

REPUTATION IN THE LONG-RUN WITH IMPERFECT MONITORING¹

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 games with locally nonconflicting interests or games with strictly conflicting interests: 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. In contrast, if the type's commitment payoff is strictly less than player one's highest payoff consistent with the players' individual rationality, then the worst perfect Bayesian equilibrium payoff for a patient player one is equal to his minimax payoff.

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

¹We 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	1
1.1	Related literature and our contribution.	4
2	The model	6
2.1	Types and strategies.	7
2.2	Payoffs.	8
2.3	Repeated game and equilibrium.	8
2.4	Commitment payoff and shortfall of a strategy.	9
2.5	Class of stage games.	10
3	Reputation effects	15
3.1	The reputation result for the repeated common-interest game.	17
3.1.1	Verbal intuition.	17
3.1.2	Sketch of the proof.	18
3.2	Other commitment types.	24
3.3	Reputation effects in stage games that do not satisfy SA.	24
4	Nonreputation results	25
4.1	The intuition for Theorem 2 and Lemma 3.	28
4.2	Sketch of the construction.	29
5	Discussion	32
5.1	Full-support imperfect monitoring.	32
5.2	Imperfect monitoring in extensive-form stage games.	33
5.3	Games outside of the class \mathcal{G}	34
A	Finite automata, learning, and Lemma 1	36
A.1	Proof of Lemma 1.	40
B	Proof of Lemma 2 and Theorem 1	41
B.1	Preliminaries.	41
B.2	The reputation bound.	41
B.3	Proof of Lemma 2.	50
B.4	Proof of Theorem 1.	52
B.4.1	Review types.	52
B.4.2	Reputation bound for review types.	54

B.4.3	Constructing type ω^* .	55
B.4.4	Completing the proof of Theorem 1.	58
C	Proofs of Theorem 2 and Lemma 3	58
C.1	Notation.	59
C.2	Review strategy profiles.	59
C.2.1	The description of a review strategy profile with one review phase.	60
C.2.2	The description of an n-step review strategy profile.	65
C.3	Completing the proof of Theorem 2.	69
C.4	Dispensing with irreducibility.	70
C.5	Proof of Lemma 3.	70
C.6	Proof of the identity given by Equation 21.	71

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

¹See 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 either *locally nonconflicting interests* (LNCI) or *strictly conflicting interests* (SCI).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. There are SCI in a game if player one has an action (a Stackelberg action) such that any best response to this action yields player one his highest IR payoff and yields player two her minimax payoff.⁵ These restrictions on the stage game ensure the existence of a commitment type that satisfies the aforementioned two properties.

We turn next to the question of whether player one can still benefit from such a reputation even if some of the assumptions of our main reputation result are violated. For this analysis, we restrict attention to commitment types that play repeated-game strategies with limited complexity (i.e., finite automata).⁶ In particular, suppose that, in contrast to our no-shortfall assumption, the shortfall for the only available commitment type is positive (i.e., the type’s commitment payoff is less than player one’s highest IR payoff). In this case, we show that a patient player one’s worst equilibrium payoff is equal to his minimax. Therefore, player one guarantees only his lowest payoff if he compromises by mimicking a commitment type with positive shortfall. Taken in conjunction with our reputation result, this implies that reputation building against an equally patient rival is an all-or-nothing phenomenon: player one guarantees either his best or only his worst repeated-game payoff.

As this discussion suggests, player one can successfully build a beneficial reputation by mimicking a commitment type, which is a finite automaton, if and only if the type has no

³The 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.

⁴We 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*.

⁵The Stackelberg payoff for player one is the highest payoff he can guarantee in the stage-game through public commitment to a stage-game action (a Stackelberg action). See [Mailath and Samuelson \(2006, page 465\)](#), for a formal definition.

⁶A finite automaton is an automaton with a finite number of states.

shortfall. For our reputation result, we ensure that a commitment type with no shortfall exists by assuming that the stage game has either LNCI or SCI. Now suppose that, in contrast to this assumption, the stage-game has neither LNCI nor SCI. In this case, we show that there is no finite automaton with no shortfall. Therefore, player one’s worst equilibrium payoff is equal to his minimax regardless of which finite automaton he mimics. In other words, it is not possible for player one to successfully build a beneficial reputation by mimicking a finite automaton unless the stage game has either LNCI or SCI.

Finally, we turn to identifying stage games for which there is a finite automaton with no shortfall. We show that there is a finite automaton with no shortfall if and only if there is a stage-game action (a strong Stackelberg action) whose commitment payoff is equal to player one’s highest IR payoff. Such games are a strict subset of the games with LNCI or SCI.⁷ If there is a strong Stackelberg action, then the finite automaton that plays this action in every period has no shortfall; moreover, player one can guarantee his highest IR payoff by mimicking this type. Hence, whenever player one can successfully build a reputation, he can do so by mimicking the least complex commitment type, i.e., a simple type that plays the strong Stackelberg action in each period. In other words, added complexity does not improve a patient player one’s worst equilibrium payoff as long as the complexity is still finite.

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 discuss this assumption in more detail in section 5.1.

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

⁷ If the stage game has LNCI and if there is no stage-game action whose commitment payoff is equal to player one’s highest IR payoff, then the commitment type with no shortfall is not a finite automaton. In such games we construct the required commitment type’s strategy using an infinite number of states.

the action the commitment type would have played; otherwise, his reputation level collapses to zero.⁸ If we relax the assumption that player one’s actions are perfectly monitored, then a 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.^{9,10}

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)).¹¹ We make three main contributions to this literature. First, we provide the first reputation result for games with LNCI.¹² Previous reputation results are for either stage games with SCI (Cripps et al. (2005)) or strictly-dominant-action stage games (Chan (2000)).¹³ 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

⁸We use these dynamics in proving both reputation results and our non reputation results.

⁹Fudenberg and Levine (1992)’s learning result (Theorem 4.1) does not help in our framework with equally patient agents.

¹⁰Note 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.

¹¹By *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).

¹²Atakan 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.

¹³For a precise definition of a strictly-dominant-action stage game, see Mailath and Samuelson (2006), Page 540.

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. However, 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 SCI and 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.¹⁴

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

¹⁴Also, 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.

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 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 $t \in \{0, 1, 2, \dots\}$. 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.¹⁵ 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.¹⁶

In the stage game, the payoff for any player i is given by the function $r_i : A_1 \times Y \rightarrow \mathbb{R}$ 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).

¹⁵ 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.

¹⁶ 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 5.2.

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 $h^t = (a_1^0, y^0, a_1^1, y^1, \dots, a_1^{t-1}, y^{t-1})$ for $t > 0$, and $h^0 = \emptyset$. The set of all public histories is $H = \bigcup_{t=0}^\infty H^t$. 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 $h_2^t = (a_1^0, a_2^0, y^0, \dots, a_1^{t-1}, a_2^{t-1}, y^{t-1})$, 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 $\sigma_i : H_i \rightarrow \Delta(A_i)$, 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 $\sigma_i : H_i \rightarrow \Delta(A_i)$ and Σ_i is the set of all behavior strategies for player i .¹⁷ 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, $h^T = (y^0, a_1^0, \dots, y^{T-1}, a_1^{T-1}) \in H(\sigma_1)$ if and only if $a_1^k \in \text{supp}(\sigma_1(h^k))$ 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 $\text{Pr}_{(\sigma_1, \sigma_2)}$.

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, $\mu : H \rightarrow \Delta(\Omega)$, 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 $o : \Theta \rightarrow \Delta(A_1)$ that assigns a (possibly mixed) stage-game action to each state, and a transition function $\tau : Y \times A_1 \times \Theta \rightarrow \Theta$ 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 $\text{Pr}_{(\omega, \sigma_2)}$ for any $\sigma_2 \in \Sigma_2$. A finite

¹⁷For player one, any behavior strategy is also a public behavior strategy because $H_1 = H$.

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 after history h^t . Consequently, $\Sigma_1 \setminus \{\omega\}$ is the set of commitment types other than ω , and $\Sigma_1 \setminus w(h^t)$ 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 $\sigma_2 : \Theta \rightarrow \Delta(A_2)$, 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 $h^{-t} = (a_1^t, y^t, a_1^{t+1}, y^{t+1}, \dots)$. 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:

$$\begin{aligned} U_1(\sigma, \delta | h^t) &= U_1(\sigma_1(N), \sigma_2, \delta | h^t), \\ U_2(\sigma, \delta, \mu | h^t) &= \sum_{\omega \in \Omega \setminus \{N\}} \mu(\omega | h^t) U_2(\omega, \sigma_2, \delta | h^t) + \mu(N | h^t) U_2(\sigma_1(N), \sigma_2, \delta | h^t), \end{aligned}$$

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 $U_1^{NE}(\mu) = \liminf_{\delta \rightarrow 1} U_1^{NE}(\delta, \mu)$. 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 .¹⁸*
- (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 $U_1^C(\sigma_1|h^t) = \liminf_{\delta \rightarrow 1} U_1^C(\sigma_1, \delta|h^t)$.¹⁹

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:

¹⁸If a perfect Bayesian equilibrium (PBE) is used as the equilibrium concept, then $(\sigma_1(N)|_{h^t}, \sigma_2|_{h^t})$ is a perfect Bayesian equilibrium of $\Gamma^\infty(\mu(h^t), \delta)$ for all $h^t \in H$, not just for those histories that have positive probability under the profile σ , as in this remark.

¹⁹ Although we define the commitment payoff using $\liminf_{\delta \rightarrow 1} U_1^C(\sigma_1, \delta|h^t)$, in the context of this paper the limit $\lim_{\delta \rightarrow 1} U_1^C(\sigma_1, \delta|h^t)$ exists.

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.

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).

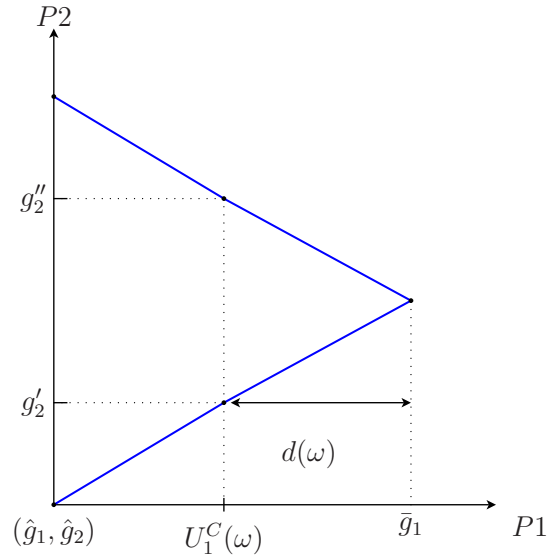


Figure 1: Shortfall of a strategy. Player two can receive any payoff between g_2' and g_2'' while player one receives $U_1^C(\omega)$.

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

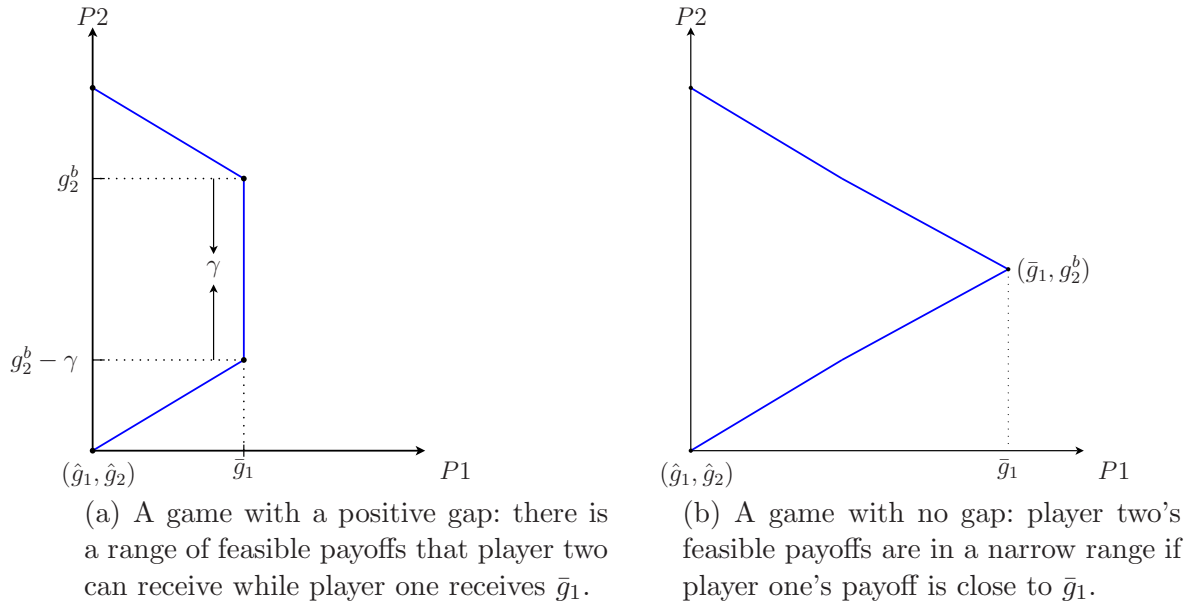


Figure 2: The gap of a game.

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 satisfy either *locally nonconflicting interests* (LNCI) or *strictly conflicting interests* (SCI), and we denote the set of such stage games by \mathcal{G} . The set of games with LNCI and the set of games with SCI are mutually exclusive. In a stage game with LNCI, in any payoff profile where player one receives his highest IR payoff, player two's payoff is strictly higher than her pure strategy minimax. In contrast, a stage game has SCI if any best reply to player one's Stackelberg action yields the highest IR payoff for player one and the minimax payoff for player two. See figure 3 for some prominent games with LNCI or SCI. Also, see figure 4 for a depiction of the set of feasible payoffs for a game with LNCI or SCI.²⁰ Although these two sets of games are quite different, both deliver the conditions we require for our reputation result. The formal definitions of LNCI and SCI are as follows:

²⁰Games with LNCI have a common-value component whereas games with SCI entail conflict. To see that games with LNCI have a common-value component, notice that in figure 4a the line segment which connects the point (\hat{g}_1, \hat{g}_2) with the point (\bar{g}_1, g_2) is strictly increasing and this line segment is on the boundary of the set of feasible and individually rational payoffs. In contrast, notice in figure 4b the line segment which connects the point (\bar{g}_1, \hat{g}_2) with the point (\hat{g}_1, \hat{g}_2) is strictly decreasing.

DEFINITION (**LNCI**) *Locally nonconflicting interests: for any $g \in G$, if $g_1 = \bar{g}_1$, then $g_2 > \hat{g}_2^p$.*

DEFINITION (**SCI**) *Strictly conflicting interests: there exists $a_1 \in A_1$ such that any best response to a_1 yields payoffs (\bar{g}_1, \hat{g}_2) .*

	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

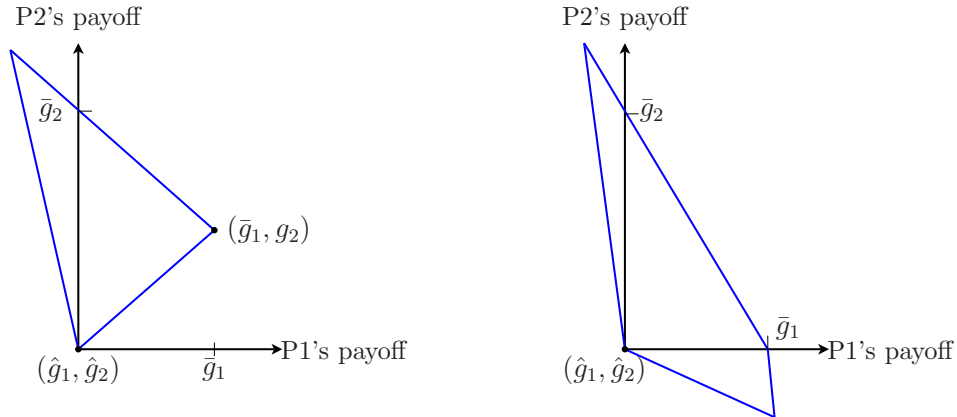
(d) The battle-of-the-sexes.

Figure 3: Stage games with LNCI (3a, 3b, and 3d) or SCI (3c).

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



(a) LNCI: the set F is bounded above and below by the lines that go through (\bar{g}_1, g_2) .

(b) SCI: the set F is bounded above by the downward sloping line that connects (\bar{g}_1, \hat{g}_2) to (\hat{g}_1, \bar{g}_2) .

Figure 4: Typical set of feasible payoffs for a game with LNCI (4a) or SCI (4b).

Other stage games that also feature prominently in our analysis are those where player one can guarantee his highest IR payoff by simply committing to a pure stage-game action (i.e., a pure strong Stackelberg action). Player one has a pure strong Stackelberg action in a stage game if there is a pure stage-game action a_1^s such that player one's public commitment to a_1^s guarantees him his highest IR payoff. The following is the definition of stage games which have pure strong Stackelberg actions:

DEFINITION (SA) *Pure 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 .*

Note that if player one has a pure strong Stackelberg action in Γ , then there is a pure strategy Nash equilibrium of Γ where player one plays the Stackelberg action a_1^s , player two best responds to a_1^s , and player one's payoff is equal to \bar{g}_1 . Also note that the set of stage games that satisfy SA is a strict subset of \mathcal{G} . The battle-of-the-sexes (figure 3d), the common-interest game (figure 3a), and the chain-store game (figure 3c) are all SA games whereas the principal-agent game (figure 3b) has LNCI but is not an SA game. 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, in the principal-agent 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 .

The games that satisfy SA are prominent in our analysis when all of player one's commitment types are finite automata. This is because if the stage game satisfies SA, 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 satisfies SA 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 satisfies SA. The following lemma, which is proved in appendix A, shows that the converse is also true: if the stage game does not satisfy SA, then a pure-strategy finite automaton with no shortfall does not exist. Nevertheless, in Theorem 1 and in section 3.3 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 satisfies SA.*

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 3.3 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 satisfy LNCI or SCI. 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.4.

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.3, 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 (8) in the appendix, which satisfies $\lim_{x \rightarrow 0} f(\omega, x) = \infty$.

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. In section 4, we further explore the cases where the shortfall for the only available commitment type is positive and the stage game Γ has a positive gap. In particular, we show that there are PBE where player one's payoff is near his minimax if either the shortfall of type ω^* or the gap of Γ is positive.

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 satisfies SA, 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 satisfy SA: 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 and SA, 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: The shortfall $d(\omega^*)$ is equal to zero by assumption. Consequently, Lemma 2 implies that if $\mu(\omega^*) > 0$, then $U_1^{NE}(\mu) \geq U_1^C(\omega^*)$. However, $U_1^C(\omega^*)$ is equal to \bar{g}_1 because $d(\omega^*)$ is equal to zero. \square

For stage games that do not satisfy SA, 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 satisfy SA 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 3.3 we discuss how we use Lemma 2 as an intermediate step to prove a reputation result for stage games that do not satisfy SA.

3.1. The reputation result for the repeated common-interest game. In this section, we provide the intuition in words and present a sketch for the proof of the reputation result in the particular case of the repeated common-interest game (figure 5). We assume that $\Omega = \{\omega^*, N\}$, where ω^* is the type which plays U in every period. When applied to this example, Corollary 1 implies that if $\mu(\omega^*) > 0$, then $U_1^{NE}(\mu) = 1$. Corollary 1 applies because the game satisfies FS and SA, and has no gap.

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

Figure 5: A common-interest game. The set of public signals is $Y = \{l, r\}$ and $\pi_r(L) = \pi_l(R) := \underline{\pi} \in (0, 1/2)$. This game satisfies SA because the action profile (U, L) gives player one his highest payoff and because player two’s unique best response to U is L . This game satisfies FS because $\underline{\pi} > 0$ and $1 - \underline{\pi} > 0$.

REMARK 2 *In contrast to what we assume, if $\underline{\pi} = 0$ in the game in figure 5, then player two’s actions are perfectly monitored and the game does not satisfy FS. In this case, our reputation result does not apply. In fact, Cripps and Thomas (1997) prove a folk theorem under perfect monitoring for this game where they show that the worst perfect Bayesian equilibrium payoff for a patient player one is equal to his minimax. For more on the use of the FS assumption, see Remark 4 below and section 5.1.*

3.1.1. *Verbal intuition.* In order to show that a sufficiently patient player one can guarantee a payoff equal to one in any NE, we will argue that player two plays R in only a payoff-insignificant number of periods against an opponent who repeatedly plays U .

If player one has played U in all previous periods, then player two assigns positive probability to player one being type ω^* and therefore attaches positive probability to player one playing U in the current period. The stage-game action L is a strict best response to U and also weakly dominates action R . Consequently, for player two to play R after observing

player one play U in all the previous periods, she must expect punishment in the continuation game if she plays L . Player one does not observe player two's action directly, but observes either signal l or signal r . Hence, if player two expects punishment for playing L , then her punishment must occur after the signal l is observed. We argue that the scope for punishment in the continuation game is limited.

The commitment type ω^* cannot punish player two because this type always plays U . Therefore, any punishment for player two must occur after player one reveals rationality. However, player one can mimic type ω^* instead of revealing rationality; hence, his continuation payoff after revealing rationality cannot differ significantly from his payoff from mimicking type ω^* . Also, because of the **FS** assumption, signal l is observed with positive probability regardless of player two's action, i.e., signal l occurs even if player two plays R . Hence, player one's payoff after signal l cannot differ significantly from his payoff from mimicking type ω^* either. Therefore, player one's payoff while punishing player two cannot differ significantly from his payoff from mimicking type ω^* . In other words, punishing player two cannot be costly for player one. Because the game satisfies **SA**, the commitment payoff of type ω^* is equal to the highest payoff of player one. Moreover, because the game has no gap, if player one's payoff is close to his highest payoff, then player two's payoffs are in a narrow range (see figure 6). Therefore, if punishments are not costly for player one, then player two's feasible continuation payoffs lie in a narrow range. Thus, the scope for punishment is limited.

