CMS-EMS
Center for Mathematical Studies in Economics
And Management Science

Discussion Paper #1506

REPUTATION WITH LONG RUN PLAYERS AND IMPERFECT OBSERVATION

ALP E. ATAKAN
Northwestern University

MEHMET EKMEKCI
Northwestern University

August 2009

JEL Classification: C73, D83.

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

NORTHWESTERN UNIVERSITY

REPUTATION WITH LONG RUN PLAYERS AND IMPERFECT OBSERVATION

ALP E. ATAKAN AND MEHMET EKMEKCI

Abstract.

Previous work shows that reputation results may fail in repeated games between two long-run players with equal discount factors. We restrict attention to an infinitely repeated game where two players with equal discount factors play a simultaneous move stage game where actions of player 2 are imperfectly observed. The set of commitment types for player 1 is taken as any (countable) set of finite automata. In this context, for a restricted class of stage games, we provide a one sided reputation result. If player 1 is a particular commitment type with positive probability and player 2's actions are imperfectly observed, then player 1 receives his highest payoff, compatible with individual rationality, in any Bayes-Nash equilibria, as agents become patient.

Keywords: Repeated Games, Reputation, Equal Discount Factor, Long-run Players, Imperfect Observation, Complicated Types, Finite Automaton
JEL Classification Numbers: C73, D83.

1. Introduction and Related Literature

We consider an infinitely repeated game where two equally patient agents play a simultaneous move stage game. Player 1's stage game actions are perfectly observed by player 2 (she) while player 2's stage game actions are imperfectly observed by player 1 (he). We present two reputation results. For our first reputation result we restrict attention to stage games where player 1 has an action (a Stackelberg action) such that any best response to this action gives player 1 his highest payoff compatible with the individual rationality of player 2. Player 1's type is private information and he can be one of many commitment types. Each commitment type is committed

Date: First draft July, 2008. This draft, August, 2009.

We would like to thank Martin Cripps, Eddie Dekel, Christoph Kuzmics and Larry Samuelson for useful discussions.

to playing a certain repeated game strategy and is a finite automaton. We further assume that there is a commitment type that plays the Stackelberg action in every period of the repeated game (the Stackelberg type). We show that a patient player 1 can guarantee his highest payoff, that is consistent with the individual rationality of player 2, in any Bayes-Nash equilibrium of the repeated game. In other words, player 1 guarantees, in any equilibrium, the payoff that he can secure by publicly pre-committing to a repeated game strategy that plays the Stackelberg action in each period.

Our second reputation result covers an expanded set of stage games. In addition to those covered by our first reputation result, we allow for any stage game in which player 2 receives a payoff that strictly exceeds her minimax value in the profile where player 1 receives his highest payoff (i.e., any *locally non-conflicting interest* game). For this result, we construct a novel repeated game strategy for player 1 (an infinitely accurate review strategy) such that any best response to this strategy gives player 1 a payoff arbitrarily close to his highest payoff if he is sufficiently patient.[1] We assume that there is a commitment type (a review type) that plays such an infinitely accurate review strategy and all the other commitment types are finite automata. We show that player 1 guarantees his highest possible payoff, in any Bayes-Nash equilibrium of the repeated game, if he is sufficiently patient. As in our first reputation result, player 1 guarantees in any equilibrium, the payoff that he can secure by publicly pre-committing to his most preferred repeated game strategy. In contrast, however, to our first reputation result, the commitment type that player 1 mimics to secure a high payoff, i.e., the review type, plays a complex repeated game strategy with an infinite number of states. In particular, the review type plays a repeated game strategy that is significantly more complicated than the other commitment types that we allow for.

A reputation result was first established for infinitely repeated games by Fudenberg and Levine (1989, 1992). They showed that if a patient player 1 plays a stage game against a myopic opponent and if there is positive probability that player 1 is a type committed to playing the Stackelberg action in every period, then in any equilibrium

[1]Our development of review strategies builds on previous work by Celantani et al. (1996). For another reference on review strategies see Radner (1981, 1985).

of the repeated game player 1 gets at least his static Stackelberg payoff.[2] However, in a game with a non-myopic opponent, player 1 may achieve a payoff that exceeds his static Stackelberg payoff by using a dynamic strategy that rewards or punishes player 2 (see Celantani et al. (1996)). Conversely, fear of future punishment or expectation of future rewards can induce player 2 to not best respond to a Stackelberg action and thereby force player 1 below his static Stackelberg payoff. The non-myopic player 2 may fear punishment either from another commitment type (Schmidt (1993) or Celantani et al. (1996)) or from player 1's normal type following the revelation of rationality (Celantani et al. (1996) section 5 or Cripps and Thomas (1997)).[3] Nevertheless, reputation results have also been established for repeated games where player 1 faces a non-myopic opponent, but one who is sufficiently less patient than player 1, by applying the techniques of Fudenberg and Levine (1989, 1992) (Schmidt (1993), Celantani et al. (1996), Aoyagi (1996), or Evans and Thomas (1997)).

Reputation results are fragile in repeated games in which a simultaneous-move stage game is played by equally patient agents and actions are perfectly observed. In particular, one-sided reputation results obtain only if the stage game is a game of *strictly conflicting interest* (Cripps et al. (2005)), or the Stackelberg action is a dominant action in the stage game (Chan (2000)).[4] For other simultaneous-move games, Cripps and Thomas (1997) show that any individually rational and feasible payoff can be sustained in perfect equilibria of the infinitely repeated game, if the players are sufficiently patient.

This paper also focuses on simultaneous-move stage games however our results sharply contrast with previous literature. We prove, if player 2's actions are imperfectly observed with full support (the *full support* assumption), then player 1 can guarantee a high payoff, whereas, with perfect observability, Cripps and Thomas (1997) and Chan (2000) demonstrated a folk theorem for a subset of the class of stage-games that we consider. In particular, we show that imposing the full support

[2]The static Stackelberg payoff for player 1 is the highest payoff he can guarantee in the stage game through public pre-commitment to a stage game action (a Stackelberg action). See Fudenberg and Levine (1989) or Mailath and Samuelson (2006), page 465, for a formal definition.

[3]Player 1 reveals rationality if he chooses a move that would not be chosen by any commitment type.

[4]A game has strictly conflicting interests (Chan (2000)) if a best reply to the Stackelberg action of player 1 yields the best feasible and individually rational payoff for player 1 and the minimax for player 2.

assumption allows us to expand the set of stage games covered by a reputation result from only strictly conflicting interest games and dominant action games to also include any locally non-conflicting interest game. Also, the result presented here improves upon the previous reputation result for strictly conflicting interest games (Cripps et al. (2005)) since our result allows for a rich set of commitment types.

The role of imperfectly observed actions in our reputation result is analogous to the role that imperfectly observed actions play in Celentani et al. (1996) or the role that trembles play in Aoyagi (1996). It ensures that a wide range of information sets in the repeated game are sampled. In particular, the information sets that are crucial for player 1 to build a reputation, for being the commitment type that he is trying to mimic, are sampled sufficiently often, irrespective of the strategy player 2 plays. Also, imperfectly observed actions ensures that, irrespective of the strategy player 2 plays, if player 1 is mimicking a particular commitment type, then player 2 will, with arbitrary precision, learn that player 1 is either this commitment type or a normal type.

This paper is closely related to Atakan and Ekmekci (2008) which proves a one-sided reputation result, for perfect Bayesian equilibria, under the assumption that the stage-game is a extensive form game of perfect information (i.e., all information sets are singletons). This paper shows that one can dispense with the perfect information assumption and can allow for Bayes-Nash equilibria, if player 2's actions are imperfectly observed. Also, the imperfect observation of player 2's actions enables us to take any countable set of finite automata as the set of possible commitment types whereas the set of commitment types in Atakan and Ekmekci (2008) is more restricted. The two papers are contrasted in detail in section 5.

The paper proceeds as follows: section 2 describes the repeated game; section 3 and 4 present our two main reputation results; and section 5 compares this paper with Atakan and Ekmekci (2008).

2. The Model

We consider a repeated game $\Gamma^{\infty}(\delta)$ in which a simultaneous move stage game Γ is played by players 1 and 2 in periods $t \in \{0, 1, 2, ...\}$ and the players discount payoffs using a common discount factor $\delta \in [0, 1)$.

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 2's stage game action is not observed by player 1 while player 1's action is perfectly observed by player 2. Y is the set of publicly observed outcomes of player 2's actions. After each period player 1 observes an element of Y which depends on the action player 2 played in the period. Each action profile $a_2 \in A_2$ induces a probability distribution over the publicly observed outcomes Y. Let $\pi_y(a_2)$ denote the probability of outcome $y \in Y$ if action $a_2 \in A_2$ is used by player 2 and for any $\alpha_2 \in \Delta(A_i)$ let $\pi_y(\alpha_2) = \sum_{a_2 \in A_2} \alpha_2(a_2)\pi_y(a_2)$.

Assumption 1 (Full support)**.** *For any a_2 and $a_2' \in A_2$, $supp(\pi(a_2)) = supp(\pi(a_2'))$.*

Assumption 1 implies that player 1 is never exactly sure about player 2's action. The assumption, however, does not put any limits on the degree of imperfect observability. In fact, player 1's information can be arbitrarily close to perfect information.

