -1$
D	$0, -1$	$1/2, -1$

(a) Common-interest game.

	$W(ork)$	$S(hirk)$
U	$3, 1$	$0, 2$
D	$0, 0$	$0, 0$

(b) Principal-agent game.

	$I(n)$	$O(ut)$
$F(ight)$	$-1, -2$	$4, 0$
$A(ccom.)$	$2, 1$	$4, 0$

(c) Chain-store game.

	A	B
A	$2, 1$	$0, 0$
B	$0, 0$	$1, 2$

(d) The battle-of-the-sexes.

Figure 3: Stage games with a strong Stackelberg action (3a, 3d, and 3c) or a game with LNCI but without a Stackelberg action (3b).

main implications of these restrictions, which we utilize heavily in proving our reputation result, are as follows: First, as we discussed above, if Γ has no gap, then player two's payoffs are in a narrow range whenever player one's payoff is close to $\bar{g}_1$. Second, if Γ is in $\mathcal{G}$, i.e., if Γ has a strong Stackelberg action or LNCI, then there is a type ω^* which has the following two properties:

First, ω^* has no shortfall, that is, ω^*'s commitment payoff is equal to player one's highest IR payoff. For example, in the battle-of-the-sexes (figure 3d), ω^* is the commitment type which plays A in each period. Playing A is player two's unique best response to ω^*, and hence ω^* is equal to player one's highest IR payoff. Second, the unit cost to a sufficiently patient player two of forcing a player one who is playing ω^* to receive a payoff less than $\bar{g}_1$ is strictly positive. For example, in the battle-of-the-sexes, in each period that player two forces ω^* to get a payoff of one (which is a unit short of $\bar{g}_1 = 2$) by playing B instead of A, she also loses a payoff equal to one.

Therefore, for stage games in $\mathcal{G}$ with no gap we have the following: *if player one's repeated-game payoff is close to the commitment payoff of ω^* (i.e., $\bar{g}_1$), then player two's feasible and individually rational repeated-game payoffs are in a narrow range determined by linear bounds that pass through $(\bar{g}_1, g_2^b)$*. Moreover, if player one is committed to playing strategy ω^*, then the unit cost to a patient player two of forcing him to receive a repeated-game payoff less than $\bar{g}_1$ is strictly positive for a patient player two.

The games that have a strong Stackelberg action are prominent in our analysis when all of player one's commitment types are finite automata. This is because if the stage game satisfies has a strong Stackelberg action, then there is a pure strategy finite automaton ω^* with no shortfall; moreover, choosing not to best respond to this commitment type is costly for player two. To see this, consider a game that has a strong Stackelberg action and the pure-strategy finite automaton that plays a_1^s in each period of the repeated game. It is straightforward to see that any best response to ω^* gives player one a payoff equal to $\bar{g}_1$, that is, ω^* has no shortfall. For example, in the battle-of-the-sexes (figure 3d), ω^* plays A in each period and player two's unique best response to ω^* entails playing A in each period. Moreover, choosing not to best respond to ω^* is strictly costly for player two. This is because if player two plays B instead of A in any period, then she gets zero instead of one against ω^*, i.e., the cost of choosing not to best respond is equal to one.

As we discussed above, there is a pure-strategy finite automaton with no shortfall if the stage game has a strong Stackelberg action. The following lemma, which is proved in appendix A, shows that the converse is also true: if the stage game does not have a strong Stackelberg action, then a pure-strategy finite automaton with no shortfall does not exist. Nevertheless, in Theorem 1 and in section 3.1 we show that there is an *infinite* automaton with no shortfall if the stage game is in $\mathcal{G}$.

Lemma 1 *Suppose that* Γ *satisfies FS and has no gap. There exists a pure strategy finite automaton with no shortfall if and only if* Γ *has a strong Stackelberg action.*

Proof: See Atakan and Ekmekci (2012), Lemma 1. □

For an intuition about the "only if" part of the above lemma, consider the principal-agent game (figure 3b). Player one's highest IR payoff is equal to three in this game. If player two's actions were observed without noise, then player one could obtain a payoff equal to three by using the following repeated-game strategy: player one starts the game by playing U; if player two does not play W in any period in which player one plays U, then player one punishes player two for two periods by playing D; after the two periods of punishment, player one again plays U. The best response of a sufficiently patient player two to this repeated-game strategy involves playing W in any period where player one plays U.

However, if player two's actions are monitored with noise, then for player one to commit to the strategy described in the previous paragraph does not necessarily guarantee him a high payoff. This is because player one cannot observe whether player two has played W or S when he plays U but can observe only an imperfect signal. Consequently, in certain periods player one will mistakenly punish player two, even if she played W against U; or he will mistakenly fail to punish player two, even if she played S against U. Thus, player one cannot guarantee a payoff equal to three. The situation is also similar with any other finite automaton. Any finite automaton ω whose commitment payoff is equal to three must punish player two by playing D if player two plays W against U. However, the finite automaton will punish player two even if player two plays W in each period because player two's actions are monitored with noise. Thus, player one's payoff from strategy ω will remain strictly below three even if player two plays W in each period. However, even though there is no finite automaton with no shortfall for the principal-agent game, in Theorem 1 and in section 3.1 we show that there is always an *infinite* automaton with no shortfall that facilitates reputation building if the stage game is in $\mathcal{G}$ and, consequently, for the principal-agent game.

3. Reputation effects

In this section we present our main reputation result. Recall that the set $\mathcal{G}$ contains all games that have a strong Stackelberg action or LNCI. Our main reputation result, which applies to stage games in $\mathcal{G}$ that have no gap, is as follows. The proof of this theorem is in appendix B.3.

Theorem 1 *Suppose that the stage game Γ is an element of $\mathcal{G}$, satisfies FS, and has no gap. There exists a commitment type ω^* such that if $\mu(\omega^*) > 0$ and if $\Omega_{-\omega^*}$ is a set of finite automata, then $U_1^{NE}(\mu) = U_1^C(\omega^*) = \bar{g}_1$.*

Under the stated assumption, Theorem 1 establishes that there exists a particular commitment type ω^* such that if this commitment type is available for player one to mimic (i.e., $\mu(\omega^*) > 0$) and if all the other commitment types are finite automata, then a patient player one can guarantee a payoff equal to the commitment payoff of ω^* in all NE. Moreover, the commitment payoff of ω^* is equal to player one's highest IR payoff. To establish Theorem 1, we use Lemma 2 stated below. This lemma, which is proved in appendix B.2, provides a lower bound on player one's NE payoffs as a function of the commitment payoff, the shortfall, and the prior probability of any irreducible pure-strategy finite automata.

Lemma 2 *Suppose that Γ satisfies FS and has no gap, and suppose that all the commitment types are finite automata. For any irreducible pure-strategy finite automaton $\omega \in \Omega$, if $\mu(\omega) > 0$, then*

$$U_1^{NE}(\mu) \geq U_1^C(\omega) - f(\omega, \mu(\omega))d(\omega),$$

where f is a positive-valued function as defined in equation (3) in the appendix, which satisfies $\lim_{x\to 0} f(\omega, x) = \infty$.

Proof: See Atakan and Ekmekci (2012), Theorem 1. □

To better understand Lemma 2, suppose that Γ satisfies FS and has no gap. Also, suppose that $\Omega = \{N, \omega^*\}$ where ω^* is an irreducible pure-strategy finite automaton. We will investigate the implications of Lemma 2 in two cases. First, suppose that the commitment type ω^*

has no shortfall (i.e., $d(\omega^*) = 0$ and therefore $U_1^C(\omega^*) = \bar{g}_1$). In this case, if ω^* is available (i.e., $\mu(\omega^*) > 0$), then Lemma 2 shows that player one can guarantee his highest IR payoff in any NE. In other words, Lemma 2 delivers a reputation result because it establishes that $U_1^{NE}(\mu^*) \geq U_1^C(\omega^*) = \bar{g}_1$ if ω^* is available and if $d(\omega^*) = 0$.

Now suppose that ω^* has a positive shortfall (i.e., $d(\omega^*) > 0$). In this case, the lower bound that Lemma 2 provides is vacuous if $\mu(\omega^*)$ is sufficiently small. This is because Lemma 2 implies only that $U_1^{NE}(\mu) \geq U_1^C(\omega^*) - f(\omega^*, \mu(\omega^*))d(\omega)$. However, if $\mu(\omega^*)$ goes to zero, then $f(\omega^*, \mu(\omega^*))$ approaches infinity and therefore, $U_1^C(\omega^*) - f(\omega^*, \mu(\omega^*))d(\omega)$ approaches negative infinity.

In summary, Lemma 2 delivers a reputation result (that is, the mere availability of type ω^* guarantees player one a high payoff in any NE) if ω^* has no shortfall and if Γ has no gap. Otherwise, Lemma 2 does not provide a meaningful lower bound on player one's NE payoff when the chosen commitment type is sufficiently unlikely.