Player two cannot expect significant punishment for playing L because, as we argued above, the scope for such punishment in the continuation game is limited. Also, choosing not to best respond to ω^* by playing R is costly for player two. Hence, player two cannot be incentivized to play R for a payoff-significant number of periods in any NE of the repeated game.

3.1.2. *Sketch of the proof.* First, some preliminaries: Given that there is only one commitment type in the example under consideration, we take the measure $\mu \in [0, 1]$ to denote the probability that player 1 is type ω^* . Likewise, given any public history $h^t \in H$, the posterior $\mu(h^t) \in [0, 1]$ denotes the probability that player 1 is type ω^* given public history h^t .

Let $U_{\omega^*}(\delta, \mu) = \inf\{U_1(\omega^*, \sigma_2, \delta) : \sigma_2 \text{ is part of a NE strategy profile of } \Gamma^\infty(\delta, \mu)\}$, and let $U_{\omega^*}(\mu) = \lim_{\delta \rightarrow 1} U_{\omega^*}(\delta, \mu)$. In words, $U_{\omega^*}(\mu)$ is the worst payoff that a patient player one can get if he uses strategy ω^* . Because ω^* is always a feasible strategy for player one, we have $U_1^{NE}(\delta, \mu) \geq U_{\omega^*}(\delta, \mu)$. We will show that $U_{\omega^*}(\mu) \geq 1$.

For this sketch, we also work under the hypothesis that $U_{\omega^*}(\delta, \mu)$ is a nondecreasing function of μ for each $\delta \in [0, 1)$. We do this for expositional convenience as it allows us to convey the main argument without the more technical details.²¹

²¹ In the main argument given in appendix B, we instead work with $\underline{U}(\mu, \delta) = \inf\{U_{\omega^*}(\mu', \delta) : \mu' \geq \mu\}$,

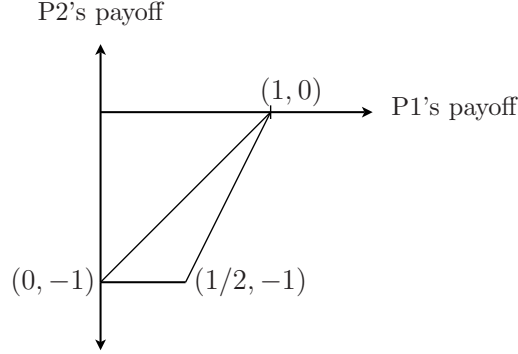


Figure 6: The set of feasible payoffs for the common-interest game. The set of feasible payoffs is given by the intersection of the half-spaces $g_1 \geq 0$, $g_2 \geq -1$, $g_2 \leq -(1 - g_1)$, and $g_2 \geq -2(1 - g_1)$.

Observe that if (g_1, g_2) is a vector of feasible payoffs for the common-interest game, then $g_2 \leq -(1 - g_1)$ and $g_2 \geq -2(1 - g_1)$ (see figure 6). The set of feasible stage-game payoffs and the set of feasible repeated-game payoffs coincide because the two players share a common discount factor. For any pair $(\sigma_1, \sigma_2) \in \Sigma_1 \times \Sigma_2$, the vector $(U_1(\sigma_1, \sigma_2, \delta), U_2(\sigma_1, \sigma_2, \delta))$ belongs to the set of feasible payoffs; it follows that

$$(1) \quad -2(1 - U_1(\sigma_1, \sigma_2, \delta)) \leq U_2(\sigma_1, \sigma_2, \delta) \leq -(1 - U_1(\sigma_1, \sigma_2, \delta)).$$

REMARK 3 *In this repeated game, the commitment payoff of type ω^* is equal to one. Therefore, if player one's NE payoff is close to the commitment payoff of type ω^* , i.e., if it is close to one, then player two's payoff conditional on facing player one's normal type, $U_2(\sigma_1(N), \sigma_2, \delta)$, is in a narrow range determined by inequality (1). For example, if the difference between player one's NE payoff and the commitment payoff of type ω^* is equal to ϵ , then player two's payoff conditional on facing player one's normal type is between -2ϵ and $-\epsilon$.*

We now use inequality (1) to establish an upper bound on player two's NE payoff in the following claim:

CLAIM 1 *For any $\mu \in [0, 1]$ and any NE strategy profile σ of the repeated common-interest game $\Gamma^\infty(\delta, \mu)$, player two's NE equilibrium payoff satisfies the following inequality:*

$$U_2(\sigma, \delta, \mu) \leq -\mu(1 - U_1(\omega^*, \sigma_2, \delta)).$$

which is nondecreasing by definition.

PROOF: Let $\sigma = (\sigma_1(N), \omega^*, \sigma_2)$ be a NE strategy profile and recall that $U_1(\sigma, \delta) = U_1(\sigma_1(N), \sigma_2, \delta)$. Inequality 1 implies that $U_2(\sigma_1(N), \sigma_2, \delta) \leq -(1 - U_1(\sigma, \delta))$. Moreover, $U_2(\sigma_1(N), \sigma_2, \delta) \leq -(1 - U_1(\sigma, \delta)) \leq 0$ because player one's highest feasible payoff is one. Inequality (1) also implies that

$$(2) \quad U_2(\omega^*, \sigma_2, \delta) \leq -(1 - U_1(\omega^*, \sigma_2, \delta)).$$

Player two's NE payoff $U_2(\sigma, \delta, \mu)$ is equal to $\mu U_2(\omega^*, \sigma_2, \delta) + (1 - \mu)U_2(\sigma_1(N), \sigma_2, \delta)$. Therefore, (2) and $U_2(\sigma_1(N), \sigma_2, \delta) \leq 0$ together prove the result. \square

For the next claim, consider a NE σ of the repeated game $\Gamma^\infty(\mu, \delta)$. Suppose that player one has played U in each period before t . Further suppose that player one plays D with positive probability in period t . Also, let player one's reputation level be $\mu' \geq \mu$ at the start of period $t + 1$, if he plays U instead of D . In the next claim we show that player two's continuation payoff, after any realization of the period t public signal, is bounded from below by a linear function of $1 - U_{\omega^*}(\mu', \delta)$. The argument for the claim is as follows: If player one is playing D with positive probability, then his payoff from playing D must be at least as large as his payoff from playing U . However, if player one plays U , then he ensures that his reputation is μ' at the start of the subsequent period, and thus guarantees a continuation payoff of $U_{\omega^*}(\mu', \delta)$ in period $t + 1$. Given this lower bound on player one's continuation payoff, the lower bound on player two's continuation payoff follows from inequality (1).

CLAIM 2 *Let σ be any NE strategy profile of the repeated common-interest game $\Gamma^\infty(\delta, \mu)$. Suppose that public history h^t has positive probability under the NE strategy profile σ and that $\mu(h^t) > 0$. Suppose also that player one plays D with positive probability in period t . Then $U_2(\sigma_1(N), \sigma_2, \delta|h^t, D, l) \geq -2(1 - U_{\omega^*}(\delta, \mu'))/\underline{\pi}$, and $U_2(\sigma_1(N), \sigma_2, \delta|h^t, D, r) \geq -2(1 - U_{\omega^*}(\delta, \mu'))/\underline{\pi}$, for any $\mu' \leq \mu(\omega^*|h^t, U, l) = \mu(\omega^*|h^t, U, r)$.*

PROOF: Let π_y^t denote the probability that signal $y \in \{l, r\}$ is observed after period t conditional on history h^t .²² If player one plays D in period t , then his payoff for period t is at most $1/2$ and his continuation payoff is equal to $\pi_l^t U_1(\sigma, \delta|h^t, D, l) + \pi_r^t U_1(\sigma, \delta|h^t, D, r)$. Alternatively, player one can play U forever, that is, he can use strategy ω^* . If player one plays strategy ω^* from period t onwards, then his payoff in period t is at least $1/2$, and his continuation payoff is at least $\pi_l^t U_{\omega^*}(\delta, \mu(\omega^*|h^t, U, l)) + \pi_r^t U_{\omega^*}(\delta, \mu(\omega^*|h^t, U, r))$. By Remark 1, the strategy profile $(\sigma_1|_{h^t}, \sigma_2|_{h^t})$ is a NE of the continuation game. Because $(\sigma_1|_{h^t}, \sigma_2|_{h^t})$ is a NE of the continuation game and because player one plays D with positive probability

²²The probability π_y^t is calculated using player two's NE strategy σ_2 , conditional on history h^t .

in period t , his repeated game payoff from playing D in period t must be at least as large as his repeated game payoff from using strategy ω^* from period t onwards. Consequently,

$$\pi_l^t U_1(\sigma, \delta|h^t, D, l) + \pi_r^t U_1(\sigma, \delta|h^t, D, r) \geq \pi_l^t U_{\omega^*}(\delta, \mu(\omega^*|h^t, U, l)) + \pi_r^t U_{\omega^*}(\delta, \mu(\omega^*|h^t, U, r)).$$

Also, $U_{\omega^*}(\delta, \mu(\omega^*|h^t, U, l)) = U_{\omega^*}(\delta, \mu(\omega^*|h^t, U, r)) \geq U_1^{NE}(\delta, \mu')$ because $\mu' \geq \mu(\omega^*|h^t, U, l) = \mu(\omega^*|h^t, U, r)$ and because U_1^{NE} is nondecreasing. Thus,

$$\pi_l^t U_1(\sigma, \delta|h^t, D, l) + \pi_r^t U_1(\sigma, \delta|h^t, D, r) \geq U_{\omega^*}(\delta, \mu').$$

By adding one to both sides of the previous inequality and rearranging, we obtain:

$$\pi_l^t(1 - U_1(\sigma, \delta|h^t, D, l)) + \pi_r^t(1 - U_1(\sigma, \delta|h^t, D, r)) \leq 1 - U_{\omega^*}(\delta, \mu').$$

Because player one's highest payoff in the game is equal to one, $1 - U_1(\sigma, \delta|h^t, D, y)$ is nonnegative for any $y \in \{r, l\}$. Therefore,

$$\pi_y^t(1 - U_1(\sigma, \delta|h^t, D, y)) \leq 1 - U_{\omega^*}(\delta, \mu'),$$

for any $y \in \{r, l\}$. Also, our assumption that $\pi_y^t \geq \underline{\pi} > 0$ for any $y \in \{r, l\}$, i.e., the *FS* assumption, implies that

$$(3) \quad 1 - U_1(\sigma, \delta|h^t, D, y) \leq (1 - U_{\omega^*}(\delta, \mu'))/\underline{\pi},$$

for any $y \in \{r, l\}$. However, because $(U_1(\sigma, \delta|h^t, D, y), U_2(\sigma_1(N), \sigma_2, \delta|h^t, D, y))$ is a feasible payoff vector for any $y \in \{r, l\}$, we can derive the following inequality from inequalities (1) and (3):

$$U_2(\sigma_1(N), \sigma_2, \delta|h^t, D, y) \geq -2(1 - U_{\omega^*}(\delta, \mu'))/\underline{\pi},$$

for any $y \in \{r, l\}$. □

REMARK 4 Notice that if the *FS* assumption is not satisfied, i.e., if $\underline{\pi} = 0$, then the lower bound given by Claim 2 is not well defined. For example, suppose that *FS* is not satisfied by the stage game. Also, suppose that $\pi_l^t = 0$, i.e., signal l occurs with zero probability after history h^t . In this case, following the reasoning in Claim 2 does not deliver a lower bound on player two's continuation payoff after public history (h^t, D, l) because this history occurs with probability zero on the equilibrium path. See section 5.1 for more on the use of the *FS*

assumption.

Fix any arbitrary reputation level $\underline{\mu} > 0$. Starting in the next paragraph below, we use Claims 1 and 2 to show that

$$(4) \quad 1 - U_{\omega^*}(\mu) \leq \frac{4\underline{\mu}}{\pi}(1 - U_{\omega^*}(\mu'))$$

for any two reputation levels $\mu' \geq \mu \geq \underline{\mu}$ such that $\mu/\mu' \geq 1 - \underline{q}$, where $\underline{q} := \pi\underline{\mu}/4$. Once we establish inequality (4), a reputation result for the common-interest game follows. This is because inequality (4) and $U_{\omega^*}(1) = 1$ together imply that $U_{\omega^*}(\mu') \geq 1$ for all $\mu' \geq \underline{\mu} > 0$. To see this, pick a nonincreasing sequence of numbers $\{\mu_n\}_{n=0}^N$ such that $\mu_0 = 1$, $\mu_N = \underline{\mu}$, and $\mu_{n+1}/\mu_n \geq 1 - \underline{q}$ for each n . We will run induction on n to show that $U_{\omega^*}(\mu_N) = U_{\omega^*}(\mu) = 1$. Notice that $\mu_0 = 1$ and $U_{\omega^*}(\mu_0) \geq 1$. Now make the induction hypothesis that $U_{\omega^*}(\mu') \geq 1$ for all $\mu' \geq \mu_n \geq \underline{\mu}$. For $\mu' = \mu_n$ and $\mu \in [\mu_{n+1}, \mu_n)$, inequality (4) and the induction hypothesis together imply that $U_{\omega^*}(\mu) \geq 1$, the induction is thus completed. Therefore, for any $\mu \geq \underline{\mu}$, we have $U_{\omega^*}(\mu) \geq 1$. Since our initial choice of $\underline{\mu}$ is arbitrary, we find that $U_{\omega^*}(\mu) \geq 1$ for any $\mu > 0$.

Now we turn to establishing inequality (4). Suppose that $\mu \geq \underline{\mu}$. Consider a NE σ of the repeated common-interest game $\Gamma^\infty(\delta, \mu)$ where player one's payoff from using strategy ω^* , $U_1(\omega^*, \sigma_2, \delta)$, is equal to $U_{\omega^*}(\delta, \mu) + \epsilon$.²³ By Claim 1, player two's NE payoff $U_2(\sigma, \delta, \mu)$ is less than or equal to $-\mu(1 - U_{\omega^*}(\delta, \mu) - \epsilon)$. We will establish inequality (4) by using the fact that the NE strategy σ_2 must give player two a payoff that is at least as great as the payoff from using an *alternative strategy* that is available to her. We consider the alternative strategy for player two that plays L until player one plays D for the first time and then reverts back to the equilibrium strategy σ_2 .

Lower bound on player two's payoff from using the alternative strategy. Suppose that player one's initial reputation level is equal to $\mu > \underline{\mu}$. Choose any reputation level μ' that is greater than μ and that satisfies $\mu/\mu' \geq 1 - \underline{q}$. Let τ denote the first period in which player one's reputation level exceeds μ' if player one plays according to strategy ω^* and player two uses the alternative strategy. We will provide a bound on player two's payoff conditional on one of the following three mutually exclusive and exhaustive events occurring: First, the event that player one plays U in every period of the repeated game. Second, the event that player one plays D for the first time in some period $k < \tau - 1$; the probability of this event is at most \underline{q} , by the definition of τ , Bayes' rule, and the choice of μ' . Third, the event that player one plays D for the first time in some period $k \geq \tau - 1$; the probability of this event

²³For each $\epsilon > 0$, such a NE exists because $U_{\omega^*}(\delta, \mu)$ is defined as the infimum over the set of NE profiles.

is at most $1 - \underline{q}$.

If player one plays U in every period, i.e., if the first event occurs, then player two's payoff from using the alternative strategy is equal to zero.

Alternatively, suppose that player one plays D for the first time in some period k , i.e., either the second or the third event occurs. Player two receives zero in each period up to period k , because she plays L and player one plays U in each such period. In period k , she receives $-(1 - \delta)$, because she plays L and player one plays D . If $k < \tau - 1$, i.e., if the second event occurs, then player two receives at least $-2(1 - U_{\omega^*}(\delta, \mu))/\underline{\pi}$ as a continuation payoff after period k . This follows from Claim 2 because player one's reputation is at least μ if he plays U in period k . If $k \geq \tau - 1$, i.e., if the third event occurs, then player two receives at least $-2(1 - U_{\omega^*}(\delta, \mu'))/\underline{\pi}$ as a continuation payoff after period k . This follows from Claim 2 because player one's reputation is at least μ' if he plays U in period k . Consequently, if player two uses the alternative strategy, then her payoff is at least

$$-\delta^k(1 - \delta) - \frac{2\delta^{k+1}}{\underline{\pi}}(\underline{q}(1 - U_{\omega^*}(\delta, \mu)) + (1 - \underline{q})(1 - U_{\omega^*}(\delta, \mu'))).$$

Simplifying this expression by using the fact that $\delta < 1$ and $1 - \underline{q} < 1$ shows that player two's payoff from using the alternative strategy is at least

$$-(1 - \delta) - \frac{2}{\underline{\pi}}(\underline{q}(1 - U_{\omega^*}(\delta, \mu)) + (1 - U_{\omega^*}(\delta, \mu'))).$$

Establishing inequality (4). Player two's payoff from her NE strategy σ_2 must be at least as great as her payoff from the alternative strategy. Therefore,

$$-(1 - \delta) - \frac{2}{\underline{\pi}}(\underline{q}(1 - U_{\omega^*}(\delta, \mu)) + (1 - U_{\omega^*}(\delta, \mu'))) \leq -\mu(1 - U_{\omega^*}(\delta, \mu)) + \epsilon.$$

By rearranging, we obtain:

$$(\mu - 2\underline{q}/\underline{\pi})(1 - U_{\omega^*}(\delta, \mu)) \leq (1 - \delta) + \mu\epsilon + \frac{2}{\underline{\pi}}(1 - U_{\omega^*}(\delta, \mu')).$$

As this inequality holds for any $\epsilon > 0$, we have

$$(\mu - 2\underline{q}/\underline{\pi})(1 - U_{\omega^*}(\delta, \mu)) \leq (1 - \delta) + \frac{2}{\underline{\pi}}(1 - U_{\omega^*}(\delta, \mu')).$$

Because the above inequality holds for every $\delta \in [0, 1)$ and because $\underline{\mu} \leq \mu$, we find that

$$(\underline{\mu} - 2\underline{q}/\underline{\pi})(1 - U_{\omega^*}(\underline{\mu})) \leq (2/\underline{\pi})(1 - U_{\omega^*}(\mu')).$$

Substituting $\underline{\pi}\mu/4$ for \underline{q} , we obtain the following inequality:

$$1 - U_{\omega^*}(\mu) \leq (4\underline{\mu}/\underline{\pi})(1 - U_{\omega^*}(\mu')).$$

3.2. Other commitment types. So far our discussion of Theorem 1, Lemma 2, and Corollary 1 has focused on the case where $\Omega = \{N, \omega^*\}$ for simplicity. However, as is evident in the statement of these results, we do not require that $\Omega = \{N, \omega^*\}$. More specifically, we place no restriction on the other commitment types beyond requiring that they be finite automata. In the proof of Lemma 2, we achieve this in the following way: First, we show that if player two assigns a small probability to the event that player one's type is in the set $\Omega_{-\omega^*}$, then the effect of the types in $\Omega_{-\omega^*}$ on player one's NE payoff is also small. Second, we show that if all the other commitment types are finite automata, and if player one mimics type ω^* for a sufficiently long number of periods t , then with large probability, player two's posterior belief that player one's type is in $\Omega_{-\omega^*}$, i.e., $\mu(\Omega_{-\omega^*} | h^t)$, becomes arbitrarily small. More importantly, the rate at which such *learning* occurs is independent of player two's strategy. Thus, by combining these two findings, we conclude that the presence of other commitment types does not disrupt our reputation result.

Our second finding (i.e., that if player one mimics type ω^* , then player two learns that player one is not playing a strategy in $\Omega_{-\omega^*}$ at a uniform rate) plays a crucial role in our reputation result (see Lemma A.2 in the appendix). We now give the intuition for this second result. The following statements are true, regardless of which strategy player two uses, because of the FS assumption: Each recurrent state of any finite automata will be visited infinitely often. Thus, player two will observe player one's action at each state infinitely often. Therefore, if player one plays according to ω^* , then player two can reject the hypothesis that the observed sequence of play is generated by any finite automata other than ω^* with arbitrarily high probability. Moreover, learning occurs at a uniform rate because all the commitment types have finitely many states.

3.3. Reputation effects in stage games that do not satisfy SA. The proof of Theorem 1 shows that, for any stage-game in \mathcal{G} with no gap which does not satisfy SA, again 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 satisfy SA (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 $3 - \epsilon$.²⁴ 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 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.

4. NONREPUTATION RESULTS

Now we turn our focus to the following two questions: Which types, if available, facilitate reputation building for player one? In which strategic situations (i.e., for which class of stage games) can player one successfully build a reputation? In addressing these questions, we restrict attention to pure-strategy finite automata.²⁵ Under this restriction, we show that (i)

²⁴See Celentani et al. (1996) which shows that $J(\epsilon)$ and ξ can indeed be chosen in this way.

²⁵We expand on the limitations of this restriction at the end of this section.

only types with no shortfall facilitate successful reputation building (Theorem 2), and (ii) player one can build a reputation only in SA games (Theorem 3).