In the repeated game Γ^∞ players have perfect recall and can observe past outcomes. $H^t = A_1^t \times Y^t$ is the set of period $t \geq 0$ public histories and $\{a_1^0, y^0, a_1^1, y^1, ..., a_1^{t-1}, y^{t-1}\}$ is a typical element. $H_2^t = A_1^t \times A_2^t \times Y^t$ is the set of period $t \geq 0$ private histories for player 2 and $\{a_1^0, a_2^0, y^0, ..., a_1^{t-1}, a_2^{t-1}, y^{t-1}\}$ is a typical element. For player 1 $H_1^t = H^t$.

Types and Strategies. Before time 0 nature selects player 1 as a type ω from a countable set of types Ω according to common-knowledge prior μ. Player 2 is known with certainty to be a normal type that maximizes expected discounted utility. Ω contains a normal type for player 1 that we denote N. We let $\Delta(\Omega)$ denote the set of probability measures over Ω and $interior(\Delta(\Omega)) = \{\mu \in \Delta(\Omega) : \mu(\omega) > 0, \ \forall \omega \in \Omega\}$. Player 2's belief over player 1's types, $\mu : \bigcup_{t=0}^{\infty} H^t \to \Delta(\Omega)$, is a probability measure over Ω after each period t public history.

A behavior strategy for player i is a function $\sigma_i : \bigcup_{t=0}^{\infty} H_i^t \to \mathcal{A}_i$ and Σ_i is the set of all behavior strategies. A behavior strategy chooses a mixed stage game action given any period t public history. Each type $\omega \in \Omega \setminus \{N\}$ is committed to playing a particular repeated game behavior strategy $\sigma_1(\omega)$. A strategy profile $\sigma = (\{\sigma_1(\omega)\}_{\omega \in \Omega}, \sigma_2)$ lists the behavior strategies of all the types of player 1 and player 2. For any period t public history h^t and $\sigma_i \in \Sigma_i$, $\sigma_i|_{h^t}$ is the continuation strategy induced by h^t. For $\sigma_1 \in \Sigma_1$ and $\sigma_2 \in \Sigma_2$, $\Pr_{(\sigma_1,\sigma_2)}$ is the probability measure over the set of (infinite) public histories induced by (σ_1, σ_2).

Payoffs. Stage game payoffs for any player i, $r_i : A_1 \times Y \to \mathbb{R}$, depend only on publicly observed outcomes a_1 and y. A player's repeated game payoff is the normalized discounted sum of the stage game payoffs. For any infinite public history h, $u_i(h,\delta) = (1-\delta)\sum_{k=0}^{\infty}\delta^k r_i(a_1^k, y^k)$, and $u_i(h^{-t},\delta) = (1-\delta)\sum_{k=t}^{\infty}\delta^{k-t} r_i(a_1^k, y^k)$ where $h^{-t} = \{a_1^t, y^t, a_1^{t+1}, y^{t+1}, ...\}$. Player 1 and player 2's expected continuation payoff, following a period t public history, under strategy profile σ, $U_1(\sigma,\delta|h^t) = U_1(\sigma_1(N), \sigma_2, \delta|h^t)$ and

$$U_2(\sigma,\delta|h^t) = \sum_{\omega\in\Omega}\mu(\omega|h^t)U_2(\sigma_1(\omega),\sigma_2,\delta|h^t),$$

where $U_i(\sigma_1(\omega),\sigma_2,\delta|h^t) = \mathbb{E}_{(\sigma_1(\omega),\sigma_2)}[u_i(h^{-t},\delta)|h^t]$ is the expectation over continuation histories h^{-t} with respect to $\Pr_{(\sigma_1(\omega)|_{h^t},\sigma_2|_{h^t})}$. Let $U_i(\sigma,\delta) = U_i(\sigma,\delta|h^0)$. Also, let

$$U_i(\sigma,\delta|h^t, a_1, \alpha_2) = \sum_{y\in Y}\pi_y(\alpha_2)U_i(\sigma,\delta|h^t, a_1, y).$$

The Stage Game. Let

$$g_i(a_1,a_2) = \sum_{y\in Y} r_i(a_1,y)\pi_y(a_2).$$

The stage game Γ with action sets A_i and payoff function $g_i : A_1 \times A_2 \to \mathbb{R}$ is a standard normal form game. The mixed minimax payoff for player i, $\hat{g}_i = \min_{\alpha_j}\max_{\alpha_i} g_i(\alpha_i,\alpha_j)$ and the pure minimax payoff for player i, $\hat{g}_i^p = \min_{a_j}\max_{a_i} g_i(a_i,a_j)$. Define $a_1^p \in A_1$ 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 $F = co\{g_1(a_1,a_2), g_2(a_1,a_2) : (a_1,a_2) \in A_1\times A_2\}$; and the set of feasible and individually rational payoffs $G = F \cap \{(g_1,g_2) : g_1 \geq \hat{g}_1, g_2 \geq \hat{g}_2\}$. Let $\bar{g}_1 = \max\{g_1 : (g_1,g_2)\in G\}$, and $M = \max\{\max\{|g_1|,|g_2|\} : (g_1,g_2)\in F\}$.

Assumption 2. *There exists $a_1^s \in A_1$ such that any best response to a_1^s yields payoffs $(\bar{g}_1, g_2(a_1^s, a_2^b))$, where $a_2^b \in A_2$ is a best response to a_1^s. Also, $g_2 = g_2(a_1^s, a_2^b)$ for all $(\bar{g}_1, g_2) \in G$.*

Assumption 3 (Locally non-conflicting interest stage game)**.** *For any $g \in G$ and $g' \in G$, if $g_1 = g_1' = \bar{g}_1$, then $g_2 = g_2' > \hat{g}_2^p$.*

Both Assumption 2 and Assumption 3 require that the payoff profile where player 1 receives $\bar{g}_1$, i.e., his highest payoff compatible with the individual rationality of player 2, is unique. This requirement is satisfied generically. Assumption 2 further requires that there exists a pure stage game action for player 1 (a_1^s) such that any best response to this action yields player 1 a payoff equal to $\bar{g}_1$. Assumption 3 requires that player 2 receives a payoff strictly higher than her pure strategy minimax payoff in the payoff profile where player 1 receives a payoff equal to $\bar{g}_1$.[5] A game Γ satisfies Assumption 2, but not Assumption 3, only if $g_2(a_1^s, a_2^b) = \hat{g}_2$, that is, if Γ is a strictly conflicting interest game.

If Γ satisfies Assumption 2, then there exists $\rho \geq 0$ such that

$$g_2 - g_2(a_1^s, a_2^b) \leq \rho(\bar{g}_1 - g_1), \text{ for any } (g_1, g_2) \in F. \tag{1}$$

Assumption 3 implies that there exists an action profile $(a_1^s, a_2^b) \in A_1 \times A_2$ such that $g_1(a_1^s, a_2^b) = \bar{g}_1$. However, a_2^b need not be a best response to a_1^s. If Γ satisfies Assumption 2 or Assumption 3 and $g_2(a_1^s, a_2^b) > \hat{g}_2$, then there exists $\rho \geq 0$ such that

$$\left| \frac{g_2 - g_2(a_1^s, a_2^b)}{\bar{g}_1 - g_1} \right| \leq \rho, \text{ for any } (g_1, g_2) \in F. \tag{2}$$

We normalize payoffs, without loss of generality, such that

$$\bar{g}_1 = 1; \; g_1(a_1, a_2) \geq 0 \text{ for all } a \in A; \text{ and } g_2(a_1^s, a_2^b) = 0. \tag{3}$$

The repeated game where the initial probability over Ω is μ and the discount factor is δ is denoted $\Gamma^\infty(\mu, \delta)$. The analysis in the paper focuses on Bayes-Nash equilibria (NE) of the game of incomplete information $\Gamma^\infty(\mu, \delta)$. In equilibrium, beliefs are obtained, where possible, using Bayes' rule given $\mu(\cdot|h_0) = \mu(\cdot)$ and conditioning on players' equilibrium strategies.

The dynamic Stackelberg payoff, strategy and type. Let

$$U_1^s(\delta) = \sup_{\sigma_1 \in \Sigma_1} \inf_{\sigma_2 \in BR(\sigma_1, \delta)} U_1(\sigma_1, \sigma_2, \delta),$$

where $BR(\sigma_1, \delta)$ denotes the set of best responses of player 2 to the repeated game strategy σ_1 of player 1 in game $\Gamma^\infty(\delta)$. Let $\sigma_1^s(\delta)$ denote a strategy that satisfies

$$\inf_{\sigma_2 \in BR(\sigma_1^s(\delta), \delta)} U_1(\sigma_1^s(\delta), \sigma_2, \delta) = U_1^s(\delta),$$

if such a strategy exists. We call $U_1^s(\delta)$ the dynamic Stackelberg payoff and $\sigma_1^s(\delta)$ a dynamic Stackelberg strategy for player 1.[6] In words, the dynamic Stackelberg

[5]In Atakan and Ekmekci (2008) we define locally non-conflicing interest games using the mixed minimax of player 2 instead of the pure minimax as we do here. We are able to do with the less stringent formulation in Atakan and Ekmekci (2008) because the pure and mixed minimax for player 2 coincide in extensive form games of perfect information.