As the above discussion suggests, Lemma 2 depends on the existence of a pure strategy finite automaton with no shortfall; in turn, the existence of such a finite automaton crucially depends on the properties of the stage game under consideration (Lemma 1). In particular, if the stage game has a strong Stackelberg action, then the commitment type ω^*, which plays a_1^s in every period of the repeated game, is a pure-strategy finite automaton with no shortfall. Therefore, Lemma 2 immediately delivers a reputation result for stage games with no gap that have a strong Stackelberg action: if all the commitment types are finite automata and if type ω^* is available, then player one can guarantee type ω^*'s commitment payoff which is equal to $\bar{g}_1$ in any NE. In other words, player one can guarantee his highest IR payoff if the set of commitment types is *sufficiently rich* that ω^* is available. For example, the set of types is sufficiently rich if all types which play the same action in every period are available or, more generally, if all the pure-strategy finite automata are available. The following corollary summarizes this:

COROLLARY 1 *Suppose that Γ satisfies FS, has a strong Stackelberg action, and has no gap; and suppose that all the commitment types are finite automata. Let ω^* denote the commitment type which plays a_1^s in each period of the repeated game. If $\mu(\omega^*) > 0$, then $U_1^{NE}(\mu) \geq \bar{g}_1$.*

PROOF: See Atakan and Ekmekci (2012), Corollary 1. □

For stage games that do not have a strong Stackelberg action, there is no pure-strategy finite automaton which has no shortfall (see Lemma 1). Therefore, Lemma 2 does not deliver a reputation result for such games. Nevertheless, our main reputation result, stated as Theorem 1, is for *all* stage games in $\mathcal{G}$ with no gap, and not just for those which satisfy SA. These findings are reconciled as follows: We establish the reputation result for stage games that do not have a strong Stackelberg action by first constructing a commitment type with *infinitely* many states that has no shortfall. We then show that player one can guarantee this type's commitment payoff if this particular type is available. In section 3.1 below we discuss how we use Lemma 2 as an intermediate step to prove a reputation result for stage games that do not have a strong Stackelberg action.

3.1. Games without a strong Stackelberg action. The proof of Theorem 1 shows that, for any stage-game in $\mathcal{G}$ with no gap which does not have a strong Stackelberg action, there exists a commitment type with no shortfall. In this case, however, the commitment type ω^* is an automaton with an *infinite* number of states. Moreover, Theorem 1 demonstrates that player one can guarantee a payoff equal to $\bar{g}_1$ by simply mimicking ω^*. In this section, we sketch how we construct this commitment type by describing ω^* for the principle-agent game which does not have a strong Stackelberg action (figure 3b).

As a first step in describing the infinite automaton ω^*, we describe a finite automaton ω^ϵ which plays a review strategy with shortfall $\epsilon > 0$ (see also Radner (1981, 1985) and Celentani et al. (1996)). The finite automaton ω^ϵ has two phases: a review phase and a punishment phase. Each review phase lasts for $J(\epsilon)$ periods and the automaton plays U in each period of the review phase. Each punishment phase lasts for $2J(\epsilon)$ periods and the automaton plays D in each period of the punishment phase. The automaton begins the game in the review phase. If player one's average payoff in a review phase is at least $3-\xi(\epsilon)$, where $\xi(\epsilon) > 0$ is the cutoff value for the review, then ω^ϵ enters a new review phase. Otherwise, ω^ϵ moves to a punishment phase and plays D, i.e., minimaxes player two, for $2J(\epsilon)$ periods. At the end of that punishment phase, the automaton again returns to a review phase.

Notice that, had there been perfect monitoring, a patient player two who faces ω^ϵ would have strictly preferred playing W in each period in order to avoid ever entering the punishment phase. Under imperfect monitoring a patient player two's incentives are similar to the case of perfect monitoring, but only for appropriately chosen $J(\epsilon)$ and $\xi(\epsilon)$. In particular, for any ϵ, we pick the length $J(\epsilon)$ of the review stage and the cutoff value $\xi(\epsilon)$ such that a sufficiently patient player two's best response to ω^ϵ entails entering the punishment phase after a review phase with arbitrarily small probability; a patient player one's repeated game payoff is thus at least $3-\epsilon$.[16] In other words, the commitment payoff of ω^ϵ is at least $3-\epsilon$.

The type ω^* first plays T_1 repetitions of a review strategy with shortfall ϵ where each repetition includes the review phase and, if it is triggered, the subsequent punishment phase. Then ω^* plays T_2 repetitions of the review strategy with shortfall $\epsilon/2$, and then T_n repetitions

[16] See Celentani et al. (1996) which shows that $J(\epsilon)$ and ξ can indeed be chosen in this way.

	$B(uy)$	$N(ot\ Buy)$
$H(igh)$	$2, -1$	$0, 0$
$L(ow)$	$1, 1$	$0, 0$

Figure 4: A product choice game.

of the review strategy with shortfall ϵ/n; and so on. As δ approaches one, the commitment payoff of type ω^* converges to three, i.e., the shortfall of ω^* is equal to zero. This is because, for any $n \geq 1$, the initial periods in which ω^* plays a review strategy with a shortfall more than ϵ/n become payoff-irrelevant as the discount factor approaches one.

The choice of how many repetitions T_n are played by ω^* of each review strategy with shortfall ϵ/n is delicate. In the appendix, we make the choices in a way that ensures that our reputation result applies. Intuitively, we choose the number of repetitions to ensure the following three conditions hold: first, $U_1^C(\omega^*, \delta)$ is increasing in δ; second, the cost of not best responding to this type is strictly positive for any δ; third, player two can distinguish the strategy of ω^* from any finite automaton's strategy regardless of which strategy she plays.

3.2. Games outside of the class $\mathcal{G}$. A stage game falls outside of the class $\mathcal{G}$ if the payoff profile in which player one receives his highest IR payoff is equal to his pure minimax payoff but the game does not have a strong Stackelberg Action. A prominent example of a game that falls outside of the class $\mathcal{G}$ is the product-choice game depicted in figure 4. In this game player one's highest IR payoff is equal to 1.5, player two receives her minimax payoff (zero) in the unique payoff profile in which player one gets 1.5, and the game has no gap. However, there is no action, whether pure or mixed, such that committing to it would guarantee player one a payoff equal to 1.5 in this game.[17] We discuss the repeated product-choice game to illustrate what can go wrong in games outside of the class $\mathcal{G}$.

This game is does not have a strong Stackelberg action; hence, we cannot obtain a reputation result with pure-strategy finite automata. In addition, an argument similar to Lemma 1 implies that any finite automata, including one that plays a mixed strategy, has a positive shortfall. However, consider the type ω that plays H with probability $1/2 + (1/2)^2$ in periods $\{1, ..., 4\}$, plays H with probability $1/2 + (1/2)^3$ in periods $\{5, ..., 8\}$, and more generally plays H with probability $1/2 + (1/2)^k$ in periods $\{2^{(k-1)} + 1, ..., 2^k\}$. Player two's unique best response to ω is to play B in each period. Moreover, the commitment payoff of this type $U_1^C(\omega)$ is equal to 1.5 (player one's highest IR payoff). Consequently, ω is an infinite

[17] Player one gets 1.5 if he plays H and L with equal probability and if player two best responds by playing B. However, both B and N are best responses for player two, and if she best responds by playing N instead of B, then player one's payoff is equal to zero. Therefore, player one cannot guarantee 1.5 by committing to play H and L with equal probability.

automaton with no shortfall.

Suppose that ω is the only commitment type available for player one to mimic. Even though ω has no shortfall, our reputation result does not apply here. This is because the type ω uses a mixed strategy. And, as we explained in the introduction, our approach is unable to provide a reputation bound for types that play mixed strategies.

Now, for the sake of argument, suppose that player one's stage-game action is to choose the probability $p \in [0,1]$ with which he plays H, and that player two observes his choice of p at the end of each period. Given this modification, ω plays a pure strategy.[18] Even under this modification, however, our reputation result still does not apply. This is because the cost to player two of playing action N instead of best responding to ω by playing B converges to zero as player two becomes increasingly patient. Consequently, a patient player two can resist playing a best response to ω at no cost to herself, and can thereby make it sufficiently difficult for player one to build a reputation.

Without this modification, one can also imagine a pure-strategy dynamic type that plays H in portion p of periods of a block of periods, plays L in the remaining periods of the block, and minimaxes player two for an appropriate number of periods if she fails to play B in many periods in the block. The length of the blocks and the length of the punishment periods can be carefully chosen to ensure that this dynamic type has a commitment payoff equal to 1.5. However, a patient player two can again resist best responding to this type at no cost to herself. Whether a reputation result can be established for this game is an open question.

A. Finite automata and learning

In this part of the appendix we prove some auxiliary results concerning finite automata which we repeatedly use in our subsequent arguments. Also, we prove our main learning result which we state as Lemma A.2. Our main learning result and its corollary, that we state as Corollary A.1, play central roles in the proofs of Lemma 2 and Theorem 1 presented in Appendix B.