We say that player one cannot successfully build a reputation (i.e., that there are no reputation effects) in the repeated game Γ^∞ given a set of commitment types Ω if we can find a prior with full support μ over Ω such that a patient player one's worst equilibrium payoff is arbitrarily close to his worst equilibrium in the repeated game of complete information. Or conversely, there are reputation effects if a patient player one's worst NE payoff under incomplete information is uniformly greater than his worst NE payoff under complete information. More intuitively, there are reputation effects if adding even a small amount of incomplete information improves a patient player one's worst NE payoff significantly. We continue our analysis by considering repeated games in which player one's worst NE payoff is his minimax in the benchmark model with complete information. Thus, we make the following assumption:

ASSUMPTION (NI) *Non-empty interior: The set F has a non-empty interior and there exists a payoff g_2 such that $(\hat{g}_1, g_2) \in G$.*²⁶

Therefore, the formal definition of no reputation effects is as follows:

DEFINITION (No reputation effects) *We say that there are no reputation effects in Γ^∞ , given a set of commitment types Ω , if for each $\epsilon > 0$, there exists $\mu \in \Delta^\circ(\Omega)$ and $\delta^* \in [0, 1)$ such that for any $\delta \geq \delta^*$, there is a perfect Bayesian equilibrium (PBE) σ of $\Gamma^\infty(\delta, \mu)$ where $U_1(\sigma, \delta) < \hat{g}_1 + \epsilon$.*

The following theorem, which is proved in appendix C, shows that commitment types with a shortfall cannot facilitate successful reputation building.

THEOREM 2 (Reputation effects are all-or-nothing) *Suppose Γ satisfies NI. Suppose also that $\Omega = \{N, \omega\}$ where ω is a pure-strategy finite automaton. If the shortfall of ω is positive, then there are no reputation effects in Γ^∞ .*

Suppose that the stage game has no gap. In this case Lemma 2 and Theorem 2 together fully characterize the set of pure-strategy finite automata that facilitate successful reputation

²⁶The first part of the assumption, namely that the set of feasible payoffs has a non-empty interior (i.e., the full dimensionality condition) is standard in folk theorems (see, for instance, Fudenberg et al. (1994)). This first part is the core of the assumption. The second part of NI requires that there exists a payoff g_2 such that $(\hat{g}_1, g_2) \in G$. This part is for expositional purposes. If F has a non-empty interior but there exists no g_2 such that $(\hat{g}_1, g_2) \in G$, then the goal of Theorem 2 would be to establish that a patient player one's worst NE payoff is equal to his lowest payoff compatible with both players' individual rationality. This payoff can, in general, exceed his minimax.

building for player one. In particular, suppose that $\Omega = \{N, \omega\}$ where ω is a pure-strategy finite automaton. Then Lemma 2 and Theorem 2 together imply that there are reputation effects in Γ^∞ *if and only if* ω has no shortfall.

It is also instructive to contrast Theorem 2 with reputation results obtained in repeated games with asymmetric discounting. If player two is less patient than player one, then player one's worst NE payoff is equal to the commitment payoff of type ω . Therefore, there are reputation effects as long as ω 's commitment payoff strictly exceeds player one's minimax (see Celentani et al. (1996), or Fudenberg and Levine (1989, 1992)). In contrast, Theorem 2 shows that a pure-strategy finite-automaton commitment type facilitates reputation building only if the type has no shortfall. Moreover, the lower bound attained with such a type is player one's highest IR payoff. Hence, player one guarantees either his highest IR payoff or his lowest payoff.

The following theorem shows that a pure-strategy finite automaton can facilitate reputation building only in repeated SA games.

THEOREM 3 *Suppose Γ satisfies FS and NI. Suppose also that $\Omega = \{N, \omega\}$, where ω is any pure-strategy finite automaton. If Γ does not satisfy SA, then there are no reputation effects in Γ^∞ .*

The proof of Theorem 3 follows from Lemma 1 and Theorem 2, taken together with the following lemma, which is proved in appendix C.

LEMMA 3 *Suppose Γ satisfies NI but does not satisfy SA. Suppose also that $\Omega = \{N, \omega\}$ where ω is a pure-strategy finite automaton. If the gap of Γ is positive, then there are no reputation effects in Γ^∞ .²⁷*

PROOF OF THEOREM 3: If Γ does not satisfy SA, then either Γ has a positive gap, or ω has a positive shortfall. This follows from Lemma 1. However, if ω has a positive shortfall, then Theorem 2 implies that there are no reputation effects. Alternatively, if Γ has a positive gap, then Lemma 3 implies that there are no reputation effects. \square

For stage games with no gap, Theorem 1 and Theorem 3 together fully characterize the games in which player one can successfully build a reputation by mimicking a pure-strategy finite automaton. To see this, suppose that Γ has no gap. Suppose also that $\Omega = \{N, \omega\}$, where ω is an arbitrary pure-strategy finite automaton. Notice that if the stage game satisfies SA, then there is a finite automaton ω^* with no shortfall and consequently there are

²⁷In fact, a stronger result holds, as follows: Suppose Γ satisfies NI but is not a strictly dominant action game. Suppose also that $\Omega = \{N, \omega\}$ where ω is a pure-strategy finite automaton. If the gap of Γ is positive, then there are no reputation effects in Γ^∞

reputation effects if $\omega = \omega^*$. In contrast, if the stage game does not satisfy SA, then there does not exist a finite automaton with no shortfall. Therefore, there are reputation effects for some pure strategy finite automaton *if and only if* the game satisfies SA.

Notice also that if the stage game satisfies SA, then the finite automaton that plays the strong Stackelberg action in every period has no shortfall. Moreover, player one can guarantee his highest IR payoff by simply mimicking this type. Hence, whenever player one can successfully build a reputation, he can do so by mimicking the least complex commitment type that plays the strong Stackelberg action in each period; in other words, added complexity does not improve a patient player one's worst payoff as long as the complexity is still finite.

In proving Theorem 2 and 3 we restricted attention to pure-strategy finite automata. The reasons we restrict attention to pure strategies were discussed in the introduction; our restriction to finite automata in Theorem 2 is mainly for technical convenience, and the finiteness of the automaton does not play a crucial role in the argument. We anticipate that an argument along similar lines would establish a nonreputation result for more complicated types that have a shortfall which is uniformly bounded away from zero after any history.²⁸ In contrast, in Theorem 3 the restriction to finite automata is a substantive assumption. As we showed previously, there are reputation effects in stage games in \mathcal{G} that are not SA games if we allow player one to mimic an automaton with an infinite number of states (see section 3.3). In section 5.3, we further discuss reputation effects outside the class \mathcal{G} .

4.1. The intuition for Theorem 2 and Lemma 3. The intuition for Theorem 2 and Lemma 3 is identical and is as follows: If the stage game has a positive gap or the commitment strategy has a positive shortfall, then NI and the folk theorem in Fudenberg et al. (1994) imply that in the complete information repeated game there is a wide range of PBE payoffs for player two that are all consistent with player one receiving his commitment payoff $U_1^C(\omega)$. Player one can use this range to credibly reveal rationality, still receive a payoff equal to his commitment payoff, and credibly punish or reward player two in the continuation game of complete information.

We construct an equilibrium in which player one mimics the commitment type for a long number of periods that are payoff-relevant and player two plays an action that gives player one a low payoff during these periods (*the review phase*). The action that player two plays in each period of the review phase is potentially a myopic non best response. We show that both players are incentivized to adhere to this equilibrium, given their equilibrium continuation payoffs. At the end of the review phase, player one reveals his rationality, and receives his commitment payoff in the continuation game. Hence, player one can do at least as well by

²⁸However, the argument is likely to be more cumbersome since the stationarity that the finite automata provides is lost with a more complicated commitment type.

revealing rationality instead of continuing to mimic the commitment type. After player one reveals rationality, player two receives a high continuation payoff if player one's realized payoff during the review phase was low, but receives a low continuation payoff otherwise. Because the review phase is sufficiently long to be payoff relevant, player one's payoff is strictly lower than his commitment payoff.

In the remaining steps of the construction, we ensure that player one's equilibrium payoff is close to his minimax payoff if the commitment type is sufficiently unlikely. Moreover, we apply an appropriate law of large numbers to overcome the technical difficulties introduced by imperfect monitoring. We sketch our construction for a particular example in the next subsection.

Notice that our construction relies heavily on the fact that there is a range of payoffs for player two that is consistent with player one receiving $U_1^C(\omega)$ in the periods where player one reveals his rationality. This property is a consequence of either a positive shortfall or a positive gap. If the strategy had no shortfall and the game had no gap, then there is a unique payoff for player two which is consistent with player one receiving $U_1^C(\omega)$. We now describe why our construction would not work in this case. First consider the scenario where player one's continuation payoff is equal to $U_1^C(\omega)$. In this scenario, player two receives her unique payoff consistent with $U_1^C(\omega)$ in the continuation game regardless of what she plays in the initial periods. But then player two would prefer to play a best response to ω during the initial periods in which player one mimics ω . Second, consider the scenario where player one's continuation payoff is strictly lower than $U_1^C(\omega)$. In this scenario, there is a range of continuation payoffs for player two that are feasible. Moreover, this range can be utilized to provide incentives for player two to play a non best response to ω during the initial periods in which player one mimics ω . However, in this scenario player one would never reveal rationality as required by our construction. This is because he can always guarantee $U_1^C(\omega)$ in the continuation game by mimicking ω .

4.2. Sketch of the construction. We use the game in figure 7 to illustrate the construction of an equilibrium in which player one's payoff is close to his minimax, which is zero in this game.²⁹ This game has a positive gap. Consider the repeated game with types $\Omega = \{N, \omega\}$ where ω plays U at every period. The commitment payoff of ω is 1.

For an arbitrarily chosen constant $\varepsilon > 0$, we will choose a cutoff reputation level z^* and a cutoff discount factor δ^* , and will construct a public PBE where player one's payoff is at most ε for any repeated game $\Gamma(\mu, \delta)$ where $\delta > \delta^*$ and $\mu(\omega) < z^*$.

Define $v > 0$ such that for every $g_1^* \in (\varepsilon, 1]$, there are two payoff profiles $(g_1^*, g_2^*) \in G$ and

²⁹This construction is similar to the construction presented in [Cripps and Thomas \(2003\)](#).

	L	R	M
U	1, 1	0, 0	0, 0
D	0, 0	1, 1/2	0, 0
M	0, 0	0, 0	0, 0

Figure 7: A game with a positive gap.

$(g_1^*, g_2'') \in G$ such that $v \leq g_2'' - g_2'$. Hence, the range of player two's payoffs is at least v if player one's payoff is greater than ε . In the game under consideration, v can be taken to equal $\varepsilon/2$. More generally, if the stage game has a positive gap and satisfies NI, then such a $v > 0$ exists.

We describe the equilibrium behavior of player one explicitly at histories where he has not yet revealed his rationality. At all other histories, we will not be explicit about his choice of action, but will instead pick the continuation equilibrium payoffs from the complete information repeated game's set of equilibrium payoffs. The equilibrium path of play where player one has not yet revealed his rationality consists of blocks, each of length T^* . We numerate the blocks so that block one is the last block, block two is the second-to-last block, and similarly for the rest of the blocks. Observable deviations from the prescribed behavior along the blocks will oblige player one to reveal his rationality and will trigger an indefinite repetition of (M, M) . This ensures that observable deviations are not profitable as long as player one's equilibrium payoff along the blocks is always bounded away from zero and the players are sufficiently patient, as will be the case.

Block Structure: Each block consists of two phases: the first $T^* - 1$ periods of the block constitute the review phase, and period T^* within the block is the revelation period. Player one plays U during the review phase. In a revelation period, player one plays, with equal probability, U or D . If player one plays U in the revelation period a block $k > 0$, then block $k - 1$ starts. If he plays U in the revelation period of block 1, then the continuation play is the indefinite play of (U, L) . If player one plays D in any revelation period, then the two players play an equilibrium of the complete information repeated game with a payoff profile that we will specify below.

Incentives of player two: Fix a small number $e > 0$. We now pick continuation payoffs such that player two is incentivized to ensure that player one's payoff for the review phase is below e with probability at least $1 - e$.

The continuation payoff profile, starting from the first period after a current block k , is equal to $(u_1^{rev}(k, \delta), u_2^{pass}(k, \delta))$ if player one plays D and if player one's discounted average payoff during the review phase is below e ; it is equal to $(u_1^{rev}(k, \delta), u_2^{fail}(k, \delta))$ if player one plays D but player one's discounted average payoff during the review phase is greater than e .

Suppose now that $u_2^{pass}(k, \delta) - u_2^{fail}(k, \delta) > v$. Intuitively, if player one reveals rationality at the end of the block, then player two enjoys a high continuation payoff if player one's realized payoff for the review phase is low, i.e., if she passes the review; but her continuation payoff is low otherwise. However, pushing player one's payoff below e requires player two to play a myopic nonbestresponse during the review phase. Therefore, player two faces a tradeoff between her payoff in the review phase and her continuation payoff. If the probability that player one is the commitment type at the start of block k is below a certain cutoff \bar{z} , then a patient player two's best response to the block strategy of player one would bring player one's payoffs below e with high probability.

The length of the review stage T^* is chosen such that $\delta^{T^*} = r$ where $r \in (\frac{2}{2+v(1-\bar{z})}, 1)$ is an arbitrary constant. Also, if δ exceeds a cutoff $\delta(e)$, then the length of the review stage T^* ensures that player two's best response to player one's block strategy gives player one a discounted average payoff of more than e with a probability less than e (see Appendix C.6).³⁰

Iteration over blocks and continuation payoffs: Since the choice of e was arbitrary, if player one's reputation level at the beginning of block one is less than \bar{z} , then a patient player one's equilibrium payoff at that period is at most r . If $r < \varepsilon$, then an equilibrium construction in which there is a single block achieves our desiderata.

If $r \geq \varepsilon$, then consider the continuation payoff of player one at the revelation period of block two. Suppose that player one's reputation level at the beginning of block two is z_2 . Moreover, suppose that if he plays U in the revelation period, i.e., does not reveal his rationality, then his reputation level is $z_1 < \bar{z}$.³¹ Let $u_1(block1, \delta)$ denote player one's continuation payoff at the start of the last block, i.e, block 1. If player one plays U in the revelation period of block 2, then his current period payoff is $1 - \delta$ and his continuation payoff is $\delta u_1(block1, \delta)$. Alternatively, if he plays D , then his current period payoff is zero and his continuation payoff is $\delta u_1^{rev}(\delta)$. We choose $u_1^{rev}(2, \delta)$ to ensure that player one is indifferent between playing U and D in the revelation period, and hence, $u_1^{rev}(2, \delta) = u_1(block1, \delta) + (1 - \delta)/\delta$. Notice that for δ close to one, $u_1^{rev}(2, \delta)$ is above ε and in the interior of the feasible and IR payoff set. Therefore, it is possible to pick PBE payoff profiles of the repeated game with complete information

³⁰In particular, if

$$(5) \quad \frac{1-z}{2}v\delta^{T^*} > 1 - \delta^{T^*},$$

and if T^* exceeds a cutoff $T(e)$ that only depends on e , then player two's best to player one's block strategy gives player one a discounted average payoff of more than e with a probability less than e . For any δ , we pick T^* such that $\delta^{T^*} = r < 1$, and we pick z for which inequality (5) is satisfied. Notice if δ exceeds a cutoff, then the T^* chosen in this manner exceeds $T(e)$. Notice also that inequality (5) is independent of the choice of e .

³¹More precisely, his initial reputation level is such that $z_1 = \frac{z_2}{z_2 + \frac{1-z_2}{2}} \leq \bar{z}$.

$(u_1^{rev}(2, \delta), u_2^{pass}(2, \delta))$ and $(u_1^{rev}(2, \delta), u_2^{fail}(2, \delta))$ such that $u_2^{pass}(2, \delta) - u_2^{fail}(2, \delta) \geq v$ if δ is sufficiently large. For this reason, a patient player two's best response delivers to player one a payoff of more than e during the review phase of block 2, with a probability at most e . Again, since e is arbitrary, a patient player one's payoff at the beginning of block two is at most r^2 .

Iterating this procedure k times and ratcheting down the reputation levels z_k , we can thus construct an equilibrium where a patient player one's payoff is no more than ε . The number of iterations is at most equal to the smallest integer k^* such that $r^{k^*} \leq \varepsilon$. Hence, if a patient player one's initial reputation level is below z_{k^*} , then there is an equilibrium in which player one's payoff is at most ε .

5. DISCUSSION

5.1. Full-support imperfect monitoring. In order to further motivate the use of the FS assumption, we contrast our reputation result for the common-interest game in figure 5 with the folk theorem of Cripps and Thomas (1997) for this game under perfect monitoring. The equilibrium construction in Cripps and Thomas (1997) is as follows: In the first K periods, as long as player one has not revealed rationality, player two plays R and player one uses a mixed strategy that plays D with small probability on the equilibrium path. After the first K periods, if player one has not revealed rationality, then (L, U) is played forever. If player one reveals rationality on the equilibrium path during the first K periods, then continuation payoffs are chosen to equal player one's continuation payoff had he not revealed rationality.

To ensure that player two has an incentive to play R , she is punished if she deviates by playing L , and the public signal (l, D) is observed. Notice that because of perfect monitoring, player two is punished with certainty in the event that player one reveals rationality. Punishment entails a continuation payoff for player two that is close to her minimax.³² Because player two only plays R on the equilibrium path, player one is indifferent between playing U and D in the first K periods. In this construction, player two is deterred from playing L , even if player one reveals rationality with a small probability in each period, because her continuation payoff is close to her minimax if public signal (l, D) is observed. However, if the probability that player one reveals rationality is small in each period, then it takes many periods for player one to build a reputation and K can be chosen large to ensure low payoffs for both players.

This argument hinges on choosing low continuation payoffs for player two after public signal (l, D) , during the first K periods. This does not conflict with player one's incentive to play D instead of U , even if low continuation payoffs for player two also imply low

³² After (l, D) or (r, D) , the continuation game is a repeated game of complete information and any individually rational payoff can be supported in equilibrium.

continuation payoffs for player one, after (l, D) . This is because, in the first K periods, player one expects player two to play L with probability zero and hence expects the public signal (l, D) to occur *with probability zero*. Thus, *payoffs after public signal (l, D) have no effect on player one's ex-ante incentive to play D ; consequently, player one's incentive constraint puts no restrictions on player two's continuation payoff after (l, D)* . In contrast, if the FS assumption is satisfied, then, as shown in Claim 2 and discussed in Remark 4, player one's incentive constraint implies a bound on player two's continuation payoff after any signal realization, i.e., player one's incentive to play D instead of U imposes a bound on the amount of punishment that player two can expect after choosing L . This is because the signal l occurs with positive probability regardless of which action player two chooses.

5.2. Imperfect monitoring in extensive-form stage games. In extensive-form stage games, making the assumption, as we do, that player one's stage-game actions are perfectly monitored is restrictive. This is because the assumption requires that player one's whole contingent plan be observed at the end of the period.

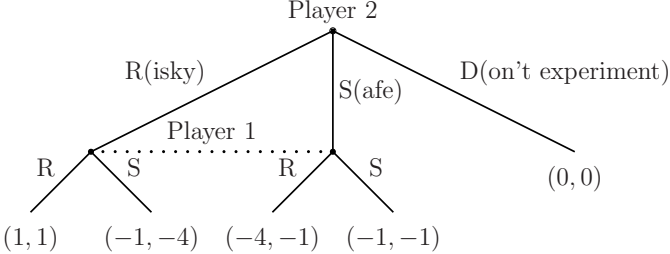


Figure 8: Experimenting with whether to play a coordination game.

Consider the example depicted in figure 8. In this game player two can choose whether to experiment or not. If she chooses not to experiment, then she guarantees a payoff equal to zero. If she chooses to experiment, then a simultaneous-move coordination game is played. The profile (R, R) is a NE of the stage game, which is better than not experimenting; and (S, S) is the other NE, which is worse than not experimenting. In this example, player one has two information sets: one after player two plays D and another after player two plays R or S . In this game our monitoring structure requires that player two see what player one would have chosen after R or S even when she actually chooses D .

However, a reputation result also holds under the following monitoring structure: First, only player one's moves are observed, i.e., if player two observes whether player one plays R or S only if signal r or s is realized. Second, regardless of what player two chooses as an action, she nevertheless ends up playing each of her three moves with probability greater than $\underline{\pi} \in (0, 1/1/4)$. Finally, when player one makes his move, he does not observe player

two's intended action but only the resulting choice of move, i.e, the signal. We depict this situation by adding a move by nature in figure 9. The essence of what we need for our reputation result is captured well by this example: we need player two to tremble and play each of her moves with positive probability, and we need this probability of a tremble to remain bounded strictly away from zero as we take the discount factor to one.

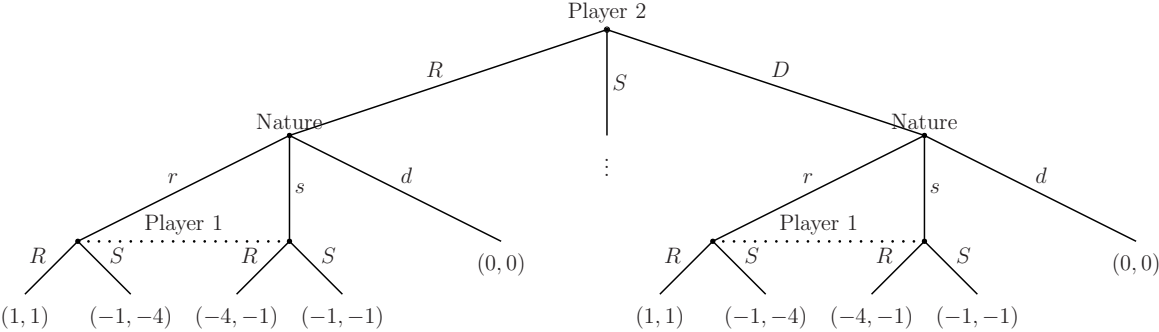


Figure 9: The stage game with nature's moves. After any move that player two makes, a similar continuation game is played where the probability with which nature mixes over $\{r, s, d\}$ is determined by player two's move. (We've suppressed the continuation game after S for brevity.) Specifically, nature mixes with the following probabilities: $\pi(r|R) = \pi(s|S) = \pi(d|D) = 1 - 2\underline{\pi}$ and $\pi(r|\text{not } R) = \pi(s|\text{not } S) = \pi(d|\text{not } D) = \underline{\pi}$.