[6]The terminology follows Aoyagi (1996) and Evans and Thomas (1997).

payoff for player 1 is the highest payoff that he can guarantee in the repeated game through public pre-commitment to a repeated game strategy (a dynamic Stackelberg strategy). The static Stackelberg payoff for player 1 is the highest payoff that he can guarantee in the stage game through public pre-commitment to a stage game action (see Fudenberg or Levine (1989, 1992) or Mailath and Samuelson (2006) for a precise definition). If Assumption 2 or Assumption 3 is satisfied by Γ, then $\lim_{\delta \to 1} U_1^s(\delta) = 1$. Also, if Γ satisfies Assumption 2, then the repeated game strategy for player 1 that plays a_1^s in each period is a dynamic Stackelberg strategy. *We let S denote the commitment type that plays a_1^s in each period, i.e, S is the Stackelberg type.* If Γ satisfies Assumption 2, then the static and dynamic Stackelberg payoffs for player 1 coincide. However, if Γ satisfies Assumption 3, but does not satisfy Assumption 2, then the dynamic Stackelberg payoff for player 1 exceeds his static Stackelberg payoff (see the discussion centered around Example 1 in section 4).

3. A Reputation Result with Finite Automata

A *finite automaton* $(\Theta, \theta_0, f, \tau)$ consists of a finite set of states Θ, an initial state $\theta_0 \in \Theta$, an output function $f : \Theta \to \Delta(A_1)$ that assigns a (possibly mixed) stage game action to each state, and a transition function $\tau : Y \times A_1 \times \Theta \to \Theta$ that assigns a state to each outcome of the stage game. Specifically, the action chosen by an automaton in period t is determined by the output function f given the period t state θ_t of the automaton. The evolution of states for the automaton is determined by the transition function τ. The Stackelberg type S, defined previously, is a particularly simple finite automaton with a single state.

In this section we assume that the set of commitment types $\Omega \setminus \{N\}$ is any set of finite automata that includes the Stackelberg type S. Also, we maintain Assumption 1 and Assumption 2. The main result of this section, Theorem 2, shows that Player 1 can guarantee a payoff arbitrarily close to one, in any NE of the repeated game, if he is sufficiently patient. Theorem 2 holds for any measure μ over the set of commitment types with $\mu(S) > 0$. Under Assumption 2 the dynamic Stackelberg payoff and static Stackelberg payoff of player 1 coincide and are equal to one. Consequently, Theorem 2 shows that a sufficiently patient player 1 guarantees his dynamic Stackelberg payoff by playing a_1^s in each period.

First we state two intermediate results, Theorem 1 and Lemma 1, that are essential for our main result. Let $\Omega_- = \Omega \setminus \{S, N\}$ denote the set of commitment types other than the Stackelberg type. Theorem 1 bounds player 1's equilibrium payoff as a function of the discount factor, $\mu(S)$, and $\mu(\Omega_-)$. In particular, the theorem implies that if $\mu(S) > 0$, and $\mu(\Omega_-)$ is sufficiently close to zero, and the discount factor is sufficiently close to one, then player 1's payoff is close to one in any NE of the repeated game. The proof of Theorem 1 is presented in this section following the statement of Theorem 2.

Theorem 1. *Assume that $S \in \Omega$. Posit Assumption 1 and Assumption 2. For any $\mu^* \in interior(\Delta(\Omega))$ and any NE profile σ of $\Gamma^\infty(\mu^*, \delta)$,*

$$U_1(\sigma, \delta) > 1 - K(\mu^*)^{\bar{n}} \max\left\{1 - \delta, \frac{\mu^*(\Omega_-)}{\mu^*(S)}\right\}.$$

where $K(\mu^*) = \frac{4\rho(1+(4+\underline{\pi})M)}{l\underline{\pi}\mu^*(S)}$ and $\bar{n} = 1 + \frac{\ln(\mu^*(S))}{\ln(1-\frac{l\pi\mu^*(S)}{4\rho})}$.

Lemma 1, our uniform learning result, shows that player 2's posterior belief that player 1 is a type in Ω_- becomes arbitrarily small, as player 1 repeatedly plays the Stackelberg action. The intuition for the result is straightforward: the full support assumption implies that each state of a finite automata will be visited repeatedly as player 1 plays a_1^s. This implies that player 2 can reject the hypothesis that the sequence of play is generated by any finite automata that plays any action other than a_1^s in any of its states with arbitrarily high probability, regardless of which strategy player 2 uses. The proof of the lemma is given in the appendix.

Lemma 1 (Uniform Learning)**.** *Assume that $\Omega \setminus \{N\}$ is any set of finite automata that contains S and posit Assumption 1. Suppose player 1 has played only a_1^s in history h_t, and let $S(h_t) \supset \{S\}$ denote the set of types that behave identical to the Stackelberg type given h_t and let $\Omega_-(h_t) \subset \Omega_-$ denote the set of commitment types not in $S(h_t)$. For any $\mu \in interior(\Delta(\Omega))$ and any $\phi > 0$, there exists a T such that, $\Pr_{(\sigma_1(S),\sigma_2)}\{h : \frac{\mu(\Omega_-(h_T)|h_T)}{\mu(S(h_T)|h_T)} < \phi\} > 1 - \phi$, for any strategy σ_2 of player 2.*

Our main result, Theorem 2, puts together our findings presented as Theorem 1 and Lemma 1: player 1 can ensure a payoff arbitrarily close to one by first manipulating player 2's beliefs so that the posterior belief of Ω_- is sufficiently low and then obtaining the high payoff outlined in Theorem 1.

Theorem 2. *Assume that $\Omega \setminus \{N\}$ is any set of finite automata that contains S. Posit Assumption 1 and Assumption 2. For any $\mu^* \in interior(\Delta(\Omega))$ and any $\gamma > 0$, there exists a $\underline{\delta} \in [0,1)$ such that if $\delta > \underline{\delta}$, then in any NE profile σ of $\Gamma^\infty(\mu^*, \delta)$,*

$$U_1(\sigma, \delta) > 1 - \gamma.$$

Proof. Pick δ_1 such that $K^{\bar{n}}(1-\delta_1) < \gamma/2$. By Lemma 1 there exists period T such that $\frac{\mu^*(\Omega_-(h_T)|h_T)}{\mu^*(S(h_T)|h_T)} < 1-\delta_1$ if player 1 plays a_1^s in each period until period T. The fact that $\mu^*(S(h_T)|h_T) \geq \mu^*(S)$ and Theorem 1 implies that player 1's payoff following history h^T is at least $1 - K^{\bar{n}} \max\left\{1-\delta, \frac{\mu^*(\Omega_-(h_T)|h_T)}{\mu^*(S(h_T)|h_T)}\right\}$. Consequently, if $\delta > \delta_1$, then $U_1(\sigma,\delta) \geq \delta^T(1-K^{\bar{n}}(1-\delta_1)) > \delta^T(1-\gamma/2)$. Consequently, we can pick $\underline{\delta} > \delta_1$ such that if $\delta > \underline{\delta}$, then $\delta^T(1-\gamma/2) > 1-\gamma$. □

The remainder of the development in this section outlines the argument for Theorem 1. We begin by introducing some definitions. Let the resistance of strategy σ_2

$$r(\sigma_2, \delta) = 1 - U_1(\sigma_1(S), \sigma_2, \delta).$$

Below we define the maximal resistance function, $R(\mu, \delta)$, which is an upper-bound on how much player 2 can resist (or hurt) type S in any NE of $\Gamma^\infty(\mu, \delta)$.

Definition 1 (Maximal resistance function)**.** *For any measure $\mu \in \Delta(\Omega)$ and $\delta \in [0,1)$ let*

$$R(\mu,\delta) = \sup\{r(\sigma_2,\delta) : \sigma_2 \textit{ is part of a NE profile } \sigma \textit{ of } \Gamma^\infty(\mu,\delta)\}.$$

In the following lemma we bound both player 1's and player 2's equilibrium payoff as a function of the maximal resistance function. We say that player 1 deviated from $\sigma_1(S)$ in the t^{th} period of infinite public history h if a_1^t is not the same as $\sigma_1(S, h^t)$. We use the fact that at any period player 1 deviates from $\sigma_1(S)$ he can instead play according to $\sigma_1(S)$ and ensure a payoff of $1 - R(\mu,\delta)$. The bound on player 1's payoff in conjunction with equation (1) or (2) then implies a similar on bound player 2's payoff.