Fix a pure strategy finite automata $\omega^* \in \Omega$ and a finite subset $W \subset \Sigma_1$. Consider a new finite set of states Θ, which is the product of the set of states of W and ω^* with typical element $\vec{\theta} = (\theta_{\omega^*}, \theta_1, ..., \theta_{|W|})$. In the following development, we fix player one's strategy ω^*, but the strategy of player 2, σ_2 varies. Notice that a period t public history h^t uniquely identifies the state $\vec{\theta}^t$ that the types are in at the start of period t. Let $a^*(\vec{\theta})$ denote the pure stage game action ω^* plays in state θ_{ω^*}. Every strategy profile (ω^*, σ_2) generates a stochastic

[18] A mixed strategy is then a probability distribution over choices of $p \in [0,1]$, i.e., a mixed strategy is an element of $\Delta([0,1])$.

process over the vector of states. In particular, the transition are given by the following equation:

$$\vec{\theta}^{t+1} = \Big(\tau(\omega, y, a^*(\vec{\theta}^t), \theta^t_\omega)\Big)_{\omega \in W \cup \{\omega^*\}} \equiv \vec{\tau}(y, \vec{\theta}^t).$$

Let $\Pr(\vec{\theta}^{t+n}|\vec{\theta}^t, h^t, \sigma_2)$ denote the probability that the state in period $t+n$ is equal to $\vec{\theta}^{t+n} \in \Theta$ given that the state in period t is equal to $\vec{\theta}^t$, the game is at history h^t and player two is using strategy σ_2. For example, when $n = 1$,

$$\Pr(\vec{\theta}^{t+1}|\vec{\theta}^t, h^t, \sigma_2) \equiv \sum\nolimits_{y \in \{y : \vec{\tau}(y, \vec{\theta}^t) = \vec{\theta}^{t+1}\}} \sum\nolimits_{a_2 \in A_2} \pi_y(a_2)\sigma_2(a_2, h^t)$$

In other words, $\Pr(\vec{\theta}^{t+1}|\vec{\theta}^t, h^t, \sigma_2)$ is the transition probability that governs the evolution of the states.

Definition A.1 (Recurrence and Transience) *A state $\vec{\theta}^*$ is* ***transient*** *if, given that the initial state is $\vec{\theta}^*$, there is a non-zero probability (in $\Pr_{(\omega^*, \sigma_2)}$) that the state $\vec{\theta}^*$ is never visited again. A state is* ***recurrent*** *if it is not transient. A subset of states $\Theta^j \subset \Theta$ is a* ***recurrent class*** *if for each $\vec{\theta}', \vec{\theta}'' \in \Theta^j$ there exists an $n > 0$ such that $\Pr(\vec{\theta}^{t+n} = \vec{\theta}'|\vec{\theta}^t = \vec{\theta}'', h^t, \sigma_2) > 0$ for all h^t, and for each $\vec{\theta}'' \in \Theta^j$ and $\vec{\theta}' \notin \Theta^j$ we have $\Pr(\vec{\theta}^{t+n} = \vec{\theta}'|\vec{\theta}^t = \vec{\theta}'', h^t, \sigma_2) = 0$ for all h^t and all $n > 0$. A subset of states $\Theta^0 \subset \Theta$ is a* ***transitory class*** *if each $\vec{\theta} \in \Theta^0$ is a transient state. A finite automaton is* ***irreducible*** *if its states form a single recurrent class.*

Lemma A.1 *Assume that Γ satisfies FS. For any $\vec{\theta}' \in \Theta^j$ and $n > 0$ we have $\Pr(\vec{\theta}^{t+n} = \vec{\theta}'|\vec{\theta}^t = \vec{\theta}(h^t), h^t, \sigma_2) > 0$ for some σ_2 and h^t if and only if $\Pr(\vec{\theta}^{k+n} = \vec{\theta}'|\vec{\theta}^k = \vec{\theta}(\hat{h}^k), \hat{h}^k, \sigma_2') > 0$ for all $k \geq 0$, σ_2', and all $\hat{h}^k$ such that $\vec{\theta}(\hat{h}^k) = \vec{\theta}(h^t)$. Consequently, the finite set of possible states Θ can be uniquely partitioned into a transitory class Θ^0, and a collection of disjoint recurrent classes Θ^j such that $\Theta = \cup_{i=0}^M \Theta^i$; and this partition is independently of σ_2.*

Proof: FS implies that the probability to transition from $\vec{\theta}^{t+1}$ to $\vec{\theta}^t$ after history h^t is bounded from below as follows:

$$\begin{aligned} \Pr(\vec{\theta}^{t+1}|\vec{\theta}^t, h^t, \sigma_2) &= \sum\nolimits_{y \in \{y : \vec{\tau}(y, \vec{\theta}^t) = \vec{\theta}^{t+1}\}} \sum\nolimits_{a_2 \in A_2} \pi_y(a_2)\sigma_2(a_2, h^t) \\ &\geq \sum\nolimits_{y \in \{y : \vec{\tau}(y, \vec{\theta}^t) = \vec{\theta}^{t+1}\}} \underline{\pi} = |\{y : \vec{\tau}(y, \vec{\theta}^t) = \vec{\theta}^{t+1}\}|\underline{\pi} \end{aligned}$$

So FS implies that $\Pr(\vec{\theta}^{t+1}|\vec{\theta}^t, h^t, \sigma_2) > 0$ if and only if $|\{y : \vec{\tau}(y, \vec{\theta}^t) = \vec{\theta}^{t+1}\}| \geq 1$. But if $|\{y : \vec{\tau}(y, \vec{\theta}^t) = \vec{\theta}^{t+1}\}| \geq 1$ then $\Pr(\vec{\theta}^{k+1}|\vec{\theta}^t(\hat{h}^k), \hat{h}^k, \sigma_2') \geq |\{y : \vec{\tau}(y, \vec{\theta}^t) = \vec{\theta}^{t+1}\}|\underline{\pi} \geq \underline{\pi} > 0$ for any $\hat{h}^k$ such that $\vec{\theta}(\hat{h}^k) = \vec{\theta}^k$. Iterating this argument generalizes the above to the case of $n > 1$. Also, see Billingsley (1995), Chapter 1, Section 8, or Stokey et al. (1989), Chapter 11.1 for more on partitioning the set of states. $\square$

DEFINITION A.2 (Speed of learning between ω^* and W) *Let*

$$\bar{p}(\omega^*, W) := \max_{\vec{\theta} \in \{p(\omega^*,\omega,\vec{\theta}^t) \neq 1, \omega \in W\}} p(\omega, \vec{\theta}),$$

where $p(\omega^*, \omega, \vec{\theta}^t) := o(\omega, a^*(\vec{\theta}^t), \theta^t_\omega)$*. That is,* $p(\omega^*, \omega, \vec{\theta}^t)$ *is the probability that type* ω *plays the same action as* ω^* *in state* $\vec{\theta}^t$ $(a^*(\vec{\theta}^t))$ *and* $\bar{p}(\omega^*, W)$ *is the maximum of* $p(\omega^*, \omega, \vec{\theta})$ *over the set of types* W *and the set of states where* $p(\omega, \omega^*, \vec{\theta})$ *differs from* 1*. Notice that* $\bar{p}(\omega^*, W) < 1$*.*[19]

Define the likelihood ratio recursively as $L^\omega_t(h) = p(\omega, \vec{\theta}^t(h))L^\omega_{t-1}(h)$ and let $L^\omega_0(h) = L^\omega_0 = \mu(\omega)/\mu(\omega^*)$. Hence, $L^\omega_t(h) = \mu(\omega|h^t)/\mu(\omega^*|h^t)$ and (L^ω_t, h^t) is a supermartingale under $\Pr_{(\omega^*,\sigma_2)}$ (Fudenberg and Levine (1992) Lemma 4.1). Also, let $L^W_0 = \mu(W)/\mu(\omega^*)$.

LEMMA A.2 *Assume that* Γ *satisfies* FS *and* $\bar{p}(\omega^*, W) \leq \xi \in [0, 1)$*. For any* $\epsilon > 0$ *and* $\phi > 0$ *there exists* $T(|\Theta|, |W|, \xi, \epsilon, \phi)$ *such that*

$$\Pr_{(\omega^*,\sigma_2)}\left\{h : \frac{\mu(W_{-\omega^*(h^t)}|h^t)}{\mu(\omega^*|h^t)} < \phi L^W_0\right\} > 1 - \epsilon,$$

for any $t > T(|\Theta|, |W|, \xi, \epsilon, \phi)$*, any* μ *such that* $\mu(\omega^*) > 0$*, and any strategy* σ_2 *of player two.*

PROOF: For any nonnegative integer k and even number l let $E(l, k)$ denote the set of infinite histories such that for any $h \in E(l, k)$ the process has entered a recurrent class $\Theta^i(h) \subset \Theta$ by period $l/2$ and all states $\vec{\theta} \in \Theta^i(h)$ have been visited at least k times by period l. For each nonnegative integer k and $\epsilon > 0$ there exists $l(k, |\Theta|, \epsilon)$ such that $\Pr_{(\omega^*,\sigma_2)}\{E(l(k, |\Theta|, \epsilon), k)\} > 1 - \epsilon$ for any σ_2. This is because FS implies that the probability the process enters a recurrent class after $|\Theta|$ periods is at least $\underline{\pi}^{|\Theta|}$. FS also implies that for any two states $\vec{\theta}'$ and $\vec{\theta}''$ in the same recurrent class Θ^j we have that $\Pr(\vec{\theta}^{t+|\Theta^j|} = \vec{\theta}'|\vec{\theta}^t = \vec{\theta}'', h^t, \sigma_2) > \underline{\pi}^{|\Theta^j|} \geq \underline{\pi}^{|\Theta|}$ for any h^t and any σ_2.