5.3. 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 SCI. A prominent example of a game that falls outside of the class \mathcal{G} is the product-choice game depicted in figure 10. 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.³³ We discuss the repeated product-choice game to illustrate what can go wrong in games outside of the class \mathcal{G} .

This game is not an SA game; hence, Theorem 3 implies that 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

³³ 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.

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

Figure 10: A product choice game.

shortfall. However, consider the type ω that plays H with probability $1/2 + (1/2)^2$ in periods $\{1, \dots, 4\}$, plays H with probability $1/2 + (1/2)^3$ in periods $\{5, \dots, 8\}$, and more generally plays H with probability $1/2 + (1/2)^k$ in periods $\{2^{(k-1)} + 1, \dots, 2^k\}$. 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 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.³⁴ 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.

³⁴ A mixed strategy is then a probability distribution over choices of $p \in [0, 1]$, i.e., a mixed strategy is an element of $\Delta([0, 1])$.

A. FINITE AUTOMATA, LEARNING, AND LEMMA 1

In this part of the appendix, in addition to Lemma 1, 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 $\vec{\theta} = (\theta_{\omega^*}, \theta_1, \dots, \theta_{|W|})$. 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 process over the vector of states. In particular, the transition are given by the following equation:

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

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 $n = 1$,

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

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 $\vec{\theta}^j \in \Theta^j$ and $n > 0$ we have $\Pr(\vec{\theta}^{t+n} = \vec{\theta}^j | \vec{\theta}^t = \vec{\theta}(h^t), h^t, \sigma_2) > 0$ for some σ_2 and h^t if and only if $\Pr(\vec{\theta}^{k+n} = \vec{\theta}^j | \vec{\theta}^k = \vec{\theta}(\hat{h}^k), \hat{h}^k, \sigma_2') > 0$ 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:

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

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 $p(\omega^*, \omega, \vec{\theta}^t) := o(\omega, a^*(\vec{\theta}^t), \theta_\omega^t)$. 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 $\bar{p}(\omega^*, W) < 1$.³⁵

Define the likelihood ratio recursively as $L_t^\omega(h) = p(\omega, \vec{\theta}^t(h)) L_{t-1}^\omega(h)$ and let $L_0^\omega(h) = L_0^\omega = \mu(\omega) / \mu(\omega^*)$. Hence, $L_t^\omega(h) = \mu(\omega | h^t) / \mu(\omega^* | h^t)$ and (L_t^ω, h^t) is a supermartingale under $\Pr_{(\omega^*, \sigma_2)}$ (Fudenberg and Levine (1992) Lemma 4.1). Also, let $L_0^W = \mu(W) / \mu(\omega^*)$.

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*

$$\Pr_{(\omega^*, \sigma_2)} \left\{ h : \frac{\mu(W_{-\omega^*}(h^t) | h^t)}{\mu(\omega^* | h^t)} < \phi L_0^W \right\} > 1 - \epsilon,$$

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

³⁵The 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 $p(\omega, \vec{\theta}) \neq 1$

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 $\vec{\theta}^j$ and $\vec{\theta}^{j'}$ in the same recurrent class Θ^j we have that $\Pr(\vec{\theta}^{j'+|\Theta^j|} = \vec{\theta}^j | \vec{\theta}^j = \vec{\theta}^{j'}, h^t, \sigma_2) > \underline{\pi}^{|\Theta^j|} \geq \underline{\pi}^{|\Theta|}$ 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^* , 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 $\omega \in W_{-\omega^*(h^t)} = W_{-\omega^*(h^{l^*})}$, 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. \square

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. \square

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*

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

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

- (i). For any stationary strategy $\sigma_2 \in O_2$, $\lim_{\delta \rightarrow 1} U_i(\omega, \sigma_2, \delta) = U_i(\omega, \sigma_2) = \hat{U}_i(\omega, \sigma_2)$.
- (ii). The set $F(\omega, \delta) = \text{co}\{(U_1(\omega, \sigma_2, \delta), U_2(\omega, \sigma_2, \delta)) : \sigma_2 \in O_2\}$ and $F(\omega) = \text{co}\{(U_1(\omega, \sigma_2), U_2(\omega, \sigma_2)) : \sigma_2 \in O_2\}$, 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 $\Pr(\theta(h^{k+|\Theta^i|})|\theta(h^k), \sigma_2(h^k)) \geq \underline{\pi}^{|\Theta^i|}$ 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

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

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

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.
 \square

A.1. Proof of Lemma 1.

PROOF: If part is straightforward since the type that plays a_1^s every period is a one state pure strategy finite automaton with no shortfall. We now turn to the only if part: Lemma A.3 implies that there is a $\delta^* < 1$, and a pure stationary strategy $o_2 : \Theta \rightarrow A_2$ which is a best response to σ_1 for any $\delta > \delta^*$. Suppose that SA is not satisfied but there exists a pure strategy finite automata repeated game strategy σ_1 such that $U_1^C(\sigma_1|h^t) = \bar{g}_1$ for some $h^t \in H(\sigma_1)$. Since the automaton is finite, there should be a recurrent class of the states of the automaton Θ^j such that $U_1^C(\sigma_1|h^t) = \bar{g}_1$ for any h^t such that $\theta(h^t) \in \Theta^j$.

We will first show that $g_1(o(\theta), o_2(\theta)) = \bar{g}_1$ and $g_2(o(\theta), o_2(\theta)) = g_2^b$ for any state $\theta \in \Theta^j$.

Case 1: $g_2^b > \hat{g}_2$. In this case $\bar{g}_1 = \max\{g_1 : (g_1, g_2) \in F\}$. Combining this with the fact that $U_1^C(\sigma_1|h^t) = \bar{g}_1$ for any h^t with $\theta(h^t) \in \Theta^j$ and FS assumption we get $g_1(o(\theta), o_2(\theta)) = \bar{g}_1$ for any $\theta \in \Theta^j$. No gap together with $g_1(o(\theta), o_2(\theta)) = \bar{g}_1$ for any state $\theta \in \Theta^j$ implies that $g_2(o(\theta), o_2(\theta)) = g_2^b$.

Case 2: $g_2^b = \hat{g}_2$. $U_1^C(\sigma_1|h^t) = \bar{g}_1$ for any h^t with $\theta(h^t) \in \Theta^j$ implies that $\lim_{\delta \rightarrow 1} U_2(\sigma_1, \sigma_2, \delta|h^t) = \hat{g}_2$. Player two can guarantee herself \hat{g}_2 in each recurrent state. Suppose that $g_2(o(\theta_1), o_2(\theta_1)) < \hat{g}_2$ for some state $\theta_1 \in \Theta^j$. Then there must exist another state $\theta_2 \in \Theta^j$ such that $g_2(o(\theta_2), o_2(\theta_2)) > \hat{g}_2$, because otherwise player two's payoff after a history h^t with $\theta(h^t) = \theta_1$, $U_2(\sigma_1, \sigma_2, \delta|h^t)$, would be below her minimax which would contradict that σ_2 is a best response to σ_1 . However, then an alternative strategy σ_2' for player two that gets at least \hat{g}_2 in any state other than θ_2 and gets $g_2(o(\theta_2), o_2(\theta_2)) > \hat{g}_2$ at state θ_2 would have the feature that $\lim_{\delta \rightarrow 1} U_2(\sigma_1, \sigma_2, \delta|h^t) > \hat{g}_2$ from the FS assumption, hence contradicting that σ_2 is a best response to σ_1 . Consequently, there is no recurrent state $\theta \in \Theta^j$ for which $g_2(o(\theta), o_2(\theta)) < \hat{g}_2$. If for every $\theta \in \Theta^j$ we have $g_2(o(\theta), o_2(\theta)) \geq \hat{g}_2$ and if we have $U_1^C(\sigma_1|h^t) = \bar{g}_1$ then $g_1(o(\theta), o_2(\theta)) = \bar{g}_1$ for any $\theta \in \Theta^j$. If $g_2^b = \hat{g}_2$, then $g_2(o(\theta), o_2(\theta)) \geq \hat{g}_2$ for any state $\theta \in \Theta^j$. However, no gap together with $g_1(o(\theta), o_2(\theta)) = \bar{g}_1$ and $g_2(o(\theta), o_2(\theta)) \geq \hat{g}_2$ for any state $\theta \in \Theta^j$ implies that $g_2(o(\theta), o_2(\theta)) = g_2^b$.

We now complete the argument. If assumption SA does not hold then for any $a_1 \in A_1$, there exists $a_2 \in A_2$ which is a best response to a_1 and $g_1(a_1, a_2) < \bar{g}_1$. Consequently, for any $\theta \in \Theta^j$, there is a $a_2(\theta) \in A_2$ such that $a_2(\theta)$ is a best response to $o(\theta)$ and $g_1(o(\theta), a_2(\theta)) < \bar{g}_1$. Since $a_2(\theta)$ is a best response to $o(\theta)$ it must be that $g_2(o(\theta), a_2(\theta)) \geq g_2(o(\theta), o_2(\theta)) = g_2^b$ for each $\theta \in \Theta^j$. Consequently, the repeated game strategy for player two obtained by replacing $o_2(\theta)$ by $a_2(\theta)$ gives player two a weakly higher repeated game payoff. However, this new strategy gives player one a payoff that is strictly lower than \bar{g}_1 for any $\theta \in \Theta^j$, which is a

contradiction. □

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:

$$(6) \quad g_2 \leq -\rho g_1, \text{ for any } (g_1, g_2) \in F; \text{ and } g_2 \geq \rho g_1, \text{ for any } (g_1, g_2) \in G.$$

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

$$(7) \quad g_2 \geq \rho g_1, \text{ for any } (g_1, g_2) \in F.$$

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

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

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 .

B.2. The reputation bound. We first prove Lemma B.1 given below in this section. We then prove Lemma 2 using Lemma B.1 and Corollary A.1.

LEMMA B.1 Fix any $\delta \in [0, 1)$, $z \in (0, 1]$, $\phi \geq 0$. 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:

$$(8) \quad \underline{U}(\delta, \phi, z) \geq -b - f(l(\omega^*), \underline{z})\epsilon, \text{ where}$$

$$(9) \quad f(l(\omega^*), \underline{z}) := \bar{K}^{\bar{n}},$$

$$(10) \quad \bar{K}(l(\omega^*)) := \frac{1}{\underline{z}l\pi} \max\{2l(\omega^*), 8\rho, 4M + l(\omega^*) + K_1 + \rho(5 + 8M)\},$$

$$(11) \quad \bar{n}(\underline{z}, l(\omega^*)) := \text{the smallest integer } j \text{ s.t. } \left(1 - \frac{\pi l(\omega^*)\underline{z}}{4\rho}\right)^{j-1} < \underline{z}.$$

We now proceed with the proof of Lemma B.1. As we will keep ϕ fixed for the rest of the development, we will drop reference to ϕ from the function $\underline{U}(\delta, \phi, z)$. The following lemma bounds player one's payoff in histories where he deviates from the strategy ω^* with positive probability.

LEMMA B.2 (Player one's incentive constraint) *Pick any $z > 0$, $\mu \in \Delta_{\omega^*, \phi, z}$, $\sigma \in NE(\Gamma^\infty(\mu, \delta))$ and public history h^t such that $h^t \in H(\omega^*)$ and $h^t \in H(\sigma_1(N))$. Let $a_1^* = \omega^*(h^t)$, $\alpha_2 = \sigma_2(h^t)$. Suppose that $\mu(\omega^*|h^t, a_1^*, y) \geq z$ for all $y \in Y$. Suppose player one plays $a_1 \neq a_1^*$ with positive probability in period t , then $U_1(\sigma, \delta|h^t, a_1, \alpha_2) \geq \underline{U}(\delta, z) - 2M\epsilon$.*

PROOF: Notice $a_1 \in \text{supp}(\sigma_1(N, h^t))$ implies that player one's repeated game payoff from playing a_1 must be at least as large as his repeated game payoff from playing strategy ω^* from period t onwards. Consequently,

$$(1 - \delta)g_1(a_1, \alpha_2) + \delta U_1(\sigma, \delta|h^t, a_1, \alpha_2) \geq (1 - \delta)g_1(a_1^*, \alpha_2) + \delta U_1(\omega^*, \sigma_2, \delta|h^t, a_1^*, \alpha_2).$$

Let θ_y denote ω^* 's state after history (h^t, a_1^*, y) . Recall $\omega|_{(h^t, a_1^*, y)}$ is the continuation strategy induced by history (h^t, a_1^*, y) . For each $y \in Y$, define measure $\mu_y \in \Delta_{\omega_{\theta_y}}$ as $\mu_y(\omega|_{(h^t, a_1^*, y)}) = \mu(\omega|h^t, a_1^*, y)$ for each $\omega \in \Sigma_1$ and $\mu_y(N) = \mu(N|h^t, a_1^*, y)$. Notice, $\mu_y(\omega_{\theta_y}^*) \geq z$ and $\mu_y(\Sigma_1 \setminus \omega_{\theta_y}^*)/\mu_y(\omega_{\theta_y}^*) \leq \phi$ for each $y \in Y$. Because $(h^t, a_1^*, y) \in H(\omega^*)$ for any $y \in Y$, Remark 1 implies that $\sigma_2|_{(h^t, a_1^*, y)}$ is part of a NE profile of $\Gamma^\infty(\delta, \mu_y)$ for any $y \in Y$. Therefore, Definition B.1 implies that $U_1(\omega^*, \sigma_2, \delta|h^t, a_1^*, \alpha_2) \geq \underline{U}(\delta, z)$. Observing that $g_1(a_1^*, \alpha_2) - g_1(a_1, \alpha_2) \geq -2M$ and rearranging we get

$$U_1(\sigma, \delta|h^t, a_1, \alpha_2) \geq -2M(1 - \delta)/\delta + U_1(\sigma, \delta|h^t, a_1^*, \alpha_2) \geq -2M(1 - \delta)/\delta + \underline{U}(\delta, z).$$

□

COROLLARY B.1 *Suppose that $\hat{g}_2 < 0$. Pick any $z > 0$, $\mu \in \Delta_{\omega^*, \phi, z}$, $\sigma \in NE(\Gamma^\infty(\mu, \delta))$ and public history h^t such that $h^t \in H(\omega^*)$ and $h^t \in H(\sigma_1(N))$. Let $a_1^* = \omega^*(h^t)$ and suppose that*

$\mu(\omega^*|h^t, a_1^*, y) \geq z$ for each $y \in Y$. Suppose player one plays $a_1 \neq a_1^*$ with positive probability in period t , then

$$U_1(\sigma, \delta|h^t, a_1, y) \geq \frac{1}{\underline{\pi}}(\underline{U}(\delta, z) - 2M\epsilon), \text{ for all } y \in Y.$$

PROOF: Let $\alpha_2 = \sigma_2(h^t)$. Notice FS implies that the probability of any signal $y \in Y$ is at least $\underline{\pi}$ under α_2 . Also, player one's highest feasible payoff is equal to zero because $\hat{g}_2 < 0$. By definition, $\underline{U}(\delta, z) - 2M(1 - \delta)/\delta \leq 0$. However, the fact that player one's highest payoff is equal to zero, $\underline{U}(\delta, z) - 2M(1 - \delta)/\delta \leq 0$, and FS together imply that player one's continuation payoff after any y , $U_1(\sigma, \delta|h^t, a_1, y)$, must be at least $\frac{1}{\underline{\pi}}(\underline{U}(\delta, z) - 2M(1 - \delta)/\delta)$. Because otherwise the expectation under α_2 cannot achieve the lower bound given in Lemma B.2. \square

The following lemma converts the bound on player one's continuation payoff into a bound on player two's continuation payoff using inequalities (6) and (7).

LEMMA B.3 (Payoff conversion) *Pick any $\mu \in \Delta_{\omega^*, \phi, z}$, $\sigma \in NE(\Gamma^\infty(\mu, \delta))$ and history h^t such that $h^t \in H(\omega)$ and $h^t \in H(\sigma_1(N))$. Let $a_1^* = \omega^*(h^t)$, $\alpha_2 = \sigma_2(h^t)$ and suppose that $\mu(\omega^*|h^t, a_1^*, y) \geq z$ for each $y \in Y$. Suppose player one plays $a_1 \neq a_1^*$ with positive probability in period t , then*

$$U_2(\sigma_1(N), \sigma_2, \delta|h^t, a_1, \alpha_2) \leq -\rho(\underline{U}(\delta, z) - 2M\epsilon).$$

Moreover, if $\hat{g}_2 < 0$, then

$$U_2(\sigma_1(N), \sigma_2, \delta|h^t, a_1, y) \geq \frac{\rho}{\underline{\pi}}(\underline{U}(\delta, z) - 2M\epsilon), \text{ for any } y \in Y.$$

PROOF: By definition we have $(U_1(\sigma, \delta|h^t, a_1, \alpha_2), U_2(\sigma_1(N), \sigma_2, \delta|h^t, a_1, \alpha_2)) \in F$ and $U_1(\sigma, \delta|h^t, a_1, \alpha_2) \geq \underline{U}(\delta, z) - 2M(1 - \delta)/\delta$, by Lemma B.2. Consequently, inequality (6) implies that

$$U_2(\sigma_1(N), \sigma_2, \delta|h^t, a_1, \alpha_2) \leq \rho(-\underline{U}(\delta, z) + 2M(1 - \delta)/\delta).$$

Similarly, if $\hat{g}_2 < 0$, then Corollary B.1 and inequality (7) implies that

$$U_2(\sigma_1(N), \sigma_2, \delta|h^t, a_1, y) \geq \frac{\rho}{\underline{\pi}}(\underline{U}(\delta, z) - 2M(1 - \delta)/\delta), \text{ for any } y \in Y.$$

\square

We will now establish upper bound (Lemma B.6) and a lower bound (Lemma B.7) on

player two's equilibrium payoffs. However, first we define and elaborate on some stopping times and events which we will utilize in establishing the upper and lower bounds.

DEFINITION B.2 (Stopping time and events) *Fix a strategy profile σ , a measure $\mu \in \Delta_{\omega_\theta^*}$ and a constant $z' \in (0, 1]$.*

- (i). **Stopping time τ_1** : *For any infinite history h , let $\tau_1(h)$ be the first period t such that h^t contains a deviation from ω_θ^* and if such a period t does not exist, then $\tau_1(h) = \infty$. In words, if $\tau_1(h) = t$, then player one's action is identical to $a^*(h^k)$ in all periods $k < t - 1$, and player one's action in period $t - 1$ is different from $a^*(h^{t-1})$. The event $\{h : \tau_1(h) \leq t\}$ is h^t -measurable for any t thus τ_1 is a stopping time.*
- (ii). **Stopping time τ_2** : *For any infinite history h , let $\tau_2(\sigma, \mu, z', h)$ be the first period t such that $\mu(\omega_\theta^*|h^t) < z'$ and $\mu(\omega_\theta^*|h^t, a^*(h^t), y) \geq z'$ for any $y \in Y$. If such a period t does not exist, then $\tau_2(\sigma, \mu, z', h) = \infty$. The event $\{h : \tau_2(\sigma, \mu, z', h) \leq t\}$ is h^t -measurable for any t thus τ_2 is a stopping time.*
- (iii). **Event E_1** : *Let $E_1(\sigma, \mu, z') = \{h : \tau_1(h) < \tau_2(\sigma, \mu, z', h)\}$.*
- (iv). **Event E_2** : *Let $E_2(\sigma, \mu, z') = \{h : \tau_1(h) \geq \tau_2(\sigma, \mu, z', h)\}$ If both $\tau_1(h) = \infty$ and $\tau_2(\sigma, \mu, z', h) = \infty$, then we say that $\tau_1(h) \geq \tau_2(\sigma, \mu, z', h)$.*

Notice that the two events E_1 and E_2 are complements of each other. The following two lemmas are intermediary results that expand on the properties of the above defined stopping times and events. We use these intermediary results in establishing the upper and lower bound on player two's equilibrium payoffs.