Lemma 2. *Posit Assumption 1 and Assumption 2. Pick any NE profile σ of $\Gamma^\infty(\mu,\delta)$ and period t public history h^t. Let $\mu'(\cdot) = \mu(\cdot|h^t, a_1^s, y)$ for any y. Suppose that $a_1 \in supp(\sigma_1(N,h^t))$ and $a_1 \neq \sigma_1(S)$, i.e., player 1 deviates from $\sigma_1(S)$ with positive* *probability. Then,*

$$U_1(\sigma,\delta|h^t,a_1,\sigma_2(h^t)) \geq 1 - R(\mu',\delta) - 2M(1-\delta).$$

Consequently,

$$|U_2(\sigma_1(N),\sigma_2|h^t,a_1,a_2)| \leq \frac{\rho}{\underline{\pi}}(R(\mu',\delta)+2M(1-\delta)), \text{ for any } a_2 \in A_2, \text{ if } g_2(a_1^s,a_2^b) > \hat{g}_2,$$

$$U_2(\sigma_1(N),\sigma_2|h^t,a_1,\sigma_2(h^t)) \leq \rho(R(\mu',\delta)+2M(1-\delta)), \text{ otherwise.}$$

Proof. Player 1's payoff from playing any $a_1 \neq a_1^s$ is at most

$$(1-\delta)M + U_1(\sigma|h^t,a_1,\sigma_2(h^t)) = (1-\delta)M + \sum_{y\in Y}\pi_y(\sigma_2(h^t))U_1(\sigma|h^t,a_1,y).$$

If player 1 instead plays a_1^s in period t, then he receives at least zero for the period and beliefs are updated to $\mu'(\cdot)$. This implies that his continuation payoff is at least $1 - R(\mu',\delta)$. Consequently, for any $a_1 \in supp(\sigma_1(h^t))$,

$$\begin{aligned}(1-\delta)M + U_1(\sigma|h^t,a_1,\sigma_2(h^t)) \geq \delta(1-R(\mu',\delta)) \geq 1 - R(\mu',\delta) - M(1-\delta)\\ U_1(\sigma|h^t,a_1,\sigma_2(h^t)) \geq 1 - R(\mu',\delta) - 2M(1-\delta).\end{aligned}$$

For any $y \in Y$, $(U_1(\sigma|h^t,a_1,y), U_2(\sigma_1(N),\sigma_2|h^t,a_1,y)) \in F$. If Γ satisfies Assumption 2, then equations (1) and (2), imply that

$$\begin{aligned}U_2(\sigma_1(N),\sigma_2|h^t,a_1,y) &\leq \rho(1-U_1(\sigma|h^t,a_1,y))\\ \sum_{y\in Y}\pi_y(\sigma_2(h^t))U_2(\sigma_1(N),\sigma_2|h^t,a_1,y) &\leq \sum_{y\in Y}\pi_y(\sigma_2(h^t))\rho(1-U_1(\sigma|h^t,a_1,y))\\ U_2(\sigma_1(N),\sigma_2|h^t,a_1,\sigma_2(h^t)) &\leq \rho(1-U_1(\sigma|h^t,a_1,\sigma_2(h^t)))\\ U_2(\sigma_1(N),\sigma_2|h^t,a_1,\sigma_2(h^t)) &\leq \rho(R(\mu',\delta)+2M(1-\delta))\end{aligned}$$

If Γ satisfies Assumption 2 and $g_2(a_1^s,a_2^b) > \hat{g}_2$, then equation (2) implies that

$$\begin{aligned}|U_2(\sigma_1(N),\sigma_2|h^t,a_1,y)| &\leq \rho(1-U_1(\sigma|h^t,a_1,y))\\ \sum_{y\in Y}\pi_y(a_2)|U_2(\sigma_1(N),\sigma_2|h^t,a_1,y)| &\leq \sum_{y\in Y}\pi_y(a_2)\rho(1-U_1(\sigma|h^t,a_1,y))\end{aligned}$$

If Γ satisfies Assumption 2 and $g_2(a_1^s,a_2^b) > \hat{g}_2$, then $1 - U_1(\sigma|h^t,a_1,y) \geq 0$. This if because $\bar{g}_1 = 1$ is also the highest payoff for player 1 in F. Assumption 1 implies that

$\pi_y(a_2) \leq \frac{\pi_y(\sigma_2(h^t))}{\underline{\pi}} < \infty$, for any $a_2 \in A_2$. Consequently, for any $a_2 \in A_2$,

$$\sum_{y\in Y} \pi_y(a_2)|U_2(\sigma_1(N), \sigma_2|h^t, a_1, y)| \leq \sum_{y\in Y} \frac{\pi_y(\sigma_2(h^t))}{\underline{\pi}}\rho(1 - U_1(\sigma|h^t, a_1, y))$$

$$\sum_{y\in Y} \pi_y(a_2)|U_2(\sigma_1(N), \sigma_2|h^t, a_1, y)| \leq \frac{\rho}{\underline{\pi}}(R(\mu', \delta) + 2M(1 - \delta))$$

$$\left|\sum_{y\in Y} \pi_y(a_2)U_2(\sigma_1(N), \sigma_2|h^t, a_1, y)\right| \leq \frac{\rho}{\underline{\pi}}(R(\mu', \delta) + 2M(1 - \delta))$$

$$|U_2(\sigma_1(N), \sigma_2|h^t, a_1, a_2)| \leq \frac{\rho}{\underline{\pi}}(R(\mu', \delta) + 2M(1 - \delta))$$

□

The following definition introduces reputation thresholds. In particular, $z_n(\delta, \phi)$ is the highest reputation level, such that $\mu(S) = z_n(\delta, \phi)$ and the resistance $R(\mu, \delta)$ exceeds $K(\mu)^n$, given an upper bound ϕ on the relative likelihood of any other commitment type, i.e., $\mu(\Omega_-)/\mu(S)$ is less than ϕ.

Definition 2 (Reputation Thresholds)**.** *For each $n \geq 0$, let*

$$z_n(\delta, \phi) = \sup\{z : \exists\mu \in \Delta(\Omega) \ s.t. \ R(\mu, \delta) \geq K(\mu)^n\epsilon, \mu(S) = z, \ \frac{\mu(\Omega_-)}{\mu(S)} \leq \phi\},$$

where $\epsilon = \max\{1 - \delta, \phi\}$ and $K(\mu)$ is the defined in Theorem 1.

The following definition introduces the maximal resistance in a ξ neighborhood of a given reputation level z and a given upper bound on the relative likelihood of another commitment type.

Definition 3. *For any $\xi > 0$ and $z \in (0, 1)$ let*

$$\bar{R}(\xi, z, \delta, \phi) = \sup\{r : \exists\mu \in \Delta(\Omega) \ s.t. \ R(\mu, \delta) \geq r, \mu(S) = z' \in [z - \xi, z], \frac{\mu(\Omega_-)}{\mu(S)} \leq \phi\}.$$

We will use Definition 2 and Definition 3 in conjunction with Lemma 2 to calculate upper and lower bounds for player 2's equilibrium payoffs. By definition, there exists μ such that $\mu(S) = z \in [z_n(\delta, \phi) - \xi, z_n(\delta, \phi)]$ and $\frac{\mu(\Omega_-)}{\mu(S)} \leq \phi$, and NE σ of $\Gamma^\infty(\mu, \delta)$ such that σ_2 has resistance of at least $\bar{R}(\xi, z_n, \delta, \phi) - \xi$. Also, by definition, $\bar{R}(\xi, z_n, \delta, \phi) \geq K^n\epsilon$. The definition of $z_n(\delta, \phi)$ and $\bar{R}(\xi, z_n, \delta, \phi) \geq K^n\epsilon$ implies that if $\mu(S) \in [z_n(\delta,\phi) - \xi, z_{n-1}(\delta,\phi)]$ and $\frac{\mu(\Omega_-)}{\mu(S)} \leq \phi$, then $R(\mu,\delta) \leq \bar{R}(\xi, z_n, \delta, \phi)$ in any NE profile σ of $\Gamma^\infty(\mu,\delta)$.

For a given μ such that $\mu(S) = z \in [z_n(\delta,\phi) - \xi, z_n(\delta,\phi)]$ and $\frac{\mu(\Omega_-)}{\mu(S)} \leq \phi$ in what follows we focus on an equilibrium of $\Gamma^\infty(\mu,\delta)$ where player 2 resists the Stackelberg type by approximately $\bar{R}(\xi, z_n, \delta, \phi)$. We compare player 2's payoff in this equilibrium with her payoff if she uses an alternative strategy that best responds to the Stackelberg strategy until player 1 reveals rationality. Resisting is costly for player 2 since there is a positive probability that she actually faces the Stackelberg type. The alternative strategy allows player 2 to avoid this cost. However, player 2 may be resisting the Stackelberg type because she expects a reward in the event that she sticks with equilibrium play and player 1 reveals rationality; or because she fears punishment in the event that she best responds to the Stackelberg strategy and player 1 reveals rationality. Lemma 4 gives an upper-bound on player 2's payoff if she sticks to the equilibrium strategy. The lemma ties player 2's expected reward to $\bar{R}(\xi, z_n, \delta, \phi)$ and $K^{n-1}(\mu)$ by using Lemma 2. Also, the lemma takes into account that player 2 bears a cost against the Stackelberg type. Lemma 5 gives a lower-bound on player 2's payoff if she uses the alternative strategy. Similarly this lemma ties player 2's expected punishment to $\bar{R}(\xi, z_n, \delta, \phi)$ and $K^{n-1}(\mu)$ by using Lemma 2.

First we define a stopping time and prove a technical lemma related to the stopping time. These are needed for the upper and lower bound calculations. In particular, we use the stopping time to define the random time player 1's reputation exceeds z_{n-1} if he starts from an initial reputation level of z_n.

Definition 4 (Stopping time)**.** *For any $\mu \in \Delta(\Omega)$, $z' \in (\mu(S), 1]$, infinite public history h, and strategy profile σ, let $T(\sigma, \mu, z', h)$ denote the first period such that $\mu(S|h^t) \geq z'$, where h^t is a period t public history that coincides with the first t periods of h.*

Suppose that $T : H \to \mathbb{Z}$ is a stopping time, then $E_{[0,T)}$ denotes the set of infinite public histories h where player 1 deviates from $\sigma_1(S)$ for the first time in some period $t \in [0, T(h))$ in history h. That is $E_{[0,T)}$ is the event that player 1 deviates from the Stackelberg strategy before random time T.