Let $k^* = K(|W|, \xi, \phi) = \frac{\ln \phi - \ln |W|}{\ln \xi}$. Pick l^* such that $\Pr_{(\sigma_1(\omega^*),\sigma_2)}\{E(l^*, k^*)\} > 1 - \epsilon$ for any σ_2. We show that choosing $T(|\Theta|, |W|, \xi, \epsilon, \phi) = l^*$ achieves the claim of the lemma.

Suppose that $h \in E(l^*, k^*)$. If $\omega \in W_{-\omega^*(h^{l^*})}$, then there exists a state $\vec{\theta}_\omega \in \Theta^i(h)$ such that $p(\omega, \vec{\theta}_\omega) \leq \bar{p}(\omega^*, W) < \xi < 1$. Because state $\vec{\theta}_\omega$ has been visited more than k^* times by time l^*,

[19] The maximum is well defined since W is a finite set and $\{\vec{\theta} : p(\omega, \vec{\theta}) \neq 1, \omega \in W\} \neq \emptyset$ because for each $\omega \in W \subset \Omega_{-\omega^*}$ there is a state such that $p(\omega, \vec{\theta}) \neq 1$

and because $L_t^{\omega}(h)$ is a supermartingale, we have $L_t^{\omega}(h) \leq \bar{p}(\omega^*, W)^{k^*} L_0^{\omega} \leq \xi^{k^*} L_0^{\omega} \leq \xi^{k^*} L_0^{W}$ for any $t \geq l^*$. Our initial choice of k^* implies that if $t > l^*$ and if $\omega \in W_{-\omega^*(h^l)} = W_{-\omega^*(h^t)}$, then $L_t^{\omega}(h) = \mu(\omega|h^t)/\mu(\omega^*|h^t) \leq \phi L_0^W/|W|$. Consequently, if $h \in E(l^*, k^*)$ and if $t > l^*$, then $\mu(W_{-\omega^*(h^t)}|h^t)/\mu(\omega^*|h^t) \leq \phi L_0^W$. Moreover, $\Pr_{(\omega^*, \sigma_2)}\{E(l^*, k^*)\} > 1 - \epsilon$ proving the result. □

COROLLARY A.1 *Assume that* Γ *satisfies FS. For any* μ *such that* $\mu(\omega^*) > 0$ *and* $\chi > 0$ *there exists* $T(\omega^*, \mu, \chi)$ *such that*

$$\Pr_{(\sigma_1(\omega^*), \sigma_2)}\{h : \mu(\Sigma_1 \setminus \omega^*(h^t)|h^t)/\mu(\omega^*|h^t) < \chi\} > 1 - \chi,$$

for any $t \geq T(\omega^*, \mu, \chi)$ *and any strategy* σ_2 *of player two.*

PROOF: Choose finite set $W \subset \Sigma_1$ such that $\mu(W) \geq 1 - \chi/2$. Set $\xi = \bar{p}(\omega^*, W)$, $\phi = \chi/2L_0^W$ and $\epsilon = \chi$. Notice that $|\Theta|$, $|W|$, L_0^W, and ξ depend only on μ (through the choice of the set W) and on ω^*. Observe that $\mu(\Sigma_1 \setminus \omega^*(h^t)|h^t)/\mu(\omega^*|h^t) \leq \mu(W \setminus \omega^*(h^t)|h^t)/\mu(\omega^*|h^t) + \chi/2$ for any h^t and apply Lemma A.2. □

LEMMA A.3 (Blackwell Optimality) *Suppose* Γ *satisfies FS and that* $\omega = (\Theta, \theta_0, o, \tau)$ *is an irreducible pure strategy finite automaton. For any* $\sigma_2 \in \Sigma_2$, *let*

$$\hat{U}_i(\omega, \sigma_2) = \limsup_{N \to \infty} \frac{1}{N} E_{(\omega, \sigma_2)} \sum_{t=0}^{N} g_i(a_1^t, a_2^t),$$

i.e., $\hat{U}_i(\omega, \sigma_2)$ *is player i's long-run average payoff. Let* $O_2 = \{\sigma_2 : \Theta \to A_2\}$ *denote the finite set of pure stationary strategies for player two. Let* $F(\omega, \delta) = co\{(U_1(\omega, \sigma_2, \delta), U_2(\omega, \sigma_2, \delta)) : \sigma_2 \in \Sigma_2\}$ *and* $F(\omega) = co\{(\hat{U}_1(\omega, \sigma_2), \hat{U}_2(\omega, \sigma_2)) : \sigma_2 \in \Sigma_2\}$.

(i). For any stationary strategy $\sigma_2 \in O_2$, $\lim_{\delta \to 1} U_i(\omega, \sigma_2, \delta) = U_i(\omega, \sigma_2) = \hat{U}_i(\omega, \sigma_2)$.

(ii). The set $F(\omega, \delta) = co\{(U_1(\omega, \sigma_2, \delta), U_2(\omega, \sigma_2, \delta)) : \sigma_2 \in O_2\}$ *and* $F(\omega) = co\{(U_1(\omega, \sigma_2), U_2(\omega, \sigma_2)) : \sigma_2 \in O_2\}$, *i.e, the finite set of vectors* $\{(U_1(\omega, \sigma_2), U_2(\omega, \sigma_2)) : \sigma_2 \in O_2\}$ *are extreme points of* $F(\omega)$.

(iii). There exist a $\delta^* \in (0, 1)$ *and a pure stationary strategy* $o_2 \in O_2$ *such that* $o_2 \in BR(\omega, \delta)$ *and* $U_1^C(\omega, \delta) = U_1(\omega, o_2, \delta)$ *for all* $\delta \in (\delta^*, 1)$.

(iv). Moreover, for any stationary σ_2, *we have* $|U_2(\omega, \sigma_2, \delta|h^t) - U_2(\omega, \sigma_2, \delta|h^k)| \leq K$, $|U_2(\omega, \sigma_2, \delta|h^t) - U_2(\omega, \sigma_2|h^k)| \leq K$, $|U_1^C(\omega, \delta|h^t) - U_1^C(\omega, \delta|h^k)| \leq K$, *and* $|U_1^C(\omega, \delta|h^t) - U_1^C(\omega|h^k)| \leq K$, *for any* $h^t, h^k \in H(\omega)$ *and any stationary* σ_2 *where* $K = (1 - \delta^{|\Theta^i|})M/(\delta \underline{\pi})^{|\Theta^i|}$.

PROOF: Part (i) follows from Bertsekas (2007), chapter 4, Proposition 1.2. Part (ii) follows from Dutta (1995), Lemma 1 because the sets $F(\omega, \delta)$ and $F(\omega)$ are the set of feasible payoffs for a discounted and undiscounted stochastic game, respectively, where the state space is Θ, the unique action available to player one in state θ is $o(\theta)$ and the transition function is τ.

For part (iii), first notice that against a fixed ω finding player two's best response is a standard discounted dynamic programming problem. Thus, a standard argument shows that a pure stationary best response exists (see Bertsekas (2007), Chapter 1). Also notice, if σ is a stationary strategy profile, then $U_i(\sigma, \delta|h^t) = U_i(\sigma, \delta|h^k)$ for any $h^t, h^k \in H(\omega)$ such that $\theta(h^t) = \theta(h^k)$. Let $U_2(\omega, \delta|h^t)$ denote player two's payoff after h^t given that she best responds to ω. Since a stationary best response exists $U_2(\omega, \delta|h^t) = U_2(\omega, \delta|h^k)$ for any $h^t, h^k \in H(\omega)$ such that $\theta(h^t) = \theta(h^k)$. Player one's commitment payoff is given by the following dynamic program: $U_1^C(\omega, \delta|h^t) = \min_{\alpha_2 \in \Delta(A_2)}(1-\delta)g_1(\omega, \alpha_2) + \delta E_y[U_1^C(\omega, \delta|h^t, \omega, y)|\alpha_2]$ subject to $(1-\delta)g_2(\omega, \alpha_2) + \delta E[U_2(\omega, \delta|\theta')|\theta, \alpha_2] = U_2(\omega, \delta|\theta)$, where $U_2(\omega, \delta|\theta)$ is player two's payoff in state θ given that she best responds. A standard argument shows that a pure stationary solution to this dynamic program exists. The existence of δ^* and o_2 follows from the existence of a Blackwell optimal policy in finite state and finite action dynamic programs. See Bertsekas (2007), Chapter 4, Proposition 2.2.