LEMMA B.4 *For any $z, z' \in (0, 1]$, $\mu \in \Delta_{\omega_\theta^*, z}$, and any strategy profile σ , $\Pr_{\sigma, \mu}[E_1(\sigma, \mu, z')] \leq \max\{1 - z/z', 0\}$ where $\Pr_{\sigma, \mu}$ denotes player two's belief over the set of all histories if the players use strategy profile σ and if player two's prior over player one's types is μ .*

PROOF: Player one's reputation level in period t , $\mu(\omega_\theta^*|h^t)$, is a uniformly bounded martingale under $\Pr_{\sigma, \mu}$. For any finite constant T , the random variable $\tau = \tau_1 \wedge \tau_2 \wedge T$ is a stopping time because τ_1 and τ_2 are stopping times. Also, $\Pr_{\sigma, \mu}\{\tau < \infty\} = 1$. Consequently, the optional stopping theorem implies that $E[\mu(\omega_\theta^*|h^\tau)] = E[\mu(\omega_\theta^*|h^0)] = \mu(\omega_\theta^*|h^0) \geq z$. Also,

$$E[\mu(\omega_\theta^*|h^\tau)] = \Pr_{\sigma, \mu}\{\tau_1 < \tau_2 \wedge T\}E[\mu(\omega_\theta^*|h^\tau)|\tau_1 < \tau_2 \wedge T] + \Pr_{\sigma, \mu}\{\tau_1 \geq \tau_2 \wedge T\}E[\mu(\omega_\theta^*|h^\tau)|\tau_1 \geq \tau_2 \wedge T].$$

Notice that $E[\mu(\omega_\theta^*|h^\tau)|\tau_1 < \tau_2 \wedge T] = 0$. Therefore,

$$E[\mu(\omega_\theta^*|h^\tau)] = \Pr_{\sigma, \mu}\{\tau_1 \geq \tau_2 \wedge T\}E[\mu(\omega_\theta^*|h^\tau)|\tau_1 \geq \tau_2 \wedge T].$$

However, $E[\mu(\omega_\theta^*|h^\tau)|\tau_1 \geq \tau_2 \wedge T] \leq z'$ and hence $z \leq E[\mu(\omega_\theta^*|h^\tau)] \leq z' \Pr_{\sigma,\mu}\{\tau_1 \geq \tau_2 \wedge T\}$. Thus, the monotone convergence theorem implies that $z \leq z' \lim_{T \rightarrow \infty} \Pr_{\sigma,\mu}\{\tau_1 \geq \tau_2 \wedge T\} = z' \Pr_{\sigma,\mu}\{\tau_1 \geq \tau_2\}$. Hence, $z/z' \leq \Pr_{\sigma,\mu}\{\tau_1 \geq \tau_2\}$ and therefore $\Pr_{\sigma,\mu}\{\tau_1 < \tau_2\} \leq 1 - z/z'$. \square

LEMMA B.5 (Player two's payoff until player one deviates from ω_θ^* .) *Suppose that σ_2^b is a best response to ω_θ^* , then,*

$$\left| E_{\sigma_1, \sigma_2^b} \left[(1 - \delta) \sum_{s=0}^{\tau_1-1} \delta^s r_2(a_1^s, y^s) \right] \right| \leq \epsilon(2\rho + M), \text{ for any } \sigma_1,$$

where if $\tau_1 = \infty$, then $\sum_{s=0}^{\tau_1-1} \delta^s r_2(a_1^s, y^s) := \sum_{s=0}^{\infty} \delta^s r_2(a_1^s, y^s)$. Moreover,

$$E_{\sigma_1, \sigma_2} \left[(1 - \delta) \sum_{s=0}^{\tau_1-1} \delta^s r_2(a_1^s, y^s) \right] \leq \epsilon(2\rho + M) \text{ for any } \sigma_1 \text{ and } \sigma_2.$$

PROOF: Note that $P_{\sigma_1, \sigma_2}(h = (h^t, h^{-t})|t < \tau_1) = P_{\omega_\theta, \sigma_2}(h = (h^t, h^{-t})|t < \tau_1)$, for any (σ_1, σ_2) . Consequently, we have the following equality:

$$(12) \quad E_{\omega_\theta^*, \sigma_2} \left[(1 - \delta) \sum_{s=0}^{\tau-2} \delta^s r_2(a_1^s, y^s) \right] = E_{\sigma_1, \sigma_2} \left[(1 - \delta) \sum_{s=0}^{\tau-2} \delta^s r_2(a_1^s, y^s) \right].$$

We define h^t measurable the random variable x_t as follows:

$$x_t := (1 - \delta) \sum_{s=0}^{t-1} \delta^s r_2(a_1^s, y^s) + \delta^t U_2(\omega_\theta^*, \sigma_2^b, \delta|h^t)$$

Notice that

$$\begin{aligned} x_t &= (1 - \delta) \sum_{s=0}^{t-1} \delta^s r_2(a_1^s, y^s) + E_{\omega_\theta^*, \sigma_2^b}[(1 - \delta)\delta^t r_2(a_1^t, y^t) + \delta^{t+1} U_2(\omega_\theta^*, \sigma_2^b, \delta|h^{t+1})|h^t], \\ x_t &= E_{\omega_\theta^*, \sigma_2^b}[(1 - \delta) \sum_{s=0}^{t-1} \delta^s r_2(a_1^s, y^s) + (1 - \delta)\delta^t r_2(a_1^t, y^t) + \delta^{t+1} U_2(\omega_\theta^*, \sigma_2^b, \delta|h^{t+1})|h^t]. \end{aligned}$$

Also, observe that

$$x_{t+1} = (1 - \delta) \sum_{s=0}^{t-1} \delta^s r_2(a_1^s, y^s) + (1 - \delta)\delta^t r_2(a_1^t, y^t) + \delta^{t+1} U_2(\omega_\theta^*, \sigma_2^b, \delta|h^{t+1})$$

Consequently, $x_t = E_{\omega_\theta^*, \sigma_2^b}[x_{t+1}|h^t]$. Therefore, for every $t \geq 0$, $E_{\omega_\theta^*, \sigma_2^b}[x_{t+1} - x_t|h^t] = 0$. Hence, x_t is a martingale with respect to $P_{\omega_\theta^*, \sigma_2^b}$. For any integer $n \geq 1$, let $\tau(n) := \tau_1 \wedge n$.

Since each x_t is bounded, we can apply the optional sampling theorem to obtain the following

$$\begin{aligned} U_2(\omega_\theta^*, \sigma_2^b, \delta) &= E_{\omega_\theta^*, \sigma_2} [x_{\tau(n)-1}] \\ &= E_{\omega_\theta^*, \sigma_2^b} \left[(1 - \delta) \sum_{s=0}^{\tau(n)-2} \delta^s r_2(a_1^s, y^s) + \delta^{\tau(n)-1} U_2(\omega_\theta^*, \sigma_2^b, \delta | h^{\tau(n)-1}) \right]. \end{aligned}$$

However, equation (12) implies that

$$U_2(\omega_\theta^*, \sigma_2^b, \delta) = E_{\sigma_1, \sigma_2^b} \left[(1 - \delta) \sum_{s=0}^{\tau(n)-2} \delta^s r_2(a_1^s, y^s) + \delta^{\tau(n)-1} U_2(\omega_\theta^*, \sigma_2^b, \delta | h^{\tau(n)-1}) \right].$$

Since $U_2(\omega_\theta^*, \sigma_2^b, \delta | h^t) \leq \rho\epsilon$ for every h^t , applying the triangular inequality gives that

$$\left| E_{\sigma_1, \sigma_2^b} \left[(1 - \delta) \sum_{s=0}^{\tau(n)-2} \delta^s r_2(a_1^s, y^s) \right] \right| \leq 2\rho\epsilon.$$

Since this is true for any $n \geq 1$, and since the payoffs are bounded and are discounted, by the bounded convergence theorem we have the following

$$\left| E_{\sigma_1, \sigma_2^b} \left[(1 - \delta) \sum_{s=0}^{\tau_1-2} \delta^s r_2(a_1^s, y^s) \right] \right| \leq 2\rho\epsilon.$$

Consequently,

$$\left| E_{\sigma_1, \sigma_2^b} \left[(1 - \delta) \sum_{s=0}^{\tau_1-1} \delta^s r_2(a_1^s, y^s) \right] \right| \leq 2\rho\epsilon + (1 - \delta)M.$$

Let σ_2 be any strategy of player 2. Now x_t is a supermartingale because the continuation payoffs from strategy profile $(\omega_\theta^*, \sigma_2)$ can be at most $U_2(\omega_\theta^*, \sigma_2^b, \delta | h^t)$ at any history $h^t \in H(\omega_\theta^*)$. Hence, by the optional sampling theorem we have that

$$E_{\omega_\theta^*, \sigma_2} \left[\sum_{s=0}^{\tau(n)-2} \delta^s r_2(a_1^s, y^s) + \delta^{\tau(n)-1} U_2(\omega_\theta^*, \sigma_2^b, \delta | h^{\tau(n)-1}) \right] \leq U_2(\omega_\theta^*, \sigma_2^b, \delta).$$

Since $|U_2(\omega_\theta^*, \sigma_2^b, \delta | h^t)| \leq \rho\epsilon$ for every h^t , we have that

$$E_{\omega_\theta^*, \sigma_2} \left[(1 - \delta) \sum_{s=0}^{\tau(n)-2} \delta^s r_2(a_1^s, y^s) \right] \leq \rho 2\epsilon.$$

Since n is arbitrary and payoffs are bounded and are discounted, we have that

$$E_{\omega_\theta^*, \sigma_2} \left[(1 - \delta) \sum_{s=0}^{\tau_1-2} \delta^s r_2(a_1^s, y^s) \right] \leq 2\rho\epsilon.$$

Using equation (12) we conclude that $E_{\sigma_1, \sigma_2} [(1 - \delta) \sum_{s=0}^{\tau_1-1} \delta^s r_2(a_1^s, y^s)] \leq \rho 2\epsilon + (1 - \delta)M$.
 \square

Below we use Lemma B.3, Definition B.2, Lemma B.4, and Lemma B.5 to calculate the upper and lower bound for player two's payoffs. For $z, z' \in (0, 1]$, define $q(z, z') := \max\{z/z', 0\}$.

LEMMA B.6 (Upper-bound) *Suppose $z, z' \in (0, 1]$. Let $\xi > 0$, $\theta \in \Theta^*$, $\mu \in \Delta_{\omega_\theta^*, \phi, z}$, and $\sigma \in NE(\Gamma^\infty(\mu, \delta))$ be such that $U_1(\omega_\theta^*, \sigma_2, \delta) \leq \underline{U}(\delta, z) + \xi$. For the chosen NE σ ,*

$$(13) \quad U_2(\sigma, \delta, \mu) \leq -\rho q(z, z') \underline{U}(z, \delta) - \rho \underline{U}(z', \delta) + lz(\underline{U}(\delta, z) + \xi) + K\epsilon,$$

where $K = 2\rho + 2M + 4M\rho + l + K_1$.

PROOF: Suppose that $\tau_1(h)$ is infinite, i.e., player one never deviates from ω_θ . In this case, $U_2(\sigma, \delta, \mu | \tau_1 = \infty) = U_2(\omega_\theta^*, \sigma_2, \delta, \mu) \leq (\underline{U}(\delta, z) + \xi + b)l + K_1\epsilon \leq (\underline{U}(\delta, z) + \xi)l + \epsilon l + K_1\epsilon$, by the maintained assumption of Lemma B.1.

Suppose that $\tau_1(h)$ is finite. An upper bound on player two's total payoff for the periods zero through $\tau_1(h) - 1$ is given by Lemma B.5 and is at most $\epsilon(2\rho + M)$. We will bound player two's continuation payoff after $h^{\tau_1(h)}$ in the events player one is the normal type and $E_1(\sigma, \mu, z')$ occurs, the event that player one is the normal type and the event $E_2(\sigma, \mu, z')$ occurs, and the event that player one is any other type.

If $h \in E_1(\sigma, \mu, z')$, then $\mu(\cdot | h^{\tau_1(h)-1}, a_1^*(h^{\tau_1(h)-1}), y) \in \Delta_{\omega_\theta^*, \phi, z}$ for any $y \in Y$. Consequently, player two's expected continuation payoff satisfies the following inequality

$$U_2(\sigma_1(N), \sigma_2, \delta, \mu | h^{\tau_1(h)-1}, a_1^{\tau_1(h)-1}, \sigma_2(h^{\tau_1(h)-1})) \leq \rho(-\underline{U}(z, \delta) + 2M\epsilon),$$

by Lemma B.3. If $h \in E_2(\sigma, \mu, z')$, then $\mu(\cdot | h^{\tau_1(h)-1}, a_1^*(h^{\tau_1(h)-1}), y) \in \Delta_{\omega_\theta^*, \phi, z'}$ for any $y \in Y$. Consequently, player two's expected continuation payoff is at most $\rho(-\underline{U}(z', \delta) + 2M\epsilon)$ by Lemma B.3. The probability of event $E_1(\sigma, \mu, z')$ is at most $q(z, z')$, by Lemma B.4, the probability of event $E_2(\sigma, \mu, z')$ is at most one, the probability of $\omega \neq \omega_\theta^*$ is at most $\phi \leq \epsilon$. So,

$$U_2(\sigma, \delta, \mu | \tau_1 < \infty) \leq \epsilon(2\rho + M) + \rho(-q(z, z') \underline{U}(z, \delta) - \underline{U}(z', \delta) + 4M\epsilon) + \phi M.$$

However, as we argued before, $U_2(\sigma, \delta | \tau_1 = \infty) \leq l\epsilon + K_1\epsilon + (\underline{U}(\delta, z) + \xi)l$, and the probability that $\tau_1 = \infty$ is at least $z > 0$ because $\mu \in \Delta_{\omega_\theta^*, \phi, z}$. Consequently,

$$U_2(\sigma, \delta, \mu) \leq \epsilon(2\rho + 2M + 4M\rho + l + K_1) + \rho(-q(z, z') \underline{U}(z, \delta) - \underline{U}(z', \delta)) + (\underline{U}(\delta, z) + \xi)lz,$$

because $\underline{U} \leq 0$, because $\phi \leq \epsilon$, and because ρ , ϵ , l , and M are positive constants. Thus,

$$U_2(\sigma, \delta, \mu) \leq -\rho q(z, z')\underline{U}(z, \delta) - \rho\underline{U}(z', \delta) + lz(\underline{U}(\delta, z) + \xi) + K\epsilon,$$

where $K = 2\rho + 2M + 4M\rho + l + K_1$. □

LEMMA B.7 (Lower-bound) *Suppose $z, z' \in (0, 1]$. For any $\theta \in \Theta^*$, $\mu \in \Delta_{\omega_\theta^*, \phi, z}$ and any $\sigma \in NE(\Gamma^\infty(\mu, \delta))$,*

$$(14) \quad U_2(\sigma, \delta, \mu) \geq \frac{\rho}{\pi}q(z, z')\underline{U}(z, \delta) + \frac{\rho}{\pi}\underline{U}(z', \delta) - K\epsilon,$$

where $K = \rho(3 + \frac{4M}{\pi}) + 2M$.

PROOF: Notice that $\hat{g}_2 \leq 0$. This is because $g_2^b = 0$. If $\hat{g}_2 = 0$, then inequality (14) trivially holds because the righthand side is nonpositive.

Assume $\hat{g}_2 < 0$. Pick any $\sigma \in NE(\Gamma^\infty(\mu, \delta))$, pick $\sigma_2^b \in BR(\omega^*, \delta)$. Define strategy σ_2^* as follows: $\sigma_2^*(h^k) = \sigma_2^b(h^k)$, if period k public history $h^k \in H(\omega_\theta^*)$; and $\sigma_2^*(h^k) = \sigma_2(h^k)$, otherwise. Let strategy profile $\sigma^* = (\sigma_1, \sigma_2^*)$.

Suppose that $\tau_1(h)$ is infinite, i.e., player one never deviates from ω_θ^* . In this case, $U_2(\sigma, \delta, \mu | \tau_1 = \infty) = U_2(\omega_\theta^*, \sigma_2^*, \delta, \mu) \geq -b\rho$, by the maintained assumption of Lemma B.1.

Suppose that $\tau_1(h)$ is finite. A lower bound on player two's total payoff for the periods zero through $\tau_1(h) - 1$ is given by Lemma B.5 and is at least $-\epsilon(2\rho + M)$. We will bound player two's continuation payoff after $h^{\tau_1(h)}$ in the events player one is the normal type and $E_1(\sigma^*, \mu, z')$ occurs, the event that player one is the normal type and the event $E_2(\sigma^*, \mu, z')$ occurs, and the event that player one is any other type.

If $h \in E_1(\sigma^*, \mu, z')$, then $\mu(\cdot | h^{\tau_1(h)-1}, a_1^*(h^{\tau_1(h)-1}), y) \in \Delta_{\omega_\theta^*, \phi, z}$ for any $y \in Y$. Consequently, player two's continuation payoff after history $h^{\tau_1(h)}$ is at least $\frac{\rho}{\pi}(\underline{U}(z, \delta) - 2M\epsilon)$ by Lemma B.3. If $h \in E_2(\sigma^*, \mu, z')$, then $\mu(\cdot | h^{\tau_1(h)-1}, a_1^*(h^{\tau_1(h)-1}), y) \in \Delta_{\omega_\theta^*, \phi, z'}$ for any $y \in Y$. Consequently, player two's continuation payoff after history $h^{\tau_1(h)}$ is at least $\frac{\rho}{\pi}(\underline{U}(z', \delta) - 2M\epsilon)$ by Lemma B.3. The probability of event $E_1(\sigma^*, \mu, z')$ is at most $q(z, z')$, by Lemma B.4, the probability of event $E_2(\sigma^*, \mu, z')$ is at most one, the probability of $\omega \neq \omega_\theta^*$ is at most ϕ . Also, the probability that $\tau_1(h)$ is infinite is at most one. So,

$$U_2(\sigma^*, \delta, \mu | \tau_1 < \infty) \geq -\epsilon(2\rho + M) + \frac{\rho}{\pi}(q(z, z')\underline{U}(z, \delta) + \underline{U}(z', \delta) - 4M\epsilon) - \phi M - b\rho.$$

Thus,

$$U_2(\sigma^*, \delta, \mu) \geq \frac{\rho}{\underline{\pi}} q(z, z') \underline{U}(z, \delta) + \frac{\rho}{\underline{\pi}} \underline{U}(z', \delta) - K\epsilon,$$

where $K = \rho(3 + \frac{4M}{\underline{\pi}}) + 2M$. □

LEMMA B.8 *Fix $\underline{z} \in (0, 1]$, $\mu \in \Delta_{\omega^*, \underline{z}, \phi}$. Let $\bar{q} = \underline{\pi} l \underline{z} / 4\rho$. For any $z, z' \in [\underline{z}, 1]$ such that $q(z, z') \leq \bar{q}$ the following inequality holds:*

$$(15) \quad \underline{U}(z, \delta) \geq K \underline{U}(z', \delta) - K\epsilon.$$

where $K = \frac{1}{2\underline{z}l\underline{\pi}} \max\{8\rho, 4M + l + K_1 + \rho(5 + 8M)\}$.

PROOF: The definition of $\underline{U}(\delta, z)$ implies that $\underline{U}(\delta, z) = 0$ or $\underline{U}(\delta, z) = v(\delta, z)$. If $\underline{U}(\delta, z) = 0$, then the inequality is obviously true. If $\underline{U}(\delta, z) = v(\delta, z)$, then for each $\xi > 0$, there is $\theta \in \Theta^*$, $\mu^\xi \in \Delta_{\omega_\theta^*, \phi, z}$, and $\sigma^\xi \in NE(\Gamma^\infty(\mu^\xi, \delta))$ such that $U_1(\omega_\theta^*, \sigma^\xi, \sigma) \leq \underline{U}(\delta, z) + \xi$. Lemma B.6 implies that $U_2(\sigma^\xi, \delta, \mu^\xi)$ satisfies inequality (13). Also, Lemma B.7 implies that $U_2(\sigma^\xi, \delta, \mu^\xi)$ satisfies inequality (14). Combining the two inequalities gives

$$(\underline{U}(z, \delta) + \xi) l z \geq 2 \frac{\rho}{\underline{\pi}} (q(z, z') \underline{U}(z, \delta) + \underline{U}(z', \delta)) - K\epsilon,$$

Since this equation is true for each ξ we get

$$\underline{U}(z, \delta) l z \geq 2 \frac{\rho}{\underline{\pi}} (q(z, z') \underline{U}(z, \delta) + \underline{U}(z', \delta)) - K\epsilon.$$

The fact that $z \geq \underline{z}$, $q(z, z') \leq \bar{q}$, and $\underline{U} \leq 0$ together imply the following:

$$\underline{U}(z, \delta) \left(l \underline{z} - 2 \frac{\bar{q} \rho}{\underline{\pi}} \right) \geq 2 \frac{\rho}{\underline{\pi}} \underline{U}(z', \delta) - K\epsilon.$$

Substituting $\underline{\pi} l \underline{z} / 4\rho$ for \bar{q} in the previous inequality and rearranging implies the following inequality:

$$\underline{U}(z, \delta) \geq 4 \frac{\rho}{\underline{\pi} l \underline{z}} \underline{U}(z', \delta) - \frac{K}{2 l \underline{z}} \epsilon.$$

Redefining the constant $K = \frac{1}{2\underline{z}l\underline{\pi}} \max\{8\rho, 4M + l + K_1 + \rho(5 + 8M)\}$ delivers the result. □

PROOF OF LEMMA B.1: If $z' = 1$, then $\underline{U}(z', \delta) \geq U_1^C(\omega^*, \delta) \geq -b$. Consequently, if $z \geq \underline{z}$, $z' = 1$, and $z \in [(1 - \bar{q}), 1]$, then inequality (15) implies that $\underline{U}(z, \delta) \geq -(b + \epsilon)K$. (Recall

that $\bar{q} = \underline{\pi}l\underline{z}/4\rho$.)

We will show that if $z \geq \underline{z}$ and if $z \in [(1 - \bar{q})^n, (1 - \bar{q})^{n-1}]$, then $\underline{U}(z, \delta) \geq -bK^{n-1} - \epsilon \sum_{j=1}^{n-1} K^j$ by using induction on n . We make the inductive hypothesis that if $z \geq \underline{z}$ and if $z \in [(1 - \bar{q})^{k-1}, (1 - \bar{q})^{k-2}]$, then we have $\underline{U}(z, \delta) \geq -bK^{k-1} - \epsilon \sum_{j=1}^{k-1} K^j$.