Lemma 3. *For any $\mu \in \Delta(\Omega)$, $z' \in (\mu(S), 1]$, and any strategy profile σ,*

$$\Pr_{(\sigma_1(N),\sigma_2)}[E_{[0,T(\sigma,\mu,z')-1)}] < 1 - \frac{\mu(S)}{z'}.$$

Proof. See the Appendix. □

Below we use Lemma 2, Definition 2, Definition 3, Definition 4 and Lemma 3 to calculate the upper and lower bound for player 2's payoffs.

Lemma 4 (Upper-bound)**.** *Posit Assumption 1 and Assumption 2. Pick $\mu \in \Delta(\Omega)$ such that $\mu(S) = z \in [z_n(\delta,\phi) - \xi, z_n(\delta,\phi)]$ and $\frac{\mu(\Omega_-)}{\mu(S)} \le \phi$, and pick NE profile σ of $\Gamma^\infty(\mu,\delta)$ such that $r(\delta,\sigma_2) \ge \bar{R}(\xi, z_n, \delta, \phi) - \xi$. For the chosen σ,*

(4)

$$U_2(\sigma,\delta) \le \rho(q(\delta,\phi,n,\xi)\bar{R}(\xi,z_n,\delta,\phi) + \epsilon K(\mu)^{n-1} + 4M\epsilon) + M\epsilon - (\bar{R}(\xi,z_n,\delta,\phi) - \xi)(z_n(\delta,\phi) - \xi)l,$$

where $q(\delta,\phi,n,\xi) = 1 - (z_n(\delta,\phi) - \xi)/z_{n-1}(\delta,\phi)$ and $l > 0$ is a positive constant such that $g_2(a_1^s, a_2) < -l$ for any a_2 that is not a best response to a_1^s.

Proof. By Assumption 2 and normalization (3) there is a constant $l > 0$. Choose equilibrium σ such that $r(\sigma_2,\delta) \ge \bar{R}(\xi,z_n,\delta,\phi) - \xi$. Let $T(h) = T(\sigma,\mu,z_{n-1},h)$. Recall that $E_{[0,T-1)}$ denotes the event that player 1 reveals rationality before random time $T(h) - 1$, in other words, player 1 reveals rationality before the posterior probability that he is type S exceeds z_{n-1}. We bound player 2's payoff in the events player 1 is the normal type and $E_{[0,T-1)}$ occurs, the event that player 1 is the normal type and the event $E_{[T-1,\infty)}$ occures, the event that player 1 is the Stackelberg type, and the event that player 1 is any other type. Player 2's payoff until player 1 deviates from $\sigma_1(S)$ is at most zero by normalization (3). If $\mu(S) = z$, $\frac{\mu(\Omega_-)}{\mu(S)} \le \phi$ and player 1 has not deviated from $\sigma_1(S)$ in h^t, then Bayes' rule implies that $\frac{\mu(\Omega_-|h^t)}{\mu(S|h^t)} \le \phi$ and $\mu(S|h^t) \ge z$. So, player 2's continuation payoff after player 1 deviates from $\sigma_1(S)$ is at most $\rho(\bar{R}(\xi,z_n,\delta,\phi) + 2M(1-\delta))$ if the deviation occurs at $t < T - 1$; and is at most $\rho(K^{n-1}\epsilon + 2M(1-\delta))$ if the deviation occurs at $t \ge T - 1$, by Lemma 2. Consequently, $U_2(\sigma_1(N),\sigma_2,\delta|E_{[0,T-1)}) \le \rho(\bar{R}(\xi,z_n,\delta,\phi) + 2M(1-\delta))$ and $U_2(\sigma_1(N),\sigma_2,\delta|E_{[T-1,\infty)}) \le \rho(K^{n-1}\epsilon + 2M(1-\delta))$. Also, $U_2(\sigma_1(S),\sigma_2,\delta) \le -l(\bar{R}(\xi,z_n,\delta,\phi) - \xi)$. Player 2 can get at most M against any other type ω. The probability of event $E_{[0,T-1)}$ is at most $q(\delta,\phi,n,\xi)$, by Lemma 3, probability of event $E_{[T-1,\infty)}$ is at most one, the probability of S is equal to $\mu(S)$, the probability of

$\omega \in \Omega_-$ is ϕ. So,

$$U_2(\sigma, \delta) \leq \rho(q(\delta, \phi, n, \xi)\bar{R}(\xi, z_n, \delta, \phi) + \epsilon K^{n-1} + 4M\epsilon) + M\epsilon - (\bar{R}(\xi, z_n, \delta, \phi) - \xi)(z_n(\delta, \phi) - \xi)l.$$

□

Lemma 5 (Lower-bound). *Posit Assumption 1 and Assumption 2. Suppose that* $\mu(S) = z \in [z_n(\delta, \phi) - \xi, z_n(\delta, \phi)]$ *and* $\frac{\mu(\Omega_-)}{\mu(S)} \leq \phi$. *In any NE* σ *of* $\Gamma^\infty(\mu, \delta)$,

$$U_2(\sigma, \delta) \geq -\frac{\rho}{\underline{\pi}}\left(\bar{R}(\xi, z_n, \delta, \phi)q(\delta, \phi, n, \xi) + K(\mu)^{n-1}\epsilon + 4M\epsilon\right) - M\epsilon, \tag{5}$$

where $q(\delta, \phi, n, \xi) = 1 - (z_n(\delta, \phi) - \xi)/z_{n-1}(\delta, \phi)$.

Proof. Pick any NE σ of $\Gamma^\infty(z, \delta)$ and suppose that $g_2(a_1^s, a_2^b) > \hat{g}_2$. The strategy σ_2^* plays a_2^b after any period k public history h^k, if there is no deviation from $\sigma_1(S)$ in h^k, and coincides with NE strategy σ_2 if player 1 has deviated from $\sigma_1(S)$ in h^k. Strategy profile $\sigma^* = (\{\sigma_1(\omega)\}_{\omega \in \Omega}, \sigma_2^*)$ and let $T(h) = T(\sigma^*, \mu, \mu'(S), h)$. We again look at the events $E_{[0,T-1)}$, $E_{[T-1,\infty)}$, the event that player 1 is type S and the event $\omega \in \Omega_-$. Player 2's payoff until player 1 deviates from $\sigma_1(S)$ is zero by normalization (3). Consequently, $U_2(\sigma^*, \delta | E_{[0,T-1)}) \geq -\frac{\rho}{\underline{\pi}}(\bar{R}(\xi, z_n, \delta, \phi) + 2M(1-\delta))$ and $U_2(\sigma^*, \delta | E_{[T-1,\infty)}) \geq -\frac{\rho}{\underline{\pi}}(K^{n-1}\epsilon + 2M(1-\delta))$. $U_2(\sigma_1(\omega), \sigma_2^*) \leq M$ for any $\omega \in \Omega$. Also, $U_2(\sigma_1(S), \sigma^*) = 0$ by the definition of σ_2^*. So,

$$U_2(\sigma, \delta) \geq U_2(\sigma^*, \delta) \geq -\frac{\rho}{\underline{\pi}}(q(\delta, \phi, n, \xi)\bar{R}(\xi, z_n, \delta, \phi) + K^{n-1}\epsilon + 4M\epsilon) - \epsilon M.$$

If $g_2(a_1^s, a_2^b) = \hat{g}_2$, then $U_2(\sigma, \delta) \geq \hat{g}_2 = 0 \geq -\frac{\rho}{\underline{\pi}}(q(\delta, \phi, n, \xi)\bar{R}(\xi, z_n, \delta, \phi) + K^{n-1}\epsilon + 4M\epsilon) - \epsilon M$. □

Below we use the fact that the upper-bound provided in Lemma 4 must exceed the lower-bound given in Lemma 5 to complete our proof.

Proof of Theorem 1. Combining the upper and lower bounds for $U_2(\sigma, \delta)$, given by equations (4) and (5), and simplifying by canceling ϵ delivers

$$(z_n(\delta, \phi) - \xi)l\frac{\bar{R}(\xi, z_n, \delta, \phi) - \xi}{\epsilon} \leq 2\frac{\rho}{\underline{\pi}}\left(\frac{q(\delta, \phi, \xi)\bar{R}(\xi, z_n, \delta, \phi)}{\epsilon} + K(\mu)^{n-1} + 4M\right) + 2M.$$

Let $q_n(\delta, \phi) = 1 - z_n(\delta, \phi)/z_{n-1}(\delta, \phi)$. $\bar{R}(\xi, z_n, \delta, \phi) \in [0, 1]$ for each ξ, we pick any convergent subsequence and let $\lim_{\xi \to 0} \bar{R}(\xi, z_n, \delta, \phi) = \bar{R}(z_n, \delta, \phi)$. Taking $\xi \to 0$

implies that $q(\delta,\phi,n,\xi) \to q_n(\delta,\phi)$ and

$$z_n(\delta,\phi) l \bar{R}(z_n,\delta,\phi)/\epsilon \leq 2\frac{\rho}{\underline{\pi}}(q_n(\delta,\phi)\bar{R}(z_n,\delta,\phi)/\epsilon + K(\mu)^{n-1} + 4M) + 2M.$$