Part (iv). Lemma A.1 implies that $\Pr(\theta(h^{k+|\Theta|})|\theta(h^k), \sigma_2(h^k)) \geq \underline{\pi}^{|\Theta|}$ for any $\theta(h^t),\ \theta(h^k) \in \Theta$. Let history h^k be such that $U_i(\omega, \sigma_2, \delta|h^k) = \max_{\{h^l : \theta(h^l) \in \Theta\}} U_i(\omega, \sigma_2, \delta|h^l)$ and let history h^t be such that $U_i(\omega, \sigma_2, \delta|h^t) = \min_{\{h^l : \theta(h^l) \in \Theta\}} U_i(\omega, \sigma_2, \delta|h^l)$. We have the following two equations

$$U_i(\omega, \sigma_2, \delta|h^t) \geq -(1-\delta^{|\Theta|})M + \underline{\pi}^{|\Theta|}\delta^{|\Theta|}U_i(\omega, \sigma_2, \delta|h^k) + \delta^{|\Theta|}(1-\underline{\pi}^{|\Theta|})U_i(\omega, \sigma_2, \delta|h^t)$$
$$U_i(\omega, \sigma_2, \delta|h^k) \leq (1-\delta^{|\Theta|})M + \underline{\pi}^{|\Theta|}\delta^{|\Theta|}U_i(\omega, \sigma_2, \delta|h^t) + \delta^{|\Theta|}(1-\underline{\pi}^{|\Theta|})U_i(\omega, \sigma_2, \delta|h^k).$$

Solving delivers the result. The argument for $|U_1^C(\omega, \delta|h^t) - U_1^C(\omega, \delta|h^k)|$ is identical because there is a stationary strategy for player two that delivers player one his commitment payoff. Also, see Bertsekas (2007), Chapter 4, Proposition 1.2 for the remainder of the inequalities. □

B. Proof of Lemma 2 and Theorem 1

B.1. Preliminaries. Fix a stage game Γ that satisfies FS and that has no gap. Normalize payoffs such that $(\bar{g}_1, g_2^b) = (0, 0)$. For this game, there exists a finite constant $\rho \geq 0$ such that the following inequalities holds:

$$(1) \qquad g_2 \leq -\rho g_1, \text{ for any } (g_1, g_2) \in F; \text{ and } g_2 \geq \rho g_1, \text{ for any } (g_1, g_2) \in G.$$

In addition, if $\hat{g}_2 < 0$, then the following inequality also holds:

$$(2) \qquad g_2 \geq \rho g_1, \text{ for any } (g_1, g_2) \in F.$$

Fix a pure strategy finite automata $\omega^* = (\Theta^*, \theta^*, o^*, \tau^*)$. Let $\omega^*_\theta = (\Theta^*, \theta, o^*, \tau^*)$, that is, ω^*_θ is a pure strategy finite automaton which is identical to ω^* except that it may have a different initial state $\theta \in \Theta^*$. Recall that the set of commitment types is a countable subset of the set of repeated game strategies of player one Σ_1. For any $z \in (0,1]$ and $\phi \geq 0$, let $\Delta_{\omega^*,\phi,z}$ denote the set of all measures μ over $\{N\} \cup \Sigma_1$ with countable support such that $\mu(\omega^*) \geq z$ and $\mu(\Sigma_1 \setminus \{\omega^*\})/\mu(\omega^*) \leq \phi$, let $\Delta_{\omega^*,z}$ denote the set of all measures μ over $\{N\} \cup \Sigma_1$ with countable support such that $\mu(\omega^*) \geq z$, and let Δ_{ω^*} denote the set of all measures μ over $\{N\} \cup \Sigma_1$ with countable support such that $\mu(\omega^*) > 0$.

Definition B.1 *For any $z \in (0,1]$, $\phi \geq 0$, and $\delta \in [0,1)$ let $\underline{U}(\delta, \phi, z) = \min\{v(\delta, \phi, z), 0\}$ where*

$$v(\delta, \phi, z) = \inf\{U_1(\omega^*_\theta, \sigma_2, \delta) : \theta \in \Theta^*,\ \mu \in \Delta_{\omega^*_\theta,\phi,z},\ \sigma_2 \textit{ is part of a NE of } \Gamma^\infty(\delta, \mu)\}.$$

In words, $\underline{U}(\delta, \phi, z)$ is player one's worst payoff if he plays strategy ω^*_θ for some θ, if player two plays an equilibrium strategy, if the probability of ω^*_θ is at least z, and if the relative likelihood of the other commitment types is at most ϕ.

Lemma B.1 *Fix any $\delta \in [0,1)$, $\underline{z} \in (0,1]$, $\phi \geq 0$. Let $b \geq 0$ be a constant such that $U_1^C(\omega^*_\theta, \delta) > -b$ for all $\theta \in \Theta^*$. Define $\epsilon := \max\{b, (1-\delta^{|\Theta^*|})/(\delta\underline{\pi})^{|\Theta^*|}, \phi\}$. Assume that there exists constants $l(\omega^*) > 0$ and $K_1(\omega^*) > 0$ such that if $U_1(\omega^*_\theta, \sigma_2, \delta) \leq -b - x$ for any $x > 0$, then $U_2(\omega^*_\theta, \sigma_2, \delta) \leq -l(\omega^*)x + K_1(\omega^*)\epsilon$. If $z \geq \underline{z}$, then we have the following inequality:*

$$(3) \qquad \underline{U}(\delta, \phi, z) \geq -b - f(l(\omega^*), \underline{z})\epsilon, \textit{ where}$$

$$(4) \qquad f(l(\omega^*), \underline{z}) := \bar{K}^{\bar{n}},$$

$$(5) \qquad \bar{K}(l(\omega^*)) := \frac{1}{\underline{z} l \underline{\pi}} \max\{2l(\omega^*), 8\rho, 4M + l(\omega^*) + K_1 + \rho(5 + 8M)\},$$

$$(6) \qquad \bar{n}(\underline{z}, l(\omega^*)) := \textit{ the smallest integer } j \textit{ s.t. } \left(1 - \frac{\underline{\pi} l(\omega^*) \underline{z}}{4\rho}\right)^{j-1} < \underline{z}.$$

Proof: See Atakan and Ekmekci (2012). □

B.2. Proof of Lemma 2. We will first show, in the following lemma, that any irreducible finite automaton satisfies the hypothesis of Lemma B.1. Then we will then use Lemma B.1 and Corollary A.1 to establish Lemma 2.

Lemma B.2 (Unit cost lemma) *If ω^* is a irreducible finite automaton, then $U_1^C(\omega_\theta^*, \delta) \geq -b$ where $b = d(\omega^*) + M(1-\delta^{|\Theta^*|})/(\delta\underline{\pi})^{|\Theta^*|}$ for any $\theta \in \Theta^*$. Moreover, there exists a constant $l(\omega^*) > 0$ such that if $U_1(\omega_\theta^*, \sigma_2, \delta) \leq -b-x$ for some $\theta \in \Theta^*$ and $x > 0$, then $U_2(\omega_\theta^*, \sigma_2, \delta) \leq -lx + K_1\epsilon$ where $K_1 = \rho + lM + 1$.*

Proof: See proof of Lemma B.2 in Atakan and Ekmekci (2012). □

Notice that the bound in Lemma B.1 depends on both ϕ and δ. The learning result in Corollary A.1 implies that the likelihood of other commitment types becomes arbitrarily small if player one mimics type ω^* for a sufficiently long number of periods. To prove Lemma 2 we will use the learning result in Corollary A.1 and take the limit as δ goes to 1 to show that the bound in Lemma B.1 can be written independent of ϕ at the limit.

Proof of Lemma 2: See proof of Theorem 1 in Atakan and Ekmekci (2012). □

B.3. Proof of Theorem 1. For games that have a strong Stackelberg action, the proof of Theorem 1 follows immediately from Lemma 2 as described in the main text. In this section, we instead assume that the stage game Γ is in $\mathcal{G}$ but does not have a strong Stackelberg action.

In order to prove Theorem 1, we construct the commitment type ω^* which is an automaton with an infinite number of states with no shortfall. Recall that for a game that does not have a strong Stackelberg action there is no finite automaton with no shortfall. In constructing the infinite automaton ω^*, first we describe a finite automaton that we term a "review type" in the next section, second we establish a reputation bound for this review type (Lemma B.5) which is a strengthened version of Lemma B.1, third we construct type ω^* using an infinite sequence of review types, and finally we prove the bound for ω^* that is claimed in Theorem 1.

B.3.1. *Review types.* Here we describe a pure strategy finite automaton review type with shortfall at most ϵ which we denote as ω_ϵ. If a stage game has LNCI, then there exists a positive integer P and a positive constant $l > 0$ such that

$$(7) \qquad g_2(a_1^s, a_2) + Pg_2(a_1^p, a_2') < -Ml(P+1)$$

for any $a_2 \in A_2$ such that $g_1(a_1^s, a_2) < 0$ and $a_2' \in A_2$.

In the following we first consider a KJ-fold finitely repeated game $\Gamma^{KJ}(\delta)$.[20] We partition Γ^{KJ} into blocks of length J, $\Gamma^{J,k}$, $k = 1, ..., K$. Let u_i^k denote player i's time average payoff in

[20]This development closely follows Celentani et al. (1996), Lemma 4. Also, see the lemma's proof in that paper's appendix.

block $\Gamma^{J,k}$ and let $u_i^{KJ}(\delta)$ denote player i's discounted payoff in the KJ-fold finitely repeated game $\Gamma^{KJ}(\delta)$. Let σ_1^{KJ} be the following strategy: in block $\Gamma^{J,1}$ player one plays a_1^s in each period. We call a block where player one chooses to play a_1^s in each period a review phase. In the beginning of block $\Gamma^{J,2}$, player one reviews play in the previous block. If $u_1^1 \geq -\eta$, then player one again chooses to play a_1^s in each period of block $\Gamma^{J,2}$ and so on. If for any k, $u_1^k < -\eta$, then player 1 plays action a_1^p, for the next P repetitions of $\Gamma^{J,k}$ and then plays a_1^s in $\Gamma^{J,k+P+1}$. We call the blocks where player one chooses to play a_1^p in each period a "punishment phase".