If $z \geq \underline{z}$, $z' = (1 - \bar{q})^{k-1}$, and $z \in [(1 - \bar{q})^k, (1 - \bar{q})^{k-1}]$, then inequality (15) and the inductive hypothesis together imply that

$$\underline{U}(z, \delta) \geq K \left(-bK^{k-1} - \epsilon \sum_{j=1}^{k-1} K^j \right) - K\epsilon.$$

Hence, $z \geq \underline{z}$ and $z \in [(1 - \bar{q})^k, (1 - \bar{q})^{k-1}]$, then $\underline{U}(z, \delta) \geq -bK^k - \epsilon \sum_{j=1}^k K^j$, completing the induction.

The definition of \bar{n} implies that $\underline{z} \in [(1 - \bar{q})^{\bar{n}-1}, (1 - \bar{q})^{\bar{n}-2}]$, and consequently, $\underline{U}(\underline{z}, \delta) \geq -bK^{\bar{n}-1} - \epsilon \sum_{j=1}^{\bar{n}-1} K^j$. If $K \geq 2$, then $\sum_{j=1}^{\bar{n}-1} K^j \leq K^{\bar{n}}$. Consequently, if $z \geq \underline{z}$, then $\underline{U}(z, \delta) \geq -b - \bar{K}^{\bar{n}}$ where $\bar{K} = \frac{1}{\underline{z}l\underline{\pi}} \max\{2l, 8\rho, 4M + l + K_1 + \rho(5 + 8M)\}$. \square

B.3. 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.9 (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: Lemma A.3, part (iv) implies that $U_1^C(\omega^*) = U_1^C(\omega_\theta^*)$ and consequently $d(\omega^*) = d(\omega_\theta^*)$. Applying Lemma A.3, part (iv) once again shows that $U_1^C(\omega_\theta^*, \delta) \geq -b$ for any $\theta \in \Theta^*$.

Pick mixed stationary strategy $\sigma_2^* \in \Delta(O_2)$ such that $(U_1(\omega_\theta^*, \sigma_2^*, \delta), U_2(\omega_\theta^*, \sigma_2^*, \delta)) = (U_1(\omega_\theta^*, \sigma_2, \delta), U_2(\omega_\theta^*, \sigma_2, \delta))$. Such a strategy exists by Lemma A.3, part (i). Lemma A.3 part (iv) together with $U_1(\omega_\theta^*, \sigma_2^*, \delta) \leq -b - x$ implies that $U_1(\omega_\theta^*, \sigma_2^*) \leq -b - x + M\epsilon$. Consider the finite linear program

$$v(y) = \max_{p \in \Delta(O_2)} \sum_{\sigma_2 \in O_2} p(\sigma_2) U_2(\omega_\theta^*, \sigma_2)$$

subject to $\sum_{\sigma_2 \in O_2} p(\sigma_2) U_1(\omega_\theta^*, \sigma_2) \leq U_1^C(\omega^*) - y$.

Lemma A.3, part (iv) implies that $U_1^C(\omega^*) = U_1^C(\omega_\theta^*)$. Hence, the definition of the commitment payoff, $U_1^C(\omega_\theta^*)$, implies that the constraint binds for all $y > 0$, and $v(0) = U_2(\omega_\theta^*, \sigma_2^b)$

where σ_2^b is a best response to ω_θ^* . Consequently, a standard result from linear programming implies that there exists $l > 0$ such that $v(y) \leq v(0) - ly$ for all $y \geq 0$ (see for example [Vohra \(2005\)](#), Theorem 4.5.1). Furthermore, $v(0) = U_2(\omega_\theta^*, \sigma_2^b)$ implies that $v(y) \leq v(0) - ly$ for all y .

Substituting $-x + M\epsilon$ for y implies that $U_2(\omega_\theta^*, \sigma_2^*) \leq v(0) - lx + lM\epsilon = U_2(\omega_\theta^*, \sigma_2^b) - lx + lM\epsilon$. Notice $U_2(\omega_\theta^*, \sigma_2^b) \leq b\rho$ by inequality (6). Consequently, $U_2(\omega_\theta^*, \sigma_2^*) \leq b\rho - lx + lM\epsilon$. However, Lemma A.3 part (iv) implies that $U_2(\omega_\theta^*, \sigma_2^*, \delta) \leq b\rho - lx + lM\epsilon + M\epsilon$ and therefore $U_2(\omega_\theta^*, \sigma_2, \delta) \leq b\rho - lx + lM\epsilon + M\epsilon$. Consequently, $U_2(\omega_\theta^*, \sigma_2, \delta) \leq -lx + K_1\epsilon$ where $K_1 = \rho + lM + 1$. \square

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: Fix $\mu \in \Delta_{\omega^*}$ and let $z = \mu(\omega^*)$. Pick $\xi > 0$. Let $\phi^\xi > 0$ be such that $\phi^\xi f(z, l(\omega^*)) \leq \xi/3$ and $\phi^\xi M \leq \xi/3$. Let $T^* = T(\omega^*, \mu^*, \phi^\xi)$ be such that for any $t \geq T^*$ we have

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

for any strategy σ_2 . Such a T^* exists by Corollary A.1. Pick δ^ξ such that $(1 - \delta^{T^*+1})M < \xi/3$ for all $\delta \geq \delta^\xi$.

For any $\sigma \in NE(\Gamma^\infty(\delta, \mu))$ and any history $h^{T^*} \in H(\omega^*)$ such that $\mu(\Sigma_1 \setminus \omega^*(h^{T^*})|h^{T^*})/\mu(\omega^*|h^{T^*}) < \phi^\xi$ by Lemma B.1 we have

$$U_1(\omega_\theta, \sigma_2, \delta|h^{T^*}) \geq -d(\delta) - f(z, l(\omega^*)) \max\{(1 - \delta^{|\Theta^*|})/(\delta\underline{\pi})^{|\Theta^*|}, d(\delta), \phi^\xi\},$$

where θ denotes ω^* 's state after history $h^{T^*} \in H(\omega^*)$ and $d(\delta) = d(\omega^*) + M(1 - \delta^{|\Theta^*|})/(\delta\underline{\pi})^{|\Theta^*|}$. The probability of $\mu(\Sigma_1 \setminus \omega^*(h^{T^*})|h^{T^*})/\mu(\omega^*|h^{T^*}) < \phi^\xi$ is at least $1 - \phi^\xi$ under $\Pr_{(\omega^*, \sigma_2)}$ for any σ_2 . Consequently, for any $\sigma \in NE(\Gamma^\infty(\delta, \mu))$, we have

$$U_1(\omega^*, \sigma_2, \delta) \geq (1 - \phi^\xi) \left(-(1 - \delta^{T^*+1})M + \delta^{T^*+1} (-d(\delta) - f(z, l(\omega^*))\epsilon(\delta, \xi)) \right) - \phi^\xi M,$$

where $\epsilon(\delta, \xi) = \max\{(1 - \delta^{|\Theta^*|})/(\delta\underline{\pi})^{|\Theta^*|}, d(\delta), \phi^\xi\}$. As $(-d(\delta) - f(z, l(\omega^*))\epsilon(\delta, \xi)) \leq 0$ and $-(1 - \delta^{T^*+1})M \leq 0$ the following holds

$$U_1(\omega^*, \sigma_2, \delta) \geq -(1 - \delta^{T^*+1})M - d(\delta) - f(z, l(\omega^*))\epsilon(\delta, \xi) - \phi^\xi M,$$

for any $\sigma \in NE(\Gamma^\infty(\delta, \mu))$. As $(1 - \delta^{T^*+1})M < \xi/3$ and $\phi^\xi M \leq \xi/3$ we find that

$$U_1(\omega^*, \sigma_2, \delta) \geq -d(\delta) - f(z, l(\omega^*)) \max\{(1 - \delta^{|\Theta^*|})/(\delta\underline{\pi})^{|\Theta^*|}, d(\delta), \phi^\xi\} - 2\xi/3,$$

for any $\sigma \in NE(\Gamma^\infty(\delta, \mu))$ and any $\delta \geq \delta^\xi$. Also, because $\phi^\xi f(z, l(\omega^*)) \leq \xi/3$ we find that

$$U_1(\omega^*, \sigma_2, \delta) \geq -d(\delta) - f(z, l(\omega^*)) \max\{(1 - \delta^{|\Theta^*|})/(\delta\underline{\pi})^{|\Theta^*|}, d(\delta)\} - \xi,$$

for any $\sigma \in NE(\Gamma^\infty(\delta, \mu))$ and any $\delta \geq \delta^\xi$. However, because ω^* is a feasible strategy for player one $U_1(\sigma, \delta) \geq U_1(\omega^*, \sigma_2, \delta)$. Consequently,

$$U_1(\sigma, \delta) \geq -d(\delta) - f(z, l(\omega^*)) \max\{(1 - \delta^{|\Theta^*|})/(\delta\underline{\pi})^{|\Theta^*|}, d(\delta)\} - \xi,$$

for any $\sigma \in NE(\Gamma^\infty(\delta, \mu))$ and any $\delta \geq \delta^\xi$. Hence,

$$U_1^{NE}(\delta, \mu) \geq -d(\delta) - f(z, l(\omega^*)) \max\{(1 - \delta^{|\Theta^*|})/(\delta\underline{\pi})^{|\Theta^*|}, d(\delta)\} - \xi,$$

for any $\delta \geq \delta^\xi$. Because ξ is arbitrary and because $\lim_{\delta \rightarrow 1} d(\delta) = d(\omega^*)$, we conclude that $U_1^{NE}(\mu) \geq -d(\omega^*) - f(z, l(\omega^*))d(\omega^*)$. \square

B.4. Proof of Theorem 1. For games that satisfy SA, 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 satisfy SA, i.e., Γ has LNCI or SCI but does not satisfy SA. However, because any game with SCI also satisfies SA by definition, our focus here is on games with LNCI.

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 with LNCI that does not satisfy SA 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.12) 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.4.1. 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

$$(16) \quad g_2(a_1^s, a_2) + P g_2(a_1^p, a_2') < -Ml(P + 1)$$

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 $\Gamma^{KJ}(\delta)$.³⁶ We partition Γ^{KJ} into blocks of length J , $\Gamma^{J,k}$, $k = 1, \dots, K$. Let u_i^k denote player i 's time average payoff in 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.10 *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. \(1996\)](#). \square

DEFINITION B.3 (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.9](#) which holds for *any* review type.

LEMMA B.11 (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.10](#). Part (i) follows immediately from Lemma [B.10](#). Proof of part (ii) is as follows: The fact that inequality [\(16\)](#)

³⁶This development closely follows [Celentani et al. \(1996\)](#), Lemma 4. Also, see the lemma's proof in that paper's appendix.

holds implies that there exists a $\delta^* < 1$ such that

$$(17) \quad \sum_{t=0}^{J(\epsilon)-1} \delta^t g_2(a_1^s, a_2) + \sum_{t=J(\epsilon)}^{J(\epsilon)+J(\epsilon)P-1} \delta^t g_2(a_1^p, a_2') < -lM J(\epsilon)(P+1)$$

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 $h^{t+J(\epsilon)-1} = \{a_1^0, y^0, a_1^1, y^1, \dots, a_1^{t+J(\epsilon)-1}, y^{t+J(\epsilon)-1}\}$, let $i(h^{t+J(\epsilon)-1}) = 1$, 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 (17). 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 $\delta > \delta_\epsilon$. \square

B.4.2. Reputation bound for review types. In the following we establish a reputation bound, Lemma B.12, 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.4 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.12 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.11 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.11 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. \square

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.11) for establishing the above bound for $\omega(n, T)$ is always the same for any n and any T .

B.4.3. 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, $T_1, T_2, \dots, T_n, \dots$ which we will pick recursively. We will also simultaneously pick a sequence of intervals of discount factors, $[\underline{\delta}_1, \bar{\delta}_1], [\underline{\delta}_2, \bar{\delta}_2], \dots$, such that $\lim_{i \rightarrow \infty} \underline{\delta}_i = 1$.

DEFINITION B.5 $\Gamma^N(\delta, \mu)$ 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.6 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.11. (i.e., the $\delta_{\epsilon/n}$ in Lemma B.11 that exists for the review type with shortfall at most ϵ/n). Notice that $\delta(n, T) \geq \delta_{\epsilon/n}$.

LEMMA B.13 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 $\lim_{k \rightarrow \infty} \xi_k = 0$, $\delta_k \in [\underline{\delta}, \bar{\delta}]$, $\mu_k \in \Delta_{\omega(n,T),z_k}$, $W_k \in W^{n,n}$ 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 $\mu \in \Delta_{\omega(n,T),z}$.³⁷ We have $\sigma \in NE(\Gamma^\infty(\mu, \delta))$ since $\lim_{k \rightarrow \infty} \xi_k = 0$ and all other terms converge to a limit.³⁸

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.12. \square

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.13. By Lemma B.13 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 5 *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 $\Gamma^\infty(\mu, \delta)$*

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

³⁷We use the Euclidean distance for ξ_k, δ_k , and **inherent product topology** for convergence of μ_k and the strategies σ_k .

³⁸This is standard see for instance Myerson (1991) page 144, Theorem 3.4 or Fudenberg and Levine (1986) Proposition 4.1.

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.13. 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 $U_1(\sigma, \delta) \geq -4\epsilon - \frac{\epsilon + \mu(\Omega \setminus W)}{z} f(z)$. \square

Now we choose T_n and the interval $[\underline{\delta}_n, \bar{\delta}_n]$ given $\{T_1, \dots, T_{n-1}\}$ and $\{[\underline{\delta}_1, \bar{\delta}_1], \dots, [\underline{\delta}_{n-1}, \bar{\delta}_{n-1}]\}$. 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 $D_{\epsilon/n}$ 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.14 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.13. We again obtain a contradiction to Lemma B.12. 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.11 hold and thus Lemma B.12 applies for $D_{\epsilon/n}$. \square

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.13 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.14 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.14, $\xi^{**} > 0$ be the cutoff $\xi([\underline{\delta}_n, \bar{\delta}_n])$ obtained in Lemma B.13, and let $\xi_n := \min\{\xi^*, \xi^{**}, \epsilon/n\} > 0$. First, Lemma B.13 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 $\sigma_N \in NE_{\xi_n}(\Gamma^N(\mu, \delta))$

$$(18) \quad U_1(\sigma_N, \delta) > -3\epsilon/n - f(z) \frac{\epsilon/n + \mu(\Omega \setminus W)}{z}.$$

Second, Lemma B.14 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 $\sigma_N \in NE_{\xi_n}(\Gamma^N(\mu, \delta))$

$$(19) \quad U_1(\sigma_N, \delta) > -3\epsilon/(n-1) - f(z) \frac{\epsilon/(n-1) + \mu(\Omega \setminus W)}{z}.$$

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.15 *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*

$$(20) \quad U_1(\sigma, \delta) > -4\epsilon/(n-1) - f(z) \frac{\epsilon/(n-1) + \mu(\Omega \setminus W)}{z}.$$

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 $\delta \leq \bar{\delta}_n$, 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 (18) and (19) 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 $\delta \leq \bar{\delta}_n$. Hence, $U_1(\sigma, \delta) \geq U_1(\sigma_N, \delta) - \epsilon/(n-1)$. \square

B.4.4. 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 $n' \geq n$, $\frac{4\epsilon}{n'-1} + f(z) \frac{\epsilon/(n'-1) + \mu(\Omega \setminus W)}{z} < \chi$. This follows from the fact that Ω is a countable set of finite automata. Also, by Lemma B.15, $U_1(\sigma, \delta) \geq -\frac{4\epsilon}{n-1} - f(z) \frac{\epsilon/(n-1) + \mu(\Omega \setminus W)}{z}$, for any $\delta \geq \underline{\delta}_n$, and any $\sigma \in NE(\Gamma^\infty(\mu, \delta))$. Therefore, we have the following inequalities:

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

Since χ is arbitrary, $\lim_{\delta \rightarrow 1} \inf_{\sigma \in NE(\Gamma^\infty(\mu, \delta))} U_1(\sigma, \delta) = U_1^{NE}(\mu) = 0$. \square

C. PROOFS OF THEOREM 2 AND LEMMA 3

We start by proving Theorem 2. The proof of Lemma 3 is similar and we indicate the changes in Theorem 2 that delivers lemma 3. We normalize the minimax payoff of player 1

to zero, i.e., $\hat{g}_1 = 0$. Moreover we normalize player one’s highest IR payoff, $\bar{g}_1 = 1$.³⁹ Given a finite automaton ω , recall that ω_θ denotes an automaton which is identical to ω except for its initial state. We will first *assume* that ω_θ is irreducible. We dispense this assumption in subsection C.4.

C.1. Notation. We denote the set of perfect public equilibrium (PPE) payoffs of $\Gamma^\infty(\delta)$ by $V(\delta)$. We define the minimum payoff of player 1 across all payoff profiles in $V(\delta)$ as

$$U_1^{low}(\delta) := \min\{u_1 \mid \exists u_2 \in \mathbb{R} \text{ s.t. } u = (u_1, u_2) \in V(\delta)\}.$$

Define $u_2^+(u_1, \delta) = \max\{u_2 : (u_1, u_2) \in V(\delta)\}$ and $u_2^-(u_1, \delta) = \min\{u_2 : (u_1, u_2) \in V(\delta)\}$ as the highest and lowest equilibrium payoffs of player two consistent with player one receiving a payoff equal to u_1 .⁴⁰ For any $\theta \in \Theta$, let $U_2(\omega_\theta, \delta) := \max_{\sigma_2 \in \Sigma_2} U_2(\omega_\theta, \sigma_2, \delta)$ denote player 2’s best payoff when her opponent plays the automaton strategy that starts at state $\theta \in \Theta$.

Let $\varepsilon > 0$ be the target upper bound on player one’s equilibrium payoff in our construction. We choose $\varepsilon \leq U_1^C(\omega)$ because otherwise there is nothing to prove. We pick $d > 0$ such that $d < \min_{\theta \in \Theta} \lim_{\delta \rightarrow 1} 1 - U_1^C(\omega_\theta, \delta)$.⁴¹ Therefore d is a lower bound on the shortfalls of all commitment types ω_θ for $\theta \in \Theta$.

Assumption NI and Fudenberg et al. (1994)’s Folk Theorem together imply that there exists a $v(\varepsilon, d) > 0$ and a $\delta^*(\varepsilon, d) < 1$ such that for every $\delta > \delta^*(\varepsilon, d)$, we have $u_2^+(u_1, \delta) - u_2^-(u_1, \delta) > v(\varepsilon, d)$ for every $u_1 \in [\varepsilon, 1 - d]$.

C.2. Review strategy profiles. We now describe a particular set of strategy profiles that we call review strategy profiles and denote by R . Subsets of this set are indexed by $\delta \in (0, 1)$ as well as the following set of parameters: $z \in [0, 1]$ which denotes the probability of the commitment type ω , $n \in \mathbb{N}$ which denotes the maximum number of review phases the strategy has, $e \in \mathbb{R}$ which denotes the cutoff level that defines “passing” and “failing” the test (described further below), and $r \in (0, 1)$ which is a measure of the length of the test phase. Hence $R(\delta, z, n, e, r)$ denotes a typical set of review strategy profiles. With every such set, there is an associated integer, $T(\delta)$, which is the largest integer such that $\delta^T \geq r$.⁴² The number T is the duration of each test phase of the strategy profile.

³⁹This normalization is without loss as long as $\bar{g}_1 > \hat{g}_1$ in Γ . We do not consider the uninteresting case where $\hat{g}_1 = \bar{g}_1$.

⁴⁰ The PPE payoff set $V(\delta)$ is compact, hence the max and min of the sets exist.

⁴¹This limit exists by Lemma A.3.

⁴² $T(\delta) := \max\{t \in \mathbb{N} : \delta^t \geq r\}$, where the number r is the last argument in the set R . To save on notation, we sometimes refer to $T(\delta)$ as simply T .

C.2.1. *The description of a review strategy profile with one review phase.* We first describe a review strategy profile $\sigma \in R(\delta, z, n = 1, e, r)$ on the equilibrium path. Then, we describe the off equilibrium path behavior of a review strategy profile σ . Throughout this description, we required that if player two's strategy σ_2 is part of a review strategy profile $\sigma = (\sigma_1, \sigma_2)$, then σ_2 is a best response to σ_1 at every history $h_2^t \in H_2$.

Equilibrium path play and phases. A review strategy profile has three phases. The first is the test phase which begins in period 0 and lasts through period $T - 1$. The second is the revelation phase and takes place in period T . The third is the execution phase and includes all periods after period T .

Test Phase: For $t \in \{0, 1, \dots, T-1\}$, $\sigma_1(h^t) = o(\theta^t)$, i.e., player one mimics the commitment strategy during the test phase. At the end of the test phase, we say that player two *failed the test* if $\frac{1-\delta}{1-\delta^T} \sum_{s=0}^{T-1} \delta^s r_1(a_1^s, y^s) > e$, i.e., if player one's payoff during the test phase exceeds the threshold e . Otherwise, we say that player two *passed the test*.

Revelation phase (or period): In period T , $\sigma_1(h^T)(\tilde{a}_1(\theta^T)) = \sigma_1(h^T)(o(\theta^T)) = 1/2$ where $\tilde{a}_1 : \Theta \rightarrow A_1$ is a function such that $\tilde{a}_1(\theta) \neq o(\theta)$ for every $\theta \in \Theta$ which we term a *revelation rule* for player one. In words, the normal type of player one, with equal probability, either reveals rationality by playing the revelation action $\tilde{a}_1(\theta^T)$ or alternatively mimics the commitment strategy by playing $o(\theta^T)$.