Rearranging,

$$q_n(\delta,\phi) \geq \frac{z_n(\delta,\phi) l \underline{\pi}}{2\rho} - \frac{K(\mu)^{n-1}\epsilon}{\bar{R}(z_n,\delta,\phi)} - \frac{4M\epsilon}{\bar{R}(z_n,\delta,\phi)} - \frac{M\epsilon\underline{\pi}}{\rho\bar{R}(z_n,\delta,\phi)}.$$

Also, $\bar{R}(\xi, z_n,\delta,\phi) \geq K(\mu)^n\epsilon$ for each ξ implies that $\bar{R}(z_n,\delta,\phi) \geq K(\mu)^n\epsilon$. Consequently,

$$q_n(\delta,\phi) \geq \frac{z_n(\delta,\phi) l \underline{\pi}}{2\rho} - \frac{K(\mu)^{n-1}}{K(\mu)^n} - \frac{4M}{K(\mu)^n} - \frac{M\underline{\pi}}{\rho K(\mu)^n}.$$

Substituting in our initial choice of $K(\mu)$ implies that $q_n(\delta,\phi) \geq \frac{\mu^*(S) l \underline{\pi}}{4\rho} \equiv \underline{q} > 0$, for any $z_n(\delta,\phi) \geq \mu^*(S)$. So, $z_n(\delta,\phi) \geq \mu^*(S)$ implies that $1 - z_n(\delta,\phi)/z_{n-1}(\delta,\phi) \geq \underline{q} > 0$ for all $\delta < 1$, $\phi > 0$ and $n = 0, 1, ..., \infty$. Then, for each $\delta < 1$ and $\phi > 0$, we have $z_{n(\underline{q})}(\delta,\phi) < \mu^*(S)$, where $n(\underline{q})$ is the smallest integer j such that $(1-\underline{q})^j < \mu^*(S)$. By definition $n(\underline{q}) \leq \bar{n}$. Consequently, if $\mu(S) \geq z_{\bar{n}(\underline{q})}(\delta,\phi)$ and $\frac{\mu(\Omega_-)}{\mu(S)} \leq \phi$, then $R(\mu,\delta) \leq K(\mu)^{\bar{n}}\epsilon$. So, $R(\mu^*,\delta) \leq K(\mu^*)^{\bar{n}}\{1-\delta, \frac{\mu^*(\Omega_-)}{\mu^*(S)}\}$. □

4. A reputation result with review strategies

In this section we extend our reputation result to any stage game that satisfies Assumption 3. If Γ satisfies Assumption 3, but does not satisfy Assumption 2, then the dynamic Stackelberg payoff of player 1 exceeds his static Stackelberg payoff. Consider the following example.

Example 1. A game that satisfies Assumption 3, but not Assumption 2.

$_1\backslash^2$	L	R
U	$3,1$	$0,2$
D	$0,0$	$0,0$

In this game player 1's static Stackelberg payoff is equal to zero whereas his dynamic Stackelberg payoff is equal to three. If player 2's actions were observed without noise, then player 1 could obtain a payoff of three by using the following repeated game strategy: player 1 starts the game by playing U; in any period if player 2 does not play L when player 1 plays U, then player 1 punishes player 2 for two periods by playing D;

after the two periods of punishment player 1 again plays U. The unique best response for a sufficiently patient player 2 to this repeated game strategy of player 1 involves playing L in each period. However, if player 2's action are observed imperfectly, i.e., if Assumption 1 holds, then a pre-commitment to the strategy described before does not guarantee player 1 a high payoff. This is because player 1 cannot observe whether player 2 has played L or R when he plays U but can observe only a imperfect signal. Consequently, in certain periods player 1 may mistakenly punish player 2, even if she played L against U; or mistakenly fail to punish player 2, even if she played R against U.

Under Assumption 1 player 1 can achieve his dynamic Stackelberg payoff by using review strategies (see Radner (1981, 1985) and Celantani et al. (1996)) which statistically test whether player 2 is playing L frequently over a sufficiently long sequence of periods where player 1 plays U. In the next subsection we discuss review strategies that can be represented as finite automata. In Lemma 6 we show that, for any accuracy level $\epsilon > 0$, there is a finite review strategy such that any best response to this review strategy gives player 1 a payoff of at least $3 - \epsilon$, if the agents are sufficiently patient.

In Theorem 3 we bound the payoff player 1 can achieve by mimicking a commitment type that is a finite automaton who plays a review strategy. However, we show that a tight reputation bound cannot be established by mimicking commitment types that are finite automata. So, our main reputation result in this section, Theorem 4, considers a commitment type with infinitely many states. Theorem 4 maintains Assumption 1 and Assumption 3 and assumes that the there are finitely many other commitment types that are each finite automata. The theorem uses the reputation bound established in Theorem 3 and shows that there exists a review type (with infinitely many states) such that player 1 can achieve a payoff arbitrarily close to his dynamic Stackelberg payoff by mimicking this review type, if he is sufficiently patient.

4.1. **Review strategies.** We begin by describing a repeated game review strategy with accuracy ϵ denoted $\sigma_1(D_\epsilon)$. Assumption 3 and normalization (3) implies that there exists a positive integer P and a positive constant $l > 0$ such that

$$g_2(a_1^s, a_2) + Pg_2(a_1^p, a_2')) < -l(P+1) \tag{6}$$

for any $a_2 \in A_2$ such that $g_1(a_1^s, a_2) < 1$ and $a_2' \in A_2$.

In order to describe strategy $\sigma_1(D_\epsilon)$, we follow Celentani et al. (1996) and first consider a KJ-fold finitely repeated game $\Gamma^{KJ}(\delta)$. We partition Γ^{KJ} into blocks of length J, $\Gamma^{J,k}$, $k = 1, ..., K$. Let u_i^k denote player i's time average payoff in 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^* be the following strategy: in block $\Gamma^{J,1}$ player 1 plays a_1^s in each period. We call a block where player 1 chooses to play a_1^s in each period a review phase. In the beginning of block $\Gamma^{J,2}$, player 1 reviews play in the previous block. If $1 - u_1^1 < \eta$, then player 1 again chooses to play a_1^s in each period of block $\Gamma^{J,2}$ and so on. If for any k, $1 - u_1^k \geq \eta$, 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 1 chooses to play a_1^p in each period a "punishment phase".

Lemma 6. *Given $\epsilon > 0$ there are numbers η, K and J with $P/K < \epsilon$ and discount factor δ^* such that for any $\delta > \delta^*$ and for any best response σ_2^* to σ_1^* in $\Gamma^{KJ}(\delta)$, the following is satisfied:*

(*i*) *If player 1 chooses a_1^s in each period of block $\Gamma^{J,k}$, $k = 1, ., ., K - P$, then $Pr(1 - u_1^k < \epsilon) > 1 - \epsilon$.*
(*ii*) *The fraction of stages k in which player 1 uses his punishment strategy is smaller than ϵ with probability $1 - \epsilon$.*
(*iii*) *In the game $\Gamma^{KJ}(\delta)$ player 1's discounted payoff $u_1^{KJ}(\sigma_1^*, \sigma_2^*, \delta) > 1 - \epsilon$.*

Proof. This construction is directly from Celentani et al. (1996) Lemma 4 and a proof can be found there. □

4.2. **A reputation bound for review strategies with accuracy ϵ.** The infinitely repeated game review strategy with accuracy ϵ, $\sigma_1(D_\epsilon)$, plays a_1^s in each period in the J_ϵ period review phase. If the time average payoff of player 1 for the J_ϵ period review phase is at least $1 - \eta_\epsilon$ then $\sigma_1(D_\epsilon)$ remains in the review phase for the next J_ϵ periods. Otherwise, $\sigma_1(D_\epsilon)$ moves to the punishment phase and plays a_1^p, for the next PJ_ϵ periods. At the end of the punishment phase, the strategy again returns to the review phase. The strategy starts the game in the review phase. The commitment type that plays strategy $\sigma_1(D_\epsilon)$ is denoted D_ϵ. The commitment type D_ϵ is a finite automaton. Also, we can define such a finite automaton review type for any given level of accuracy.

Lemma 7. *Posit Assumption 3. For each $\epsilon > 0$, there exists $\eta_\epsilon > 0$, J_ϵ and $\delta_\epsilon \in [0, 1)$ that satisfies $(1 - \delta_\epsilon)J_\epsilon < \epsilon$ such that for all $\delta > \delta_\epsilon$*

(i) If $\sigma_2 \in BR(\sigma_1(D_\epsilon), \delta)$, then $U_1(\sigma_1(D_\epsilon), \sigma_2, \delta) > 1 - \epsilon$ and $|U_2(\sigma_1(D_\epsilon), \sigma_2, \delta)| < \rho\epsilon$,

(ii) If $U_1(\sigma_1(D_\epsilon), \sigma_2, \delta) = 1 - \epsilon - r$ and $r > 0$, then $U_2(\sigma_1(D_\epsilon), \sigma_2, \delta) \leq \rho\epsilon - lr$.

Proof. The choice of J_ϵ and η_ϵ and the bound of player 1's payoff in item (i) follows from Lemma 6. The bound on player 2's payoff follows from Assumption 3 and equation (2). Proof of item (ii) is in the appendix. □

The previous lemma implies that for all discount factors that exceed the cutoff δ_ϵ, player 1's payoff is at least $1 - \epsilon$, if player 1 uses strategy $\sigma_1(D_\epsilon)$ and player 2 plays a best response to $\sigma_1(D_\epsilon)$. Also, this is true for each ϵ. Consequently, player 1's dynamic Stackelberg payoff converges to one as $\delta \to 1$.