Lemma B.3 *Given $\epsilon > 0$ there are numbers $\eta(\epsilon)$, $K(\epsilon)$, $J(\epsilon)$ and discount factor $\delta(\epsilon)$ such that for any $\delta > \delta(\epsilon)$ and for any best response σ_2^* to $\sigma_1^{K(\epsilon)J(\epsilon)}$ in $\Gamma^{KJ}(\delta)$ player one's discounted payoff $u_1^{K(\epsilon)J(\epsilon)}(\sigma_1^{K(\epsilon)J(\epsilon)}, \sigma_2^*, \delta) > -\epsilon$.*

Proof: This construction is directly taken from Celentani et al. (1996) Lemma 4. A proof can be found in the appendix of Celentani et al. (1996). □

Definition B.2 (Review type) *Let σ_1^* denote the repeated game strategy that infinitely repeats the strategy $\sigma_1^{K(\epsilon)J(\epsilon)}$, that is, σ_1^* plays according to $\sigma_1^{K(\epsilon)J(\epsilon)}$ in periods 1 through $K(\epsilon)J(\epsilon)$, then again plays according to $\sigma_1^{K(\epsilon)J(\epsilon)}$, in periods $K(\epsilon)J(\epsilon)+1$ through $2K(\epsilon)J(\epsilon)$ and so on. The type ω_ϵ is the finite automaton which implements σ_1^* with a minimal number of states.*

The following lemma is a strengthened version of Lemma B.2 which holds for *any* review type.

Lemma B.4 (Unit cost lemma for the review type) *For each $\epsilon > 0$, there exists $\delta_\epsilon \in [0, 1)$ such that for all $\delta > \delta_\epsilon$*

(i) $U_1^C(\omega_\epsilon, \delta) > -\epsilon$,

(ii) If $U_1(\omega_\epsilon, \sigma_2, \delta) = -\epsilon - r$ and $r > 0$, then $U_2(\omega_\epsilon, \sigma_2, \delta) \leq \rho\epsilon - lr$.

Proof: Pick $\delta_\epsilon > \delta(\epsilon)$ where $\delta(\epsilon)$ is the cutoff identified in Lemma B.3. Part (i) follows immediately from Lemma B.3. Proof of part (ii) is as follows: The fact that inequality (7) holds implies that there exists a $\delta^* < 1$ such that

$$(8) \qquad \sum_{t=0}^{J_\epsilon - 1} \delta^t g_2(a_1^s, a_2) + \sum_{t=J(\epsilon)}^{J(\epsilon)+J(\epsilon)P-1} \delta^t g_2(a_1^p, a_2') < -lMJ(\epsilon)(P+1)$$

for all $\delta > \delta^*$. Also, pick δ_ϵ to be strictly greater than δ^*, i.e., $\delta_\epsilon > \max\{\delta(\epsilon), \delta^*\}$. For public history $h^{t+J(\epsilon)-1} = \{a_1^0, y^0, a_1^1, y^1, ..., a_1^{t+J(\epsilon)-1}, y^{t+J(\epsilon)-1}\}$, let $i(^{t+J(\epsilon)-1}) = 1$, if $\sum_{j=t}^{t+J(\epsilon)-1} \delta^{j-t} g_1(a_1^j, y^j) <$

$-\eta(\epsilon)$ and period t is the start of a review stage; and $i(h^t) = 0$, otherwise. If $i(h^{t+J(\epsilon)-1}) = 1$, then player 1 receives at least $-M$ in period t through period $t + J(\epsilon) + J(\epsilon)P - 1$. Consequently, $U_1(\omega_\epsilon, \sigma_2, \delta) \geq -\eta(\epsilon) - J(\epsilon)(1+P)(1-\delta)M(\mathbb{E}_{\omega_\epsilon,\sigma_2}[\sum_{t=0}^{\infty} \delta^t i(h^t)])$. By construction $\eta(\epsilon) < \epsilon$ and so $(1-\delta)\mathbb{E}_{(\omega_\epsilon,\sigma_2)}[\sum_{t=0}^{\infty} \delta^t i(h^t)] \geq r/J(\epsilon)(1+P)M$. If $i(h^{t+J(\epsilon)-1}) = 1$, then player two receives a total discounted payoff of at most $-J(\epsilon)(P+1)l(1-\delta)$ for periods t through $t + J(\epsilon)(P+1) - 1$, if $\delta > \delta_\epsilon$ by equation (8). In any block where player one receives at least $-\eta(\epsilon)$, player two receives at most $\rho\eta(\epsilon) < \rho\epsilon$. Consequently, $U_2(\omega_\epsilon, \sigma_2) \leq \epsilon\rho - J(\epsilon)(1+P)l(1-\delta)\mathbb{E}_{(\omega_\epsilon,\sigma_2)}[\sum_{t=0}^{\infty} \delta^t i(h^t)] \leq \epsilon\rho - lr$, if $\delta > \delta_\epsilon$. □

B.3.2. *Reputation bound for review types.* In the following we establish a reputation bound, Lemma B.5, for the review type described above. The reputation bound is similar to Lemma 2 and the proof of the bound also uses Lemma B.1 and Lemma A.2 as the main building blocks.

DEFINITION B.3 *For any integer $n \geq 1$, define W^n as the set of all finite automaton which have fewer states than $\omega_{\epsilon/n}$ and define $W^{n,n} = \{W : W \subset W^n, |W| \leq n\}$ as the set of all subsets of W^n with cardinality not more than n.*

The following lemma presents the reputation bound for the review type. We use this lemma extensively in constructing ω^*.

LEMMA B.5 *Given $n \geq 1$ and $T \geq 1$, suppose that $\omega(n,T)$ is a finite automaton whose strategy coincides with $\omega_{\epsilon/n}$ after period T. There exists a $\delta(n,T) < 1$ such that for any $z > 0$, any $\mu \in \Delta_{\omega(n,T),z}$, any $\delta \geq \delta(n,T)$, any set $W \subset W^{n,n}$, and any $\sigma \in NE(\Gamma^\infty(\mu,\delta))$ the following inequality is satisfied:*

$$U_1(\sigma,\delta) > -2\epsilon/n - \frac{\epsilon/n + \mu(\Omega \setminus W)}{z} f(l,z),$$

where l is the constant given in Lemma B.4 and f is the function defined in Lemma B.1.

PROOF: Clearly, there is a cutoff δ such that for all discount factors that exceed this cutoff, the conclusions of Lemma B.4 hold for $\omega(n,T)$.

Notice that $\bar{p}(W^n, \omega_{\epsilon/n}) = \xi > 0$, this is because $\omega_{\epsilon/n} \notin W^n$ and because W^n is a compact set. Consequently, for any set $W \subset W^{n,n}$, $\bar{p}(W, \omega_{\epsilon/n}) \geq \xi$. Let N be such that

$$\Pr_{(\omega^*,\sigma_2)}\left\{h : \frac{\mu(W|h^t)}{\mu(\omega(n,T)|h^t)} \frac{\mu(\omega(n,T))}{\mu(W)} < \epsilon/n\right\} > 1 - \epsilon/n,$$

for any $t \geq N$. Such a N exists by Lemma A.2 and only depends on ϵ, T, and n. The result then follows from Lemma B.1 and is analogous to the proof of Lemma 2. □

In what follows, we drop the reference to l in $f(l,z)$ because the uniform cost l we use (as defined in Lemma B.4) for establishing the above bound for $\omega(n,T)$ is always the same for any n and any T.

B.3.3. *Constructing type ω^*.* The type ω^* starts by playing a strategy that coincides with the strategy of the review type with shortfall at most ϵ, i.e., ω_ϵ, for T_1 periods, then plays a strategy that coincides with the strategy of the review type with shortfall at most $\epsilon/2$ for T_2 periods, and plays a strategy that coincides with the strategy of the review type with shortfall at most ϵ/n for T_n periods, and so on. Therefore, this type is identified by a sequence of period lengths, $T_1, T_2, .., T_n, ...$ which we will pick recursively. We will also simultaneously pick a sequence of intervals of discount factors, $[\underline{\delta}_1, \bar{\delta}_1], [\underline{\delta}_2, \bar{\delta}_2], ...$, such that $\lim_{i\to\infty} \underline{\delta}_i = 1$.

Definition B.4 *$\Gamma^N(\delta,\mu)$ is a N period repeated game, where the stage game is Γ. The types in $\Gamma^N(\delta,\mu)$ belong to the set Ω_N that is obtained as follows: For every $\omega \in \Omega$, there exists a corresponding ω_N whose strategy coincides with the strategy of ω during the finitely repeated game, i.e., ω_N is the projection of the infinitely repeated game strategy ω on the first N periods. Moreover, the probability of ω_N in the beginning of the finitely repeated game is $\mu(\omega)$.*

Definition B.5 *For any $\xi > 0$, let $NE_\xi(\Gamma^N(\mu,\delta))$ denote the set of ξ Bayes-Nash equilibria of the finitely repeated game $\Gamma^N(\mu,\delta)$ (see Radner (1981)).*

For the following, recall that $\delta_{\epsilon/n}$ is the cutoff level of the discount factor that achieves the desiderata of Lemma B.4. (i.e., the $\delta_{\epsilon/n}$ in Lemma B.4 that exists for the review type with shortfall at most ϵ/n). Notice that $\delta(n,T) \geq \delta_{\epsilon/n}$.