Execution phase: The play in the execution phase depends on the action taken by player one in the revelation period, i.e., a_1^T . If $a_1^T = o(\theta^T)$, then $\sigma_1|_{h^{T+1}}$ coincides with $\omega_{\theta^{T+1}}$ on the equilibrium path.⁴³ In words, if at the revelation phase player one has not revealed rationality, then player one will mimic the commitment type forever. Moreover, we require that $\sigma_2|_{h^{T+1}}$ satisfies $U_1(\omega_{\theta^{T+1}}, \sigma_2|_{h^{T+1}}) = U_1^C(\omega_{\theta^{T+1}}, \delta)$, i.e., $\sigma_2|_{h^{T+1}}$ is the best response to $\omega_{\theta^{T+1}}$ which minimizes player one's continuation payoff.

If $a_1^T = \tilde{a}_1(\theta^T)$, then we consider two alternatives. First, if player two has failed the test, then continuation play is given by a strategy profile $\sigma|_{h^{T+1}} = (\sigma_1|_{h^{T+1}}, \sigma_2|_{h^{T+1}})$ which sustains the payoff profile $(U_1^R(\theta^T, \delta), U_2(\theta^T, fail, \delta))$ in the complete information game $\Gamma^\infty(\delta)$, where $U_1^R(\theta^T, \delta)$ is chosen as the continuation payoff that makes player one indifferent at time T between playing $\tilde{a}_1(\theta_T)$ and playing $o(\theta_T)$, and $U_2(\theta^T, fail, \delta) = u_2^-(U_1^R(\theta^T, \delta), \delta)$. We further elaborate on $U_1^R(\theta^T, \delta)$ in the paragraph below. Second, if player two passed the test, then continuation play is given by a strategy profile $\sigma|_{h^{T+1}} = (\sigma_1|_{h^{T+1}}, \sigma_2|_{h^{T+1}})$ which sustains the payoff profile $(U_1^R(\theta^T, \delta), U_2(\theta^T, pass, \delta))$ in the complete information game $\Gamma^\infty(\delta)$ where $U_2(\theta^T, pass, \delta) = u_2^+(U_1^R(\theta^T, \delta), \delta)$.

We now calculate $U_1^R(\theta^T, \delta)$. Let $U_1^{NR}(\theta_i, \delta)$ denote player one's continuation payoff, at

⁴³ $\sigma_1|_{h^{T+1}}$ differs from $\omega_{\theta^{T+1}}$ at histories off the equilibrium path. We describe the strategy off the equilibrium path in the next subsection.

the beginning of the revelation phase, from not revealing rationality.⁴⁴ Hence:

$$U_1^{NR}(\theta_i, \delta) := (1 - \delta)g_1(o(\theta_i), \sigma_2(h^T)) + \delta E_{(\omega_{\theta_i}, \sigma_2)}\{U_1(\omega_{\theta'}, \delta)|\theta_i\}$$

where θ' is the state of the automaton at the beginning of period $T + 1$ under the strategy profile $(\omega_{\theta_i}, \sigma_2)$. Let $U_1^{REV}(\theta_i, \delta)$ denote player one's continuation payoff from revealing rationality at the revelation phase. Hence:

$$U_1^{REV}(\theta_i, \delta) := (1 - \delta)g_1(\tilde{a}_1(\theta_i), \sigma_2(h^T)) + \delta U_1^R(\theta_i, \delta)$$

where the payoff $U_1^R(\theta_i, \delta)$ is the continuation payoff player one gets starting from period $T + 1$ if he reveals rationality. The continuation payoff $U_1^R(\theta_i, \delta)$ is the payoff that makes player one indifferent at time T between revealing rationality by playing $\tilde{a}_1(\theta_i)$ and not revealing rationality by playing $o(\theta_i)$. More precisely,

$$U_1^R(\theta_i, \delta) = E_{(\omega_{\theta_i}, \sigma_2)}\{U_1(\omega_{\theta'}, \delta)|\theta_i\} + \frac{1 - \delta}{\delta}\{g_1(o(\theta_i), \sigma_2(h^T)) - g_1(\tilde{a}_1(\theta_i), \sigma_2(h^T))\}$$

This choice of continuation payoffs ensures that $U_1^{NR}(\theta_i, \delta) = U_1^{REV}(\theta_i, \delta)$. Hence, player one is indifferent between revealing rationality or not revealing rationality. Moreover, $U_1^R(\theta_i, \delta)$ is in the strict interior of $V(\delta)$ for δ sufficiently high and therefore $u_2^+(U_1^R(\theta_i, \delta), \delta)$ and $u_2^-(U_1^R(\theta_i, \delta), \delta)$ are well defined.

Off equilibrium play. Note that the full support assumption ensures that player two's deviations are not detectable, hence we only consider player one's deviations. Deviations by player one from the equilibrium path described above will trigger continuation play that gives player one a payoff of $U_1^{low}(\delta)$. Note a deviation by player one at or before time $t = T$ reveals his rationality and hence the continuation game is a game of complete information. If $a_1^T = o(\theta^T)$, then a deviation in period $t > T$ also reveals player one's rationality and again a strategy profile that yields a continuation payoff of $U_1^{low}(\delta)$ is played. If $a_1^T = \tilde{a}_1(\theta^T)$, then behavior after a deviation is already captured by the strategy profile of the complete information game.

Payoff bounds for review strategy profiles. The results of this section deliver bounds on the players' payoffs if they play a review strategy profile. These bounds are implied by the fact that player two's strategy is a best response to player one's strategy as well as the structure of the review strategy profile. Using these bounds, we verify that the review strategy profile is a PPE in the subsequent subsection.

The lemmata C.1 and C.2 together show that player one's discounted average payoffs

⁴⁴ This payoff function depends on σ . We suppress this dependence to keep the notation easier.

during the review phase are at most e with a probability at most e if $r > \frac{2M}{2M + \frac{1}{2}(1-z)ve}$ where $v := v(\varepsilon, d)$. Lemma C.3 uses Lemmata C.1 and C.2 to calculate an upper bound on a sufficiently patient player one's payoff from the review strategy profile. This bound is strictly lower than the commitment payoff.

The next lemma says that player two's payoffs differ across passing and failing the test during the review phase by at least $(1-z)v/2$.

LEMMA C.1 $\exists \delta^* < 1$ s.t. for $\delta > \delta^*$ and for all $\sigma(\delta) \in R(\delta, z, 1, e, r)$ we have:

$$\frac{1-z}{2} \left(\min_{\theta} U_2(\theta, pass, \delta) - \max_{\theta} U_2(\theta, fail, \delta) \right) + \frac{1+z}{2} \left(\min_{\theta} U_2(\omega_{\theta}, \delta) - \max_{\theta} U_2(\omega_{\theta}, \delta) \right) \geq \frac{(1-z)v}{2}$$

PROOF: The proof is in two steps. The first step shows that the second term of the left hand side of the inequality disappears when δ goes to one. The second step argues that the first term of the left hand side is bounded below by v as δ goes to one.

Step 1: $\lim_{\delta \rightarrow 1} |\min_{\theta} U_2(\omega_{\theta}, \delta) - \max_{\theta} U_2(\omega_{\theta}, \delta)| = 0$. This step follows from Lemma A.3.

Step 2: $\liminf_{\delta \rightarrow 1} (\min_{\theta} U_2(\theta, pass, \delta) - \max_{\theta} U_2(\theta, fail, \delta)) > v$. First note that any compact and convex subset of the set of strictly IR payoffs is contained in the set of equilibrium payoffs of $\Gamma^{\infty}(\delta)$ as $\delta \rightarrow 1$, by Fudenberg et al. (1994). By the full support assumption, we have $\lim_{\delta \rightarrow 1} \max_{(\theta_i, \theta_j) \in \Theta \times \Theta} (U_1^C(\omega_{\theta_i}, \delta) - U_1^C(\omega_{\theta_j}, \delta)) = 0$ and $\lim_{\delta \rightarrow 1} \sup_{\theta_i \in \Theta} U_1^C(\omega_{\theta_i}, \delta) \leq 1 - d$. Hence, using the definition of $U_1^R(\theta_i, \delta)$ implies that:

$$U_1^R(\theta_i, \delta) = E_{(\omega_{\theta_i}, \sigma_2)} \{U_1^C(\omega_{\theta'}, \delta) | \theta_i\} + \frac{1-\delta}{\delta} \{g_1(o(\theta_i), \sigma_2(h^T)) - g_1(\tilde{a}_1(\theta_i), \sigma_2(h^T))\}$$

we find that $\lim_{\delta \rightarrow 1} \max_{(\theta_i, \theta_j) \in \Theta \times \Theta} |U_1^R(\theta_i, \delta) - U_1^R(\theta_j, \delta)| = 0$ and that $\lim_{\delta \rightarrow 1} \sup_{\theta_i} U_1^R(\theta_i, \delta) \leq 1 - d$. These two inequalities and the compactness of the equilibrium payoff set imply that,

$$\begin{aligned} \lim_{\delta \rightarrow 1} \left(\min_{\theta} U_2(\theta, pass, \delta) - \max_{\theta} U_2(\theta, pass, \delta) \right) &= \lim_{\delta \rightarrow 1} \left(\min_{\theta} U_2(\theta, fail, \delta) - \max_{\theta} U_2(\theta, fail, \delta) \right) \\ &= 0. \end{aligned}$$

And since the set of equilibrium payoff set converges to the set of IR payoff set, we have $\liminf_{\delta \rightarrow 1} (\min_{\theta} U_2(\theta, pass, \delta) - \max_{\theta} U_2(\theta, fail, \delta)) > v$. \square

LEMMA C.2 If $r > \frac{2M}{2M + \frac{1}{2}(1-z)ve}$, then

$$\lim_{\delta \rightarrow 1} \sup_{\sigma(\delta) \in R(\delta, z, 1, e, r)} P_{\sigma(\delta)} \left\{ \left(\frac{1-\delta}{1-\delta^T} \right) \sum_{s=0}^{T(\delta)-1} \delta^s r_1(a_1^s, y^s) > e \right\} < e.$$

PROOF: Let $a_2^{nasty} : A_1 \rightarrow A_2$ be the mapping that satisfies $u_1(a_1, a_2^{nasty}(a_1)) \leq 0$. Such an action for player two exists, for every $a_1 \in A_1$ since player one's minmax is 0. Let $\sigma_2^{nasty}(\delta)$ be the strategy such that $\sigma_2^{nasty}(\delta)(h^t) = a_2^{nasty}(o(\theta^t))$ for any public history h^t .

First, we state the following identity without an argument. We prove this identity in section C.6 using a weak version of the law of large numbers.

$$(21) \quad \lim_{\delta \rightarrow 1} P_{\sigma_1(\delta), \sigma_2^{nasty}(\delta)} \left\{ \left(\frac{1-\delta}{1-\delta^T} \right) \sum_{s=0}^{T(\delta)-1} \delta^s r_1(a_1^s, y^s) > e \right\} = 0.$$

We'll argue on the way to a contradiction that, if

$$\limsup_{\delta \rightarrow 1} P_{\sigma(\delta)} \left\{ \left(\frac{1-\delta}{1-\delta^T} \right) \sum_{s=0}^{T(\delta)-1} \delta^s r_1(a_1^s, y^s) > e \right\} > e,$$

then a deviation from $\sigma_2(\delta)$ during the first $T(\delta)$ periods to $\sigma_2^{nasty}(\delta)$, and starting from $T(\delta)$ sticking with $\sigma_2(\delta)$ is a profitable deviation, provided δ is sufficiently high.

For a strategy profile σ , let

$$P_{\sigma(\delta)}(fail) := P_{\sigma(\delta)} \left\{ \left(\frac{1-\delta}{1-\delta^T} \right) \sum_{s=0}^{T(\delta)-1} \delta^s r_1(a_1^s, y^s) > e \right\}$$

and $P_{\sigma(\delta)}(pass) := 1 - P_{\sigma(\delta)}(fail)$.

Let $loss(\delta)$ of the strategy $\sigma_2^{nasty}(\delta)$ be the difference in player 2's expected discounted payoffs during the periods $t \in \{0, 1, \dots, T(\delta) - 1\}$ between employing the strategy $\sigma_2(\delta)$ and $\sigma_2^{nasty}(\delta)$ against the commitment type strategy. Note that, the difference in player 2's payoffs when she follows the strategy $\sigma_2(\delta)$ and $\sigma_2^{nasty}(\delta)$ is at most $2M$ at every period of the first T periods. So the loss, $loss(\delta)$ is at most $2M(1 - \delta^{T(\delta)})$. The gain of using the strategy $\sigma_2^{nasty}(\delta)$,

$gain(\delta)$, however will be at least:

$$\begin{aligned}
gain(\delta) &\geq P_{\sigma_1(\delta), \sigma_2^{nasty}(\delta)}(fail) \left(\frac{1-z}{2} \right) \delta^{T(\delta)+1} \min_{\theta} U_2(\theta, fail, \delta) \\
&\quad + P_{\sigma_1(\delta), \sigma_2^{nasty}(\delta)}(pass) \left(\frac{1-z}{2} \right) \delta^{T(\delta)+1} \min_{\theta} U_2(\theta, pass, \delta) \\
&\quad + \left(\frac{1+z}{2} \right) \delta^{T(\delta)+1} \left(\min_{\theta} U_2(\omega_{\theta}, \delta) - \max_{\theta} U_2(\omega_{\theta}, \delta) \right) \\
&\quad - P_{\sigma(\delta)}(fail) \left(\frac{1-z}{2} \right) \delta^{T(\delta)+1} \max_{\theta} U_2(\theta, fail, \delta) \\
&\quad - P_{\sigma(\delta)}(pass) \left(\frac{1-z}{2} \right) \delta^{T(\delta)+1} \max_{\theta} U_2(\theta, pass, \delta)
\end{aligned}$$

We will now show that $\liminf_{\delta \rightarrow 1} gain(\delta) - \limsup_{\delta \rightarrow 1} loss(\delta) > 0$. This will complete the proof since showing this amounts to showing that employing $\sigma_2^{nasty}(\delta)$ is a profitable deviation from $\sigma_2(\delta)$ which contradicts that $\sigma_2(\delta)$ is a best response to player 1's strategy.

We have

$$\lim_{\delta \rightarrow 1} \left(\min_{\theta} U_2(\omega_{\theta}, \delta) - \max_{\theta} U_2(\omega_{\theta}, \delta) \right) = 0$$

because of the full support and irreducibility assumptions, as shown before. Also,

$$\lim_{\delta \rightarrow 1} P_{\sigma_1(\delta), \sigma_2^{nasty}(\delta)}(pass) = 1,$$

i.e., the strategy σ_2^{nasty} passes the review test with a probability that approaches to one (this follows from equation 21). Moreover, by the definition of U_2 , full support, and irreducibility we have

$$\begin{aligned}
&\lim_{\delta \rightarrow 1} \left(\max_{\theta} U_2(\theta, pass, \delta) - \min_{\theta} U_2(\theta, pass, \delta) \right) \\
&= \lim_{\delta \rightarrow 1} \left(\max_{\theta} U_2(\theta, fail, \delta) - \min_{\theta} U_2(\theta, fail, \delta) \right) \\
&= 0.
\end{aligned}$$

By step 2 in the proof of Lemma C.1, we have that $\lim_{\delta \rightarrow 1} (\min_{\theta} U_2(\theta, pass, \delta) - \max_{\theta} U_2(\theta, fail, \delta)) \geq v$. Also, $\lim_{\delta \rightarrow 1} \delta^{T(\delta)+1} = r$, because $T(\delta)$ was chosen as the largest integer that satisfies $\delta^{T(\delta)} \geq r$.

Hence, if the supposition that $\limsup_{\delta \rightarrow 1} P_{\sigma(\delta)}(fail) > e$ is true, then $\liminf_{\delta \rightarrow 1} gain(\delta) \geq \frac{(1-z)rve}{2}$. On the other side, $\limsup_{\delta \rightarrow 1} loss(\delta) \leq 2M(1-r)$. But if $r > \frac{2M}{2M + \frac{1}{2}(1-z)ve}$, then $\limsup_{\delta \rightarrow 1} loss(\delta) < \liminf_{\delta \rightarrow 1} gain(\delta)$, completing the proof. \square

LEMMA C.3 *If $r > \frac{2M}{2M + \frac{1}{2}(1-z)ve}$, then $\limsup_{\delta \rightarrow 1, \sigma \in R(\delta, z, 1, e, r)} U_1(\sigma, \delta) \leq r(1-d) + (1-d)$*

$r)(e(2 - e))$.

PROOF: For $\sigma \in R(\delta, z, 1, e, r)$, using Lemma C.2 we obtain the following inequality:

$$\begin{aligned} U_1(\sigma, \delta) &\leq E_\sigma \left\{ (1 - \delta) \sum_{s=0}^{T(\delta)-1} \delta^s r_1(a_1^s, y^s) + (1 - \delta)\delta^{T(\delta)} 2M + \delta^{T(\delta)+1} U_1^R(\theta^{T+1}, \delta) \right\} \\ &\leq e(1 - \delta^{T(\delta)}) + (1 - e)(1 - \delta^{T(\delta)})e + (1 - \delta)\delta^{T(\delta)} 2M + \delta^{T(\delta)+1} \max_\theta U_1^R(\theta^{T(\delta)+1}, \delta) \end{aligned}$$

Hence, $\limsup_{\delta \rightarrow 1} U_1(\sigma, \delta) \leq (1 - r)(e(2 - e)) + r(1 - d)$. \square

Verification that $\sigma(\delta) \in R(\delta, z, 1, e, r)$ is an equilibrium.

Now we verify that any $\sigma(\delta) \in R(\delta, z, 1, e, r)$ is a PPE of $\Gamma^\infty(z, \delta)$ if $r > \frac{2M}{2M + \frac{1}{2}(1-z)ve}$ if the players are sufficiently patient.

LEMMA C.4 *If $r > \max\left\{\frac{2M}{2M+\varepsilon}, \frac{2M}{2M+\frac{1}{2}(1-z)ve}\right\}$, then there exists $\delta^* < 1$ such that for every $\delta > \delta^*$ any $\sigma \in R(\delta, z, 1, e, r)$ is a PPE of $\Gamma^\infty(z, \delta)$.*

PROOF: Let $O(\delta)$ represent terms for which $\limsup_{\delta \rightarrow 1} O(\delta) = 0$. Note that $\lim_{\delta \rightarrow 1} U_1^{low}(\delta) = 0$, because the folk theorem of Fudenberg et al. (1994) says that player one's lowest PPE payoff in $\Gamma^\infty(\delta)$ goes to his minimax. However, for any $\sigma \in R(\delta, z, 1, e, r)$ player one's payoff on the equilibrium path of σ is bounded below by $-2M(1 - r) + rU_1^C(\omega) + O(1 - \delta) \geq -2M(1 - r) + r\varepsilon + O(1 - \delta)$, which is strictly positive if $r > \frac{2M}{2M+\varepsilon}$ and if δ is sufficiently close to one. Hence deviations from σ_1 are deterred by the threat of $U_1^{low}(\delta)$ when δ is sufficiently large. Therefore, σ_1 is a best response to σ_2 at every history. Also, player two's strategy is by definition a best response to player one's strategy since $\sigma \in R(\delta, z, 1, e, r)$. \square

C.2.2. *The description of an n -step review strategy profile.* We describe $\sigma \in R(\delta, z, n, e, r)$ recursively, i.e., for $n > 1$, when stating the properties of the n step review strategy after the first revelation phase, we use the term $(n - 1)$ step review strategy profile when referring to the continuation strategy profile. Throughout this description, we required that if player two's strategy σ_2 is part of a review strategy profile $\sigma = (\sigma_1, \sigma_2)$, then σ_2 is a best response to σ_1 at every history $h_2^t \in H_2$.

Let $\sigma^{NR} \in R\left(\delta, \frac{z}{z+\frac{1-z}{2}}, n - 1, e, r\right)$ be an $n - 1$ step review strategy profile such that $U_1(\sigma^{NR}) \in [\varepsilon, 1 - d]$. If there is no such σ^{NR} , then let $R(\delta, z, n, e, r) := R\left(\delta, \frac{z}{z+\frac{1-z}{2}}, n - 1, e, r\right)$. Let $\tilde{a}_1 : \Theta \rightarrow A_i$ be a revelation rule with $\tilde{a}_1(\theta) \neq o(\theta)$ for every $\theta \in \Theta$.

Phases on the equilibrium path.

Test phase: For $t \in \{0, 1, \dots, T - 1\}$, $\sigma_1(h^t) = o(\theta^t)$, i.e., player 1 fully mimics the commitment strategy for the first T periods. Also, if $\left(\frac{1-\delta}{1-\delta^T}\right) \sum_{s=0}^{T-1} \delta^s r_1(a_1^s, y^s) > e$, then player two fails the test. Otherwise, player two passes the test.

Revelation phase: At $t = T$, $\sigma_1(h^T)(\tilde{a}_1(\theta^T)) = \sigma_1(h^T)(o(\theta^T)) = \frac{1}{2}$. In words, the normal type of player one reveals rationality by playing the revelation action $\tilde{a}_1(\theta^T)$, or mimics the commitment strategy with equal probabilities.

Execution phase: If $a_1^T = o(\theta^T)$, then $\sigma_1(h^t) = o(\theta^t)$ for $t \in \{T + 1, \dots, \tau - 1\}$ where τ is the first time greater than or equal to $T + 1$ for which $\theta^\tau = \theta_0$.⁴⁵ Moreover, $\sigma|_{h^\tau} = \sigma^{NR}$.