The resistance of strategy σ_2 is given by $r(\sigma_2, \delta) = \max\{1 - \epsilon - U_1(\sigma_1(D_\epsilon), \sigma_2, \delta), 0\}$ and let $R(\mu, \delta)$ denote the maximal resistance in game $\Gamma^\infty(\mu, \delta)$ as outlined in Definition 1. The following theorem gives a reputation bound for player 1's payoff. The theorem is similar to Theorem 1, but player 1 builds a reputation by mimicking review type D_ϵ instead of type S. The proof of the theorem follows the argument for Theorem 1 very closely.

Theorem 3 (Payoff Bound for Player 1)**.** *Posit Assumption 1 and Assumption 3. For any $\epsilon > 0$, if $D_\epsilon \in \Omega$ and $\mu^* \in interior(\Delta(\Omega))$, then for any NE profile σ of $\Gamma^\infty(\mu^*, \delta)$,*

$$U_1(\sigma, \delta) > 1 - \epsilon - K(\mu^*(D_\epsilon)) \max\{J_\epsilon(1 - \delta), \mu^*(\Omega_-), \epsilon\}.$$

where $\Omega_- = \Omega \setminus \{N, D_\epsilon\}$ and $K(\mu^*(D_\epsilon)) = \frac{P}{\mu^*(D_\epsilon)} \left(\frac{4\rho(1+(4+\underline{\pi})M)}{l\underline{\pi}\mu^*(D_\epsilon)}\right)^{\left(1+\frac{\ln(\mu^*(D_\epsilon))}{\ln(1-\frac{l\underline{\pi}\mu^*(D_\epsilon)}{4\rho})}\right)}$.

Proof. See the appendix. □

4.3. **A reputation result with review strategies.** A uniform learning argument, similar to Lemma 1, in conjunction with Theorem 3, implies, for any game that satisfies Assumption 1 and Assumption 3, that

$$\lim_{\delta \to 1} U_1(\sigma(\delta, \mu^*), \delta) \geq 1 - 3\epsilon/2 - K(\mu^*(D_\epsilon))\epsilon,$$

where $\sigma(\delta, \mu^*)$ is any NE profile of $\Gamma^\infty(\delta, \mu^*)$. However, as $\mu(D_\epsilon) \to 0$, $K(\mu(D_\epsilon))\epsilon \to \infty$ and this bound becomes vacuous. This is in stark contrast to games that satisfy Assumption 1 and Assumption 2 for which Theorem 2 implies that

$$\lim_{\mu(S)\to 0} \lim_{\delta \to 1} U_1(\sigma(\delta, \mu), \delta) = 1,$$

where $\sigma(\delta, \mu)$ is any NE profile of $\Gamma^\infty(\delta, \mu)$. Consequently, Theorem 3, unlike Theorem 2, provides only a weak reputation bound for player 1's payoffs. Also, mimicking more and more accurate review types will not help to strengthen Theorem 3. This is because the probability mass that μ^* places on more and more accurate commitment types may converge to zero relatively fast. This again implies a non binding reputation bound, i.e, for a fixed measure μ^*, it is possible that $\lim_{\epsilon\to 0} K(\mu^*(D_\epsilon))\epsilon > 1$. In order to provide a tight reputation bound in what follows we construct an infinitely accurate review type that we denote by D_∞. Our reputation result for stage games that satisfy Assumption 3, Theorem 4, shows that a patient player 1 can guarantee a payoff arbitrarily close to one by mimicking type D_∞.

Type D_∞ plays a review strategy with accuracy ϵ for the first T_1 periods, then plays a review strategy with accuracy $\epsilon/2$ for T_2 periods, plays a review strategy with accuracy ϵ/n for T_n periods, and so on. As the review strategy that D_∞ plays increases in accuracy (i.e., ϵ/n becomes small), so does the number of periods that the commitment type plays the particular review strategy, i.e, $T_{n-1} < T_n$. For any finite T periods, there is a cutoff $\delta(T)$ such that, for all discount factors that exceed this cutoff, the payoff in the first T periods is negligible for player 1. So, a sufficiently patient player 1 can mimic type D_∞, for a long but finite period of time, without impacting his payoff, weakly increase his reputation, and also obtain any required accuracy in the continuation game. In the following theorem we make this line of reasoning argument exact.

Theorem 4. *Posit Assumption 1 and Assumption 3 and assume that Ω is a finite set. There exists a strategy $\sigma_1(D_\infty)$ such that for any $\gamma > 0$ there exists $\underline{\delta} \in [0, 1)$ such that if $\delta \in [\underline{\delta}, 1)$, $\mu^* \in interior(\Delta(\Omega \cup \{D_\infty\}))$, then*

$$U_1(\sigma, \delta) > 1 - \gamma,$$

in any NE profile σ of $\Gamma^\infty(\mu^*, \delta)$.

Proof. See the Appendix. □

In order to provide some intuition for Theorem 4, we first describe a commitment type $D_{\epsilon/2}$: the type plays a strategy that coincides with D_ϵ up to time T_1, then plays a review strategy with accuracy $\epsilon/2$ forever. Theorem 3 implies that player 1's continuation payoff after time T_1 from mimicking $D_{\epsilon/2}$ is at least $1-\epsilon/2-K(\mu(D_{\epsilon/2}))\max\{\epsilon/2, J_{\epsilon/2}(1-\delta), \mu(\Omega_-)\}$. We pick the number of periods T_1 such that for all δ greater than a cutoff δ_1 player 1's payoff

$$U_1(\sigma,\delta) > 1-2\epsilon-K(\mu(D_{\epsilon/2}))\max\{\mu(\Omega_-),\epsilon\}$$

in any NE profile σ of $\Gamma^\infty(\mu,\delta)$. We show that T_1 can indeed be chosen in this way in Lemma 9 given in the appendix. Lemma 9 applies a limit theorem of Fudenberg and Levine (1983, 1986) to show that player 1's ϵ NE payoff in the finite repeated game $\Gamma^{T_1}(\mu,\delta)$ can be approximated by player 1's payoff, from mimicking D_ϵ, in the infinite repeated game $\Gamma^\infty(\mu,\delta)$, then uses Theorem 3 to establish the above bound. Also, because player 1 can guarantee a continuation accuracy of $\epsilon/2$ after T_1 by mimicking $D_{\epsilon/2}$, there exists a cutoff $\delta_2>\delta_1$ such that for all $\delta>\delta_2$

$$U_1(\sigma,\delta) > 1-\epsilon-K(\mu(D_{\epsilon/2}))\max\{\mu(\Omega_-),\epsilon/2\}.$$

Similarly, we can define $D_{\epsilon/n}$ inductively as the type that plays a strategy that coincides with $D_{\epsilon/(n-1)}$ up to time T_{n-1} and then plays a review strategy with accuracy ϵ/n. T_{n-1} is chosen to ensure that, for all $\delta>\delta_{n-1}$

$$U_1(\sigma,\delta) > 1-2\epsilon/(n-1)-K(\mu(D_{\epsilon/n}))\max\{\mu(\Omega_-),\epsilon/(n-1)\};$$

and for all δ that exceed cutoff $\delta_n>\delta_{n-1}$,

$$U_1(\sigma,\delta) \geq 1-2\epsilon/n-K(\mu(D_{\epsilon/n}))\max\{\epsilon/n,\mu(\Omega_-)\},$$

for any NE profile σ of$(\Gamma^\infty(\delta,\mu))$. The infinitely accurate review type D_∞ that Theorem 4 considers plays according to D_ϵ up to time T_1, plays according to $D_{\epsilon/2}$ up to time T_2, plays according to $D_{\epsilon/n}$ up to time T_n and so on. The theorem shows that by mimicking D_∞ player 1 can ensure that $U_1(\sigma,\delta)\geq 1-\epsilon/n-K(\mu(D_\infty))\max\{\mu(\Omega_-),\epsilon/n\}$ for any n, if he is sufficiently patient. Also, a learning argument similar to Lemma 1 implies that as the number of period T that player 1 mimics type D_∞ gets large $\mu(\Omega_-|h^T)$ becomes arbitrarily small. Consequently, for any NE profile $\sigma(\delta,\mu)$ of

$\Gamma^{\infty}(\delta, \mu)$,

$$\lim_{\delta \to 1} U_1(\sigma(\delta, \mu), \delta) = 1, \text{ and, } \lim_{\mu(D_\infty) \to 0} \lim_{\delta \to 1} U_1(\sigma(\delta, \mu), \delta) = 1,$$

thus providing us with a tight reputation bound as in Theorem 2.

5. Relation to One-Sided Reputation Results for Repeated Games with Perfect Information

In this section we compare the findings in this paper with our closely related one-sided reputation result presented in Atakan and Ekmekci (2008). Both papers focus on two-player repeated games with equal discount factors. Atakan and Ekmekci (2008) proves a one-sided reputation result, for subgame perfect equilibria, under the assumption that the stage-game is a extensive form game of perfect information (i.e., all information sets are singletons). In this paper we assume imperfect observation with full support (Assumption 1) instead of perfect information. Under Assumption 1 we prove a reputation result (Theorem 2) in which we are able to

(1) Deal with stage-games in normal form;
(2) Weaken the equilibrium concept from subgame perfect equilibrium to Bayes-Nash;
(3) Significantly expand the set of possible commitment types.