Lemma B.6 *Suppose that $[\underline{\delta}, \bar{\delta}] \subset [\delta(n,T), 1)$. Then there exists a $\xi([\underline{\delta}, \bar{\delta}]) > 0$ and an integer $T^*([\underline{\delta}, \bar{\delta}])$ such that: for any $z > 0$, any $\mu \in \Delta_{\omega(n,T),z}$, any set $W \in W^{n,n}$, any $\delta \in [\underline{\delta}, \bar{\delta}]$, any $\xi \leq \xi([\underline{\delta}, \bar{\delta}])$, any $N \geq T^*([\underline{\delta}, \bar{\delta}])$, and any $\sigma \in NE_\xi(\Gamma^N(\mu,\delta))$ the following inequality is satisfied:*

$$U_1(\sigma,\delta) \geq -3\epsilon/n - \frac{\epsilon/n + \mu(\Omega \setminus W)}{z} f(z).$$

Proof: On the way to a contradiction, suppose that the lemma is not true. Then we can pick a convergent sequence of ξ_k, discount factors, sets of finite automata, priors, and strategy profiles $\{\xi_k, \delta_k, \mu_k, W_k, z_k, \sigma_k\}_{k=1}^\infty$ with $\xi_k > 0$ and $\lim_{k\to\infty} \xi_k = 0$, $\delta_k \in [\underline{\delta}, \bar{\delta}]$, $\mu_k \in \Delta_{\omega(n,T),z_k}$,

$W_k \in W^{n,n}$ and $\sigma_k \in NE_{\xi_k}(\Gamma^k(\mu_k, \delta_k))$ such that

$$U_1(\sigma_k, \delta_k) < -3\epsilon/n - f(z_k)\frac{\epsilon/n + \mu_k(\Omega \setminus W_k)}{z_k}.$$

Let the limit of the sequence be $\{0, \delta, \mu, W, z, \sigma\}$ satisfying $\delta \in [\underline{\delta}, \bar{\delta}]$ and $\mu \in \Delta_{\omega(n,T),z}$.[21] We have $\sigma \in NE(\Gamma^\infty(\mu, \delta))$ since $\lim_{k\to\infty} \xi_k = 0$ and all other terms converge to a limit.[22]

We will now consider two cases: First, if $z = 0$, then the right hand side of the displayed inequality in the lemma will be arbitrarily small because $f(0) = \infty$. However $U_1(\sigma_k, \delta_k)$ is bounded below a finite number since player 1's minimax is a finite number. So it cannot be that $z = 0$. Second, if $z > 0$, then $U_1(\sigma, \delta) \leq -3\epsilon/n - f(z)\frac{\epsilon/n+\mu(\Omega\setminus W)}{z}$ which contradicts Lemma B.5. □

Choosing T_1 and the interval $[\underline{\delta}_1, \bar{\delta}_1]$. We pick $\underline{\delta}_1 > \delta(1, 0)$ and $\bar{\delta}_1 > \delta_{\epsilon/2}$. Hence, the interval $[\underline{\delta}_1, \bar{\delta}_1]$ satisfies the hypothesis of Lemma B.6. By Lemma B.6 there exists a $\xi > 0$ and integer T^* such that for any $z > 0$, any $\mu \in \Delta_{\omega(1,0),z}$, any set $W \in W^{1,1}$, any $\delta \in [\underline{\delta}_1, \bar{\delta}_1]$, any $T \geq T^*$, and any $\sigma \in NE_\xi(\Gamma^T(\mu, \delta))$ the following inequality is satisfied:

$$U_1(\sigma, \delta) \geq -3\epsilon - \frac{\epsilon + \mu(\Omega \setminus W)}{z}K^{\bar{n}}.$$

We pick T_1 so that $T_1 \geq T^*$ and $\bar{\delta}_1^{T_1} M \leq \min\{\epsilon, \xi\}$. Consequently, we have the following:

Remark 2 *Let ω be an infinitely repeated game strategy that coincides with $\omega(1, 0)$ during the first T_1 periods. We claim that for any $\delta \in [\underline{\delta}_1, \bar{\delta}_1]$, any $W \in W^{1,1}$, any $z > 0$, any $\mu \in \Delta_{\omega,z}$ and any σ that is a NE profile of $\Gamma^\infty(\mu, \delta)$*

$$U_1(\sigma, \delta) \geq -4\epsilon - \frac{\epsilon + \mu(\Omega \setminus W)}{z}K^{\bar{n}}.$$

Proof: Let σ_{T_1} be the projection of σ on the first T_1 periods. Since $\bar{\delta}_1^{T_1} M \leq \xi$, σ_{T_1} is a ξ Bayes Nash equilibrium of $\Gamma^{T_1}(\mu, \delta)$. Therefore $U_1(\sigma_{T_1}, \delta) > -3\epsilon - \frac{\epsilon+\mu(\Omega\setminus W)}{z}f(z)$ by Lemma B.6. We now use the inequality $\bar{\delta}_1^{T_1} \leq \epsilon$ to argue that $|U_1(\sigma, \delta) - U_1(\sigma_{T_1}, \delta)| \leq \bar{\delta}_1^{T_1} M \leq \epsilon$ and we conclude that $U_1(\sigma, \delta) \geq -4\epsilon - \frac{\epsilon+\mu(\Omega\setminus W)}{z}f(z)$. □

Now we choose T_n and the interval $[\underline{\delta}_n, \bar{\delta}_n]$ given $\{T_1, ..., T_{n-1}\}$ and $\{[\underline{\delta}_1, \bar{\delta}_1], ..., [\underline{\delta}_{n-1}, \bar{\delta}_{n-1}]\}$. Let $\hat{T}_n = \sum_{i=1}^{n-1} T_i$. For $n = 1$, define the strategy $D_{\epsilon/n} := \omega_\epsilon$ and for $n > 1$, define $D_{\epsilon/n}$

[21] We use the Euclidean distance for ξ_k, δ_k, and **inherent product topology** for convergence of μ_k and the strategies σ_k.

[22] This is standard see for instance Myerson (1991) page 144, Theorem 3.4 or Fudenberg and Levine (1986) Proposition 4.1.

recursively as follows: $D_{\epsilon/n}$ coincides with $D_{\epsilon/(n-1)}$ up to a time $\hat{T}_n$ and then coincides with $\omega_{\epsilon/n}$.

LEMMA B.7 *Suppose that* $[\underline{\delta}, \bar{\delta}] \subset [\max\{\delta_{\epsilon/n}, \delta(n-1, \hat{T}_{n-1})\}, 1)$. *Then there exists a* $\xi([\underline{\delta}, \bar{\delta}]) > 0$ *and an integer* $T^*([\underline{\delta}, \bar{\delta}])$ *such that: for any* $z > 0$, *any* $\mu \in \Delta_{D_{\epsilon/n}, z}$, *any set* $W \in W^{n-1,n-1}$, *any* $\delta \in [\underline{\delta}, \bar{\delta}]$, *any* $\xi \leq \xi([\underline{\delta}, \bar{\delta}])$, *any* $T \geq T^*([\underline{\delta}, \bar{\delta}])$, *and any* $\sigma \in NE_\xi(\Gamma^T(\mu, \delta))$ *the following inequality is satisfied:*

$$U_1(\sigma, \delta) \geq -3\epsilon/(n-1) - \frac{\epsilon/(n-1) + \mu(\Omega \setminus W)}{z} f(z).$$

PROOF: This argument is similar to the argument for Lemma B.6. We again obtain a contradiction to Lemma B.5. We arrive at the contradiction by using the facts that for all $\delta \geq \max\{\delta_{\epsilon/n}, \delta(n-1, \hat{T}_{n-1})\}$ we have first $U_1^C(D_{\epsilon/n}, \delta) > -\epsilon/(n-1)$, and second if $U_1(D_{\epsilon/n}, \sigma_2, \delta) = -\epsilon/(n-1) - r$ and $r > 0$, then $U_2(D_{\epsilon/n}, \sigma_2, \delta) \leq \rho\epsilon/(n-1) - lr$. In other words the conclusions of Lemma B.4 hold and thus Lemma B.5 applies for $D_{\epsilon/n}$. □

We pick interval of discount factors $[\underline{\delta}_n, \bar{\delta}_n]$ *as follows:* Let $\underline{\delta}_n$ be such that $\underline{\delta}_n > \bar{\delta}_{n-1}$ and $\underline{\delta}_n > \delta(n, \hat{T}_n)$, and for the upper end, $\bar{\delta}_n > \underline{\delta}_n$, $\bar{\delta}_n > \delta_{\epsilon/(n+1)}$. Notice that the interval $[\underline{\delta}_n, \bar{\delta}_n]$ satisfies the hypothesis of Lemma B.6 by construction because $\underline{\delta}_n > \delta(n, \hat{T}_n)$. Also, the interval $[\bar{\delta}_{n-1}, \bar{\delta}_n]$ satisfies the hypothesis of Lemma B.7 by construction because $\bar{\delta}_{n-1} \geq \max\{\delta_{\epsilon/n}, \delta(n-1, \hat{T}_{n-1})\}$.