If $a_1^T = \tilde{a}_1(\theta^T)$, then there are two alternatives: First, if player two fails the test, then continuation play is given by $\sigma|_{h^{T+1}} = (\sigma_1|_{h^{T+1}}, \sigma_2|_{h^{T+1}})$ which is a strategy profile that sustains the payoff profile $u = (U_1^R(\theta^T, \delta), U_2(\theta^T, fail, \delta))$ in the complete information game $\Gamma^\infty(\delta)$, where

$$U_1^R(\theta^T, \delta) = \frac{1 - \delta}{\delta} \{g_1(o(\theta^T), \sigma_2(h^T)) - g_1(\tilde{a}_1(\theta^T), \sigma_2(h^T))\} \\ + E_{\sigma|_{h^T}} \left\{ \sum_{s=0}^{\tau-T-2} \delta^s r_1(a_1^{T+s+1}, y^{T+s+1}) + \delta^{\tau-T-1} U_1(\sigma^{NR}) | a_1^T = o(\theta^T) \right\}$$

and $U_2(\theta^T, fail, \delta) = u_2^-(U_1^R(\theta^T, \delta))$.⁴⁶

Second, if player two passes the test, then continuation play is given by $\sigma_{h^{T+1}} = (\sigma_1|_{h^{T+1}}, \sigma_2|_{h^{T+1}})$ which is a strategy profile that sustains the payoff profile $u = (U_1^R(\theta^T, \delta), U_2(\theta^T, pass, \delta))$ in the complete information game $\Gamma^\infty(\delta)$, where $U_2(\theta^T, pass, \delta) = u_2^+(U_1^R(\theta^T, \delta))$.

Off equilibrium play. Deviations by player one from the equilibrium path described above will trigger continuation play that gives player one a payoff of $U_1^{low}(\delta)$. Note that deviations by player one at or before time $t = T$, reveal his rationality, hence the continuation game is a game of complete information. Deviations at periods $t > T$ if $a_1^T = o(\theta^T)$ then reveal rationality, hence a strategy profile that yields a continuation payoff of $U_1^{low}(\delta)$ is played. If $a_1^T = \tilde{a}_1(\theta^T)$, then behavior after deviations are already captured by the strategy profile of the complete information game.

Payoff bounds for review strategy profiles. The results that follow deliver bounds on the players' payoffs if they play a review strategy profile, if the players are sufficiently patient, and if $R(\delta, z, n, e, r) \neq R\left(\delta, \frac{z}{z + \frac{1-z}{2}}, n - 1, e, r\right)$. These bounds are implied by the fact that player two's strategy is a best response to player one's strategy as well as the structure of the review strategy profile.

The following lemma shows that in the limit as players get arbitrarily patient, U_1^R and

⁴⁵We use the convention that if $\tau = T + 1$, then the set $\{T + 1, \dots, \tau - 1\} = \emptyset$.

⁴⁶Note that as in the $n = 1$ case, $U_1^R(\theta^T, \delta)$ is defined as the payoff that makes player one indifferent between revealing rationality and mimicking w , at the revelation period T . The payoff depends on the state of the automaton at period T because *i*) player two's action at T , and the revelation action depend on the state of the automaton at T *ii*) the payoffs until the state of the automaton reaches θ_0 depend on the state of the automaton at period T .

player two's continuation payoffs don't change with the state of the automaton at the revelation phase. The second lemma shows that player one's continuation payoff from mimicking the commitment type doesn't depend on the state of the automaton, in the limit as players are arbitrarily patient.

LEMMA C.5 *i)* $\lim_{\delta \rightarrow 1} \sup_{\theta_i, \theta_j \in \Theta} |U_1^R(\theta_i, \delta) - U_1^R(\theta_j, \delta)| = 0$, *ii)* $\lim_{\delta \rightarrow 1} \sup_{\theta_i, \theta_j \in \Theta} |U_2(\theta_i, \Delta, \delta) - U_2(\theta_j, \Delta, \delta)| = 0$ for $\Delta \in \{pass, fail\}$.

PROOF: *i)* Follows from Lemma A.3. *ii)* Follows from *i)* and the fact that $\lim_{\delta \rightarrow 1} V(\delta) = G$.
□

LEMMA C.6 *Let:*

$$U_i^{NR}(h^T, \delta) := E_{\sigma|h^T} \left\{ \sum_{s=0}^{\tau-T-2} \delta^s r_i(a_1^{T+s+1}, y^{T+s+1}) + \delta^{\tau-T-1} U_i(\sigma^{NR}, \delta) | a_1^T = o(\theta^T) \right\}$$

then, $\lim_{\delta \rightarrow 1} \sup_{h^T \in H^T} |U_i^{NR}(h^T, \delta) - U_i(\sigma^{NR}, \delta)| = 0$.

PROOF: The term $U_i^{NR}(h^T, \delta)$ depends on h^T only through θ^T , since the continuation strategy of player 1 coincides with the commitment strategy until the first time the state of the automaton becomes θ_0 . The expected time it takes to reach to state θ_0 is bounded by a number K (also see Lemma A.3), therefore as $\delta \rightarrow 1$, the payoffs until the state reaches θ_0 have an affect of at most $K(1 - \delta)M$ on the total discounted payoff. □

The following lemma shows that, as players get arbitrarily patient, player two gains a payoff of at least $\frac{(1-z)v}{2}$ if she passes the test instead of failing it. Note that the gain in the payoff is the multiplication of the probability that her opponent reveals rationality at the end of the test phase by v , where v is a lower bound on the difference on player two's continuation payoffs at histories followed by player one revealing his rationality.

LEMMA C.7 *There exists a cutoff $\delta^* < 1$ such that for all $\delta > \delta^*$ we have:*

$$\begin{aligned} \frac{1-z}{2} \left(\min_{\theta} U_2(\theta, pass, \delta) - \max_{\theta} U_2(\theta, fail, \delta) \right) + \frac{1+z}{2} \left(\min_{h^T} U_2^{NR}(h^T, \delta) - \max_{h^T} U_2^{NR}(h^T, \delta) \right) \\ \geq \left(\frac{1-z}{2} \right) v \end{aligned}$$

PROOF: For the second term on the LHS of the inequality, using Lemma C.6, we have that

$\lim_{\delta \rightarrow 1} \sup_{h^T \in H^T} |U_2^{NR}(h^T, \delta) - U_2(\sigma^{NR}, \delta)| = 0$, so

$$\lim_{\delta \rightarrow 1} \left| \left(\min_{h^T} U_2^{NR}(h^T, \delta) - \max_{h^T} U_2^{NR}(h^T, \delta) \right) \right| = 0.$$

For the first term on the LHS of the inequality, note that $\lim_{\delta \rightarrow 1} \sup_{\theta_i} |U_1^{NR}(\theta_i, \delta) - U_1(\sigma^{NR}, \delta)| = 0$ as shown in Lemma C.6. Since $\sigma^{NR} \in R\left(\delta, \frac{z}{z+\frac{1-z}{2}}, n-1, e, r\right)$ is chosen to satisfy $U_1(\sigma^{NR}) \in [\epsilon, 1-d]$, we have for δ sufficiently large that $u_2^+(U_1(\sigma^{NR}), \delta) - u_2^-(U_1(\sigma^{NR}), \delta) > v$ and therefore $\min_{\theta} U_2(\theta, pass, \delta) - \max_{\theta} U_2(\theta, fail, \delta) > v$ and the result follows. \square

LEMMA C.8 *Suppose that $r > \frac{2M}{2M+\frac{1}{2}(1-z)ve}$. Then there exists a cutoff $\delta^* < 1$ such that for any $\delta > \delta^*$ such that $R(\delta, z, n, e, r) \neq R\left(\delta, \frac{z}{z+\frac{1-z}{2}}, n-1, e, r\right)$ and any $\sigma \in R(\delta, z, n, e, r)$ we have*

$$P_{\sigma} \left\{ \left(\frac{1-\delta}{1-\delta^{T(\delta)}} \right) \sum_{s=0}^{T(\delta)-1} \delta^s u_1(a_1^s, y^s) > e \right\} < e + O(1-\delta).$$

PROOF: This lemma's proof is identical to the proof of Lemma C.2. \square

LEMMA C.9 *Suppose that $r > \max \frac{2M}{2M+\frac{1}{2}(1-z)ve}$, then we have either*

$$\limsup_{\delta \rightarrow 1} \left[\sup_{\sigma \in R(\delta, z, n, e, r)} U_1(\sigma, \delta) \right] \leq (1-r)(e(2-e)) + r \limsup_{\delta \rightarrow 1} \left[\sup_{\sigma \in R\left(\delta, \frac{z}{z+\frac{1-z}{2}}, n-1, e, r\right)} U_1(\sigma, \delta) \right]$$

or

$$\limsup_{\delta \rightarrow 1} \left[\sup_{\sigma \in R(\delta, z, n, e, r)} U_1(\sigma, \delta) \right] = \limsup_{\delta \rightarrow 1} \left[\sup_{\sigma \in R\left(\delta, \frac{z}{z+\frac{1-z}{2}}, n-1, e, r\right)} U_1(\sigma, \delta) \right].$$

PROOF: Fix $\delta < 1$ and $z > 0$. If $R(\delta, z, n, e, r) = R\left(\delta, \frac{z}{z+\frac{1-z}{2}}, n-1, e, r\right)$, then

$$\sup_{\sigma \in R(\delta, z, n, e, r)} U_1(\sigma, \delta) = \sup_{\sigma \in R\left(\delta, \frac{z}{z+\frac{1-z}{2}}, n-1, e, r\right)} U_1(\sigma, \delta).$$

Suppose $R(\delta, z, n, e, r) \neq R\left(\delta, \frac{z}{z+\frac{1-z}{2}}, n-1, e, r\right)$. For any $\sigma^* \in R(\delta, z, n, e, r)$, conditional on player two passing the test, player one receives a discounted payoff of at most $e(1-\delta^T)$ during the first T periods and conditional on player two failing the test at most $1(1-\delta^T)$. Also, $U_1(\sigma^{NR}, \delta) \in [\epsilon, 1-d]$ in the profile σ^* . Consequently, writing the upperbound of player

one's payoff for σ^* gives us the following inequalities:

$$\begin{aligned}
U_1(\sigma^*, \delta) &\leq (1 - \delta^T) (eP_{\sigma^*}(\text{pass}) + P_{\sigma^*}(\text{fail})) \\
&\quad + \delta^T \max_{h^T} E_{\sigma^*|_{h^T}} \left\{ \sum_{s=0}^{\tau-T-2} \delta^s r_1(a_1^{T+s+1}, y^{T+s+1}) + \delta^{\tau-T-1} U_1(\sigma^{NR}, \delta) \mid a_1^T = o(\theta^T) \right\} \\
&\leq (1 - r) (e(2 - e)) + r \sup_{\sigma \in R\left(\delta, \frac{z}{z + \frac{1-z}{2}}, n-1, e, r\right)} U_1(\sigma, \delta) + O(1 - \delta).
\end{aligned}$$

Therefore, we have either the following inequality:

$$\sup_{\sigma \in R(\delta, z, n, e, r)} U_1(\sigma, \delta) \leq (1 - r) (e(2 - e)) + r \sup_{\sigma \in R\left(\delta, \frac{z}{z + \frac{1-z}{2}}, n-1, e, r\right)} U_1(\sigma, \delta) + O(1 - \delta)$$

or $\sup_{\sigma \in R(\delta, z, n, e, r)} U_1(\sigma, \delta) = \sup_{\sigma \in R\left(\delta, \frac{z}{z + \frac{1-z}{2}}, n-1, e, r\right)} U_1(\sigma, \delta)$. Hence, the result follows by taking the limits. \square

C.3. Completing the proof of Theorem 2.

PROOF: We will show the following: For every $\epsilon > 0$, there exists a $z^* \in (0, 1)$ such that for every $z \in (0, z^*)$ there exists a cutoff $\delta^* < 1$ such that for every $\delta > \delta^*$ there exists a PPE σ of $\Gamma^\infty(z, \delta)$ for which $U_1(\sigma, \delta) \leq \epsilon$.

Note that $\lim_{\delta \rightarrow 1} U_1^{\text{low}}(\delta) = 0$, by [Fudenberg et al. \(1994\)](#)'s folk theorem. We'll show that on the equilibrium path player one's continuation payoffs are always at least $\frac{\epsilon}{2}$ for all discount factors above a critical level, and hence in this case he has no profitable deviations. To show that his continuation payoffs are at least $\frac{\epsilon}{2}$, we put a lower bound on his payoffs at the start of the review phases, and this lower bound constitutes a lower bound on his continuation payoffs at every period during the test phase. The lower bound is found by applying Lemma [C.9](#) successively starting from review strategy profiles with one phase, i.e., $n = 1$.

Let $r(\epsilon) = \max\left\{\frac{2M}{2M + \frac{1}{2}v\frac{\epsilon}{2}}, \frac{2M + \frac{\epsilon}{2}}{2M + \epsilon}\right\}$. Define the sequence of bounds $\{b_j\}_{j \in \{0, 1, \dots\}}$ recursively as $b_0 = 1 - d$ and $b_j = (1 - r(\epsilon))\frac{\epsilon}{2}(2 - \frac{\epsilon}{2}) + r(\epsilon)b_{j-1}$ for $j \geq 1$. Since $r(\epsilon) \in (0, 1)$, and $\frac{\epsilon}{2}(2 - \frac{\epsilon}{2}) < \epsilon$, there exists $n(\epsilon)$ such that $b_{n(\epsilon)} < \epsilon$ and $b_{n(\epsilon)-1} \geq \epsilon$. Now define a sequence of reputation levels $\{z_j\}_{j \in \{0, 1, \dots\}}$ as $z_0 = 1/2$, and z_i is such that $z_{i-1} = \frac{z_i}{z_i + \frac{1-z_i}{2}}$ for $i \geq 1$. We set $z^* = z_{n(\epsilon)}$.

We claim that there exists $\delta^* < 1$ such that for every $\delta > \delta^*$, any $\sigma \in R(\delta, z, n(\epsilon), \frac{\epsilon}{2}, r(\epsilon))$ is a PPE of $\Gamma^\infty(z, \delta)$ and $U_1(\sigma, \delta) \leq \epsilon$. The fact that $U_1(\sigma, \delta) \leq \epsilon$ follows from Lemma [C.9](#) applied successively, and after checking that $r(\epsilon) \geq \frac{2M}{2M + \frac{1}{2}z v \frac{\epsilon}{2}}$ for every $z \leq 1/2$, and noting that there exists $\delta^{**} < 1$ such that for $\delta > \delta^{**}$, for any $j < n(\epsilon)$, and for any $\sigma' \in R(\delta, z, j, \frac{\epsilon}{2}, r(\epsilon))$, we have $U_1(\sigma', \delta) \in (\frac{\epsilon}{2}, 1 - d)$. Player one has no profitable deviations, because on the equilibrium path player one receives a payoff of $(1 - r)\frac{\epsilon}{2}(2 - \frac{\epsilon}{2}) + r\epsilon + O(1 - \delta)$,

and this payoff is larger than $U_1^{low}(\delta)$ for sufficiently large δ . Note that player one's deviations always lead to a continuation game of complete information, hence the strategy profile that gives player one the payoff $U_1^{low}(\delta)$ may be chosen as an equilibrium of the continuation game. Player two's strategy is by construction chosen to be a best response to player one's strategy. Hence there are no profitable deviations for either player. \square

C.4. Dispensing with irreducibility. Suppose now that the finite automata is not irreducible and may have $k > 1$ recurrent classes. Notice each of these recurrent classes uniquely identifies a new irreducible finite automaton. The above construction can be made for the irreducible automata identified by each of these recurrent classes. Let $z^* > 0$ be the minimum of the reputation levels of these k automata needed for Theorem 2 to hold. Consider now the repeated game where $\mu(\omega) < z^*$ and the strategy profiles where player one plays the action $o(\theta^t)$ until the automata enters a recurrent class and from then on the continuation strategy profile is the review strategy profile constructed using our construction for that recurrent class. Since player one fully mimics the commitment type until the automaton enters a recurrent class, his reputation level does not change until the automaton enters a recurrent class. In this strategy profile, each continuation strategy profile once player one has entered in a recurrent class gives player one a payoff of at most ϵ . In this strategy profile, the payoff of player one until the first time that the automata enters a recurrent class goes to zero as δ goes to one, and the payoff of player one in the continuation game that starts from any state of the recurrent classes is at most ϵ . Hence this strategy profile gives player one a payoff not more than 2ϵ when the players are sufficiently patient. Since ϵ is arbitrary, Theorem 2 obtains.

C.5. Proof of Lemma 3. If there is a positive gap, $\gamma > 0$, and if NI is satisfied, then for every $\epsilon > 0$, there exists a $v(\epsilon) > 0$ and a $\delta^*(\epsilon) < 1$ such that for every $\delta > \delta^*(\epsilon)$, we have $u_2^+(u_1, \delta) - u_2^-(u_1, \delta) > v(\epsilon)$ for every $u_1 \in [\epsilon, 1]$.

Moreover, if Γ does not satisfy SA, then for every action $a_1 \in A_1$ of player one, there exists an action profile $a'(a_1) = (a'_1(a_1), a_2(a_1)) \in A_1 \times A_2$ such that $g_1(a'(a_1)) > g_1(a_1, a_2(a_1))$. In the description of the review strategy profile, the revelation rule is $\tilde{a}_1(\theta) = a'_1(o(\theta))$. To give the incentives to player two to play $a'_2(o(\theta))$ at the revelation periods is possible by choosing the continuation payoffs of player two after the revelation phase if player one reveals his rationality to depend on the signal $y \in Y$. Continuation payoffs need vary only by some number $C(1 - \delta)$ where C is a constant that is independent of δ to make it optimal for player two to play $a'_1(o(\theta))$. The rest of the proof is identical to the proof of Theorem 2.

C.6. Proof of the identity given by Equation 21. Recall that $T(\delta)$ is the largest integer for which $\delta^{T(\delta)} \geq r$.⁴⁷ Let $y_t = r_1(a_1^t, y^t) - E\{r_1(a_1^t, y^t)|h^t\}$, where the expectation is taken with respect to the measure $P_{\sigma_1(\delta), \sigma_2^{nasty}(\delta)}$. We remind that $E\{r_1(a_1^t, y^t)|h^t\} \leq 0$ by the definition of $\sigma_2^{nasty}(\delta)$. Let δ_T be such that $\delta_T^T = r$. Let $x_{t,T} = (1 - \delta_T)\delta_T^t y_t$ and notice that $x_{t,T}$ is a martingale with respect to the measure $P_{\sigma_1(\delta), \sigma_2^{nasty}(\delta)}$.

We will apply the LLN in Davidson (1994), Theorem 19.7. We now check that the three assumptions of the Theorem are satisfied by picking $c_{t,T} = (1 - \delta_T)\delta_T^t$: First, $\frac{x_{t,T}}{c_{t,T}} = y_t$ is uniformly integrable since the function r_1 is bounded. Second, $\limsup_{T \rightarrow \infty} \sum_{t=0}^T c_{t,T} = \limsup_{T \rightarrow \infty} 1 - \delta_T^{T+1} = 1 - r < \infty$. Third, $\lim_{T \rightarrow \infty} \sum_{t=0}^T c_{t,T}^2 = \lim_{T \rightarrow \infty} (1 - \delta_T)^2 \left(\frac{1 - \delta_T^{2(T+1)}}{1 - \delta_T^2} \right) = \lim_{T \rightarrow \infty} (1 - \delta_T) \left(\frac{1 - r^2}{2} \right) = 0$. Hence, by Davidson (1994), Theorem 19.7, $\sum_{t=0}^T x_{t,T} \rightarrow_{L^p} 0$.

Recall that $T(\delta)$ is the largest integer for which $\delta^{T(\delta)} \geq r$.

$$\sum_{s=0}^{T(\delta)-1} (1 - \delta)\delta^s y_s = \sum_{s=0}^{T(\delta)-1} x_s + \sum_{s=0}^{T(\delta)-1} ((1 - \delta)\delta^s - (1 - \delta_{T(\delta)})\delta_{T(\delta)}^s) y_s$$

Since $\lim_{\delta \rightarrow 1} \sum_{s=0}^{T(\delta)-1} \left| (1 - \delta)\delta^s - (1 - \delta_{T(\delta)})\delta_{T(\delta)}^s \right| = 0$ and y_s is bounded, we have $\sum_{s=0}^{T(\delta)-1} (1 - \delta)\delta^s y_s \rightarrow_{L^p} 0$ as δ goes to one. Hence for any $a > 0$,

$$\lim_{\delta \rightarrow 1} P_{\sigma_1(\delta), \sigma_2^{nasty}(\delta)} \left\{ (1 - \delta) \sum_{s=0}^{T(\delta)-1} \delta^s y_s > a \right\} = 0.$$

Since $y_t = r_1(a_1^t, y^t) - E\{r_1(a_1^t, y^t)|h^t\}$ and $E\{r_1(a_1^t, y^t)|h^t\} \leq 0$, we have $r_1(a_1^t, y^t) \leq y_t$. Moreover, $\lim_{\delta \rightarrow 1} 1 - \delta^{T(\delta)} = 1 - r$. Since a is arbitrary, for any $e > 0$ the following inequality holds:

$$\lim_{\delta \rightarrow 1} P_{\sigma_1(\delta), \sigma_2^{nasty}(\delta)} \left\{ \left(\frac{1 - \delta}{1 - \delta^{T(\delta)}} \right) \sum_{s=0}^{T(\delta)-1} \delta^s r_1(a_1^s, y^s) > e \right\} = 0.$$

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

⁴⁷ We thank Xiaohong Chen for directing us to Davidson (1994), Theorem 19.7 that is used in this proof.

- 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.
- DAVIDSON, J. (1994): *Stochastic Limit Theory*, Oxford Univ. Press.
- 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.
- FUDENBERG, D., D. LEVINE, AND E. MASKIN (1994): "The folk theorem with imperfect public information," *Econometrica*, 997–1039.
- 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.