We now pick T^n. Let $\xi^* > 0$ be the cutoff $\xi([\bar{\delta}_{n-1}, \bar{\delta}_n])$ obtained in Lemma B.7, $\xi^{**} > 0$ be the cutoff $\xi([\underline{\delta}_n, \bar{\delta}_n])$ obtained in Lemma B.6, and let $\xi_n := \min\{\xi^*, \xi^{**}, \epsilon/n)\} > 0$. First, Lemma B.6 implies that there exists T^{**} such that for all $\delta \in [\underline{\delta}_n, \bar{\delta}_n]$, all $N \geq T^{**} + \hat{T}_n$, all $W \in W^{n,n}$, all $\mu \in \Delta_{D_{\epsilon/n}, z}$ and all $\sigma_N \in NE_{\xi_n}(\Gamma^N(\mu, \delta))$

$$U_1(\sigma_N, \delta) > -3\epsilon/n - f(z)\frac{\epsilon/n + \mu(\Omega \setminus W)}{z}. \tag{9}$$

Second, Lemma B.7 implies that there exists T^{***} such that for all $\delta \in [\underline{\delta}_{n-1}, \bar{\delta}_n]$, all $N \geq T^{***} + \hat{T}_n$, all $W \in W^{n-1,n-1}$, all $\mu \in \Delta_{D_{\epsilon/n}, z}$ and all $\sigma_N \in NE_{\xi_n}(\Gamma^N(\mu, \delta))$

$$U_1(\sigma_N, \delta) > -3\epsilon/(n-1) - f(z)\frac{\epsilon/(n-1) + \mu(\Omega \setminus W)}{z}. \tag{10}$$

We pick T_n such that $T_n \geq \max\{T^{**}, T^{***}\}$ and $\bar{\delta}_n^{\hat{T}_n + T_n} M < \min\{\xi_n, \epsilon/n\}$.

LEMMA B.8 *Suppose that* ω *that coincides with* $D_{\epsilon/n}$ *during the periods zero through* $\hat{T}_n + T_n$. *For all* $W \in W^{n-1,n-1}$, *all* $z > 0$, *all* $\mu \in \Delta_{\omega, z}$, *all* $\delta \in [\underline{\delta}_{n-1}, \bar{\delta}_n]$, *and all* $\sigma \in NE(\Gamma^\infty(\mu, \delta))$, *we have*

$$(11) \qquad U_1(\sigma,\delta) > -4\epsilon/(n-1) - f(z)\frac{\epsilon/(n-1)+\mu(\Omega\setminus W)}{z}.$$

Proof: Let σ be a NE of $\Gamma^\infty(\mu,\delta)$ for some $\delta \in [\underline{\delta}_{n-1}, \bar{\delta}_n]$. Our choice of T_n was such that $\bar{\delta}_n^{\hat{T}_n+T_n}M < \min\{\xi_n, \epsilon/n\}$. Therefore, if $\delta \le \bar{\delta}_n$, then the projection of σ on the first $N = \hat{T}_n + T_n$ periods, σ_N is a ξ_n Bayes-Nash equilibrium of $\Gamma^N(\mu,\delta)$. Therefore, inequalities (9) and (10) together imply that $U_1(\sigma_N,\delta) > -3\epsilon/(n-1) - f(z)\frac{\epsilon/(n-1)+\mu(\Omega\setminus W)}{z}$. However, because $\bar{\delta}_n^N M < \epsilon/n$, the payoffs after period N affect player 1's payoffs by at most ϵ/n as long as $\delta \le \bar{\delta}_n$. Hence, $U_1(\sigma,\delta) \ge U_1(\sigma_N,\delta) - \epsilon/(n-1)$. □

B.3.4. *Completing the proof of Theorem 1.*

Proof: We show that if $\mu \in \Delta_{\omega^*,z}$, $z > 0$, and all the commitment types other than ω^* in the support of μ are finite automata, then, $U_1^{NE}(\mu) = 0$. Fix any $\chi > 0$. There exists an $n > 1$ and a $W \in W^{n-1,n-1}$ such that for every $n' \ge n$, $\frac{4\epsilon}{n'-1} + f(z)\frac{\epsilon/(n'-1)+\mu(\Omega\setminus W)}{z} < \chi$. This follows from the fact that Ω is a countable set of finite automata. Also, by Lemma B.8, $U_1(\sigma,\delta) \ge -\frac{4\epsilon}{n-1} - f(z)\frac{\epsilon/(n-1)+\mu(\Omega\setminus W)}{z}$, for any $\delta \ge \underline{\delta}_n$, and any $\sigma \in NE(\Gamma^\infty(\mu,\delta))$. Therefore, we have the following inequalities:

$$\begin{aligned}\lim_{\hat{\delta}\to 1} \inf_{\delta \ge \hat{\delta}, \sigma \in NE(\Gamma^\infty(\mu,\delta))} U_1(\sigma,\delta) &\ge \inf_{\delta \ge \underline{\delta}_n, \sigma \in NE(\Gamma^\infty(\mu,\delta))} U_1(\sigma,\delta) \\ &\ge -\frac{4\epsilon}{n-1} - f(z)\frac{\epsilon/(n-1)+\mu(\Omega\setminus W)}{z} \ge -\chi.\end{aligned}$$

Since χ is arbitrary, $\liminf_{\delta\to 1, \sigma\in NE(\Gamma^\infty(\mu,\delta))} U_1(\sigma,\delta) = U_1^{NE}(\mu) = 0$. □

References

Aoyagi, M. (1996): "Reputation and Dynamic Stackelberg Leadership in Infinitely Repeated Games," *Journal of Economic Theory*, 71, 378–393.

Atakan, A. and M. Ekmekci (2011): "Reputation in Long-run Relationships," *The Review of Economic Studies*, Forthcoming.

——— (2012): "Reputation in the Long-run with Imperfect Monitoring," Working paper.

Bertsekas, D. (2007): *Dynamic programming and optimal control, vol. II.*

Billingsley, P. (1995): *Probability and Measure.*

Celentani, M., D. Fudenberg, D. Levine, and W. Pesendorfer (1996): "Maintaining a Reputation Against a Long-Lived Opponent," *Econometrica*, 64, 691–704.

Chan, J. (2000): "On the Non-Existence of Reputation Effects in Two-Person Infinitely Repeated Games," Working paper.

Cripps, M., E. Dekel, and W. Pesendorfer (2005): "Reputation with Equal Discounting in Repeated Games with Strictly Conflicting Interests," *Journal of Economic Theory*, 121, 259–272.

Cripps, M., K. Schmidt, and J. Thomas (1996): "Reputation in perturbed repeated games," *Journal of Economic Theory*, 69, 387–410.

Cripps, M. and J. Thomas (1995): "Reputation and Commitment in Two-Person Repeated Games without Discounting," *Econometrica*, 1401–1419.

——— (1997): "Reputation and Perfection in Repeated Common Interest Games," *Games and Economic Behavior*, 18, 141–158.

——— (2003): "Some Asymptotic Results in Discounted Repeated Games of One-Sided Incomplete Information," *Mathematics of Operations Research*, 433–462.

Dutta, P. (1995): "A folk theorem for stochastic games," *Journal of Economic Theory*, 66, 1–32.

Evans, R. and J. Thomas (1997): "Reputation and Experimentation in Repeated Games With Two Long-Run Players," *Econometrica*, 65, 1153–1173.

Fudenberg, D. and D. Levine (1986): "Limit Games and Limit Equilibria," *Journal of Economic Theory*, 38, 261–279.

——— (1989): "Reputation and Equilibrium Selection in Games with a Patient Player," *Econometrica*, 57, 759–778.

——— (1992): "Maintaining a Reputation when Strategies are Imperfectly Observed," *Review of Economic Studies*, 59, 561–579.

Gossner, O. (2011): "Simple bounds on the value of a reputation," *Econometrica*, forthcoming.

Mailath, G. and L. Samuelson (2006): *Repeated Games and Reputations: Long-run Relationships*, Oxford, UK: Oxford University Press.

Myerson, R. (1991): *Game Theory: Analysis of Conflict*, Cambridge, MA: Harvard University Press.

Radner, R. (1981): "Monitoring cooperative agreements in a repeated principal-agent relationship," *Econometrica*, 1127–1148.

——— (1985): "Repeated principal-agent games with discounting," *Econometrica*, 1173–1198.

Schmidt, K. (1993): "Reputation and Equilibrium Characterization in Repeated Games with Conflicting Interests," *Econometrica*, 61, 325–351.

Stokey, N., R. Lucas, and E. Prescott (1989): *Recursive Methods in Economic Dynamics*, Harvard University Press.

Vohra, R. (2005): *Advanced Mathematical Economics*, Routledge Advanced Texts in Economics and Finance.