Theorem 2 allows for a richer set of commitment types. In fact, the set of commitment types is taken as any (countable) set of finite automata.[7] In contrast, Atakan and Ekmekci (2008) allows only for uniformly learnable types. A type is uniformly learnable, if that type reveals itself as different from the Stackelberg type with probability uniformly bounded away from zero, over a sufficiently long history of play, independent of the strategy used by player 2. In the current paper, Lemma 1 uses the properties of finite automata and the learning result of Fudenberg and Levine (1992) to show that, under the full support assumption, any finite automaton is uniformly learnable. In Atakan and Ekmekci (2008), in contrast, some finite automata are not uniformly learnable, and the set of commitment types allowed is more restricted. For example, consider a finite automaton (a "perverse type") that always play the Stackelberg action as long as player 2 plays a non-best response to the Stackelberg action, and minimaxes player 2 forever, if player 2 ever best responds to the

[7]We restrict attention only to countable subsets in order to avoid dealing with issues of measurability.

Stackelberg action. If the probability of such a type is sufficiently large, then this will induce player 2 to never play a best response to the Stackelberg action and so a reputation result is precluded in Atakan and Ekmekci (2008). This type however is not uniformly learnable in the setting of Atakan and Ekmekci (2008) and consequently ruled out. This is because this perverse type always plays the same action as the Stackelberg strategy if player 2 chooses a strategy that never best responds to the Stackelberg action. In contrast, with the full support assumption, any finite automata type will be learned with probability close to one by player 2. Consider a perverse type that minmaxes player 2 after some (finite) sequence of outcomes. The full support assumption implies that the sequence of outcomes that leads to the perverse type to start minimaxing player 2 will occur almost surely, regardless of the strategy chosen by player 2. Consequently, the perverse type will reveal itself to be different than the Stackelberg type with probability close to one.

Replacing the perfect information assumption with the full support assumption to obtain Theorem 2, however, comes at a cost. The reputation result in Atakan and Ekmekci (2008) covers all stage games that satisfy Assumption 2 and Assumption 3 whereas Theorem 2 only covers those games that satisfy Assumption 2. We do extend our reputation result to games that satisfy Assumption 3 in Theorem 4. But for the reputation result in Theorem 4 we need to introduce a infinitely complex review type. In contrast in Atakan and Ekmekci (2008) the Stackelberg type needed to establish the reputation result is always a finite automaton.[8] Also, we can no longer claim, as we do in Theorem 2, that we can cover a richer set of commitment types in Theorem 4 as compared to Atakan and Ekmekci (2008). This is because type D_∞ that player 1 mimics to obtain a high payoff is not a finite automata while the other commitment types we assume to be finite automata.

Appendix A. A reputation result with finite automata: omitted proofs

Proof of Lemma 1. Step 1. The set of histories is viewed as a stochastic process and $\Pr_{(\sigma_1(s),\sigma_2)}$ as the probability measure over the set of histories H generated by $(\sigma_1(S), \sigma_2)$. We show that for each finite subset $W \subset \Omega_-$ and any $\varepsilon > 0$, there exists

[8]The finite autamata used to establish the reputation result in Atakan and Ekmekci (2008) are similar to the strategy that obtains player 1's dynamic Stackelberg payoff under perfect observation of actions following Example 1.

a T such that, $\Pr_{(\sigma_1(S),\sigma_2)}\{h : \mu_1(W \cap \Omega_-(h_T)|h_T) < \varepsilon\} > 1 - \varepsilon$, for any strategy σ_2 of player 2. Proving this is sufficient for the result since W can be picked such that $\mu_1(W)$ is arbitrarily close to $\mu(\Omega_-)$ and $\mu(S(h_T)|h_T) \geq \mu(S|h_T) > \mu(S)$.

Step 2. Each ω is a finite automaton. Consider the stochastic process over states generated by σ^s. Since $(a_1)_t = a_1^s$ for all t, the transitions between states depend only on the realizations of y_2. In particular, probability to transition from θ_1 to θ_2 after history h_t is as follows:

$$p(\theta_1, \theta_2, \sigma_2(h_t)) = \sum_{\{y_2 \in Y_2 : \tau(y_2, a_1^s, \theta_1) = \theta_2\}} \pi_{y_2}(\sigma_2(h_t)) \geq \sum_{\{y_2 \in Y_2 : \tau(y_2, a_1^s, \theta_1) = \theta_2\}} \underline{\pi}$$

Observe $p(\theta_1, \theta_2, \sigma_2(h_t)) > 0$ if and only if $\sum_{\{y_2 \in Y_2 : \tau(y_2, a_1^s, \theta_1) = \theta_2\}} \underline{\pi} > 0$. Consequently, if $p(\theta_1, \theta_2, \sigma_2(h_t)) > 0$ for some σ_2 and h_t pair, then $p(\theta_1, \theta_2, \sigma_2'(h_k)) \geq \underline{\pi}$ for all σ_2' and h_k. This implies that the state space Θ_ω of any ω can be (uniquely) partitioned into transitory states (Θ_ω^0) and a collection of disjoint ergodic sets (Θ_ω^j) such that $\Theta_\omega = \cup_{i=0}^M \Theta_\omega^i$.(see Billingsley (1995), Chapter 1, Section 8, or Stokey, Lucas, and Prescott (1989), Chapter 11.1.) This partition is independent of σ_2' and h_k because if $p(\theta_1, \theta_2, \sigma_2(h_t)) > 0$ for some $\sigma_2(h_t)$, then $p(\theta_1, \theta_2, \sigma_2'(h_k)) \geq \underline{\pi}$ for all σ_2' and h_k.

Step 3. Let $E(T, K, \varepsilon)$ denote the set of histories such that for any $h \in E(T, K, \varepsilon)$, after initial history h_l, all $\omega \in W$ have entered an ergodic subset of states $\Theta_\omega^i \subset \Theta_\omega$ and all states $\theta \in \Theta_\omega^i$ have been visited at least K times by period $l+T$. Since there are only a finite number of sets that must be considered for $\omega \in W$ (i.e, $|W| \max_{\omega \in W} |\Theta_\omega|$ in total), for each K and each $\varepsilon > 0$ there exists a finite time T such that the set $E(T, K, \varepsilon)$ has measure at least $1 - \varepsilon$ under $\sigma_1(S)$, i.e., $\Pr_{\sigma_1(S)}\{E(T, K, \varepsilon)\} > 1 - \varepsilon$. As a consequence of the above step this time T can be picked independently from player 2's strategy σ_2. That is $\Pr_{\sigma_1(S)}\{E(T, K, \varepsilon)\} > 1 - \varepsilon$ for any σ_2.

Step 4. Let $p_t^\omega(h)$ denote the probability that a_1^s is played in period t after history h_t, conditional on being type ω. Also, let $L_t^\omega(h) = \frac{p_t^\omega(h)}{p_t^S(h)} L_{t-1}^\omega(h)$ and $L_0^\omega(h) = \frac{\mu_0(\omega)}{\mu_0(S)}$. By Fudenberg and Levine (1992) Lemma 4.1, $L_t^\omega(h) = \frac{\mu(\omega|h_t)}{\mu(S|h_t)}$ and (L_t^ω, H_t) is a super-martingale, under $\sigma_1(S)$. Observe $p_t^S(h) = 1$ for $\sigma_1(S) - a.e.$ history. Let $L^\omega(K, \varepsilon)$ denote the set of histories such that either $L_T^\omega(h) < \varepsilon$ or $|p_t^\omega(h) - 1| < \varepsilon$ in all but K periods for any $T > K$. Fudenberg and Levine (1992) Theorem 4.1 implies that there exists a K_ω independent of σ_2 such that $\Pr_{\sigma_1(S)}\{L^\omega(K_\omega, \varepsilon)\} > 1 - \varepsilon$.

Step 5. Let $\xi = \min_{\theta \in \{\theta \in \cup_{\omega \in W} \theta_w : f^{\omega}_{a^s_1}(\theta) \neq 1\}} (1 - f^{\omega}_{a^s_1}(\theta))$. That is, $f^{\omega}_{a^s_1}(\theta)$ is the probability that type ω plays a^s_1 in state θ and ξ is the minimum of the set of numbers $1 - f^{\omega}_{a^s_1}(\theta)$ over the set θ such that $f^{\omega}_{a^s_1}(\theta)$ is different than 1. The minimum is well defined since each ω is a finite automaton, W is a finite set and $\{\theta \in \cup_{\omega \in W} \Theta_w : f^{\omega}_{a^s_1}(q) \neq 1\} \neq \emptyset$ because $W \subset \Omega_-$.

Step 6. Pick $\frac{\varepsilon}{2|W|} < \xi$. Pick K such that $K > K_\omega$ and $\Pr_{\sigma_1(S)}\{L^\omega(K_\omega, \frac{\varepsilon}{2|W|})\} > 1 - \frac{\varepsilon}{2|W|}$ for all $\omega \in W$. Pick T such that $\Pr_{\sigma_1(S)}\{E(T, K, \frac{\varepsilon}{2})\} > 1 - \frac{\varepsilon}{2}$. Consequently,

$$\Pr_{\sigma_1(S)}\{E(T, K, \frac{\varepsilon}{2}) \cap (\cap_{\omega \in W} L^\omega(K_\omega, \frac{\varepsilon}{2|W|}))\} > 1 - \varepsilon.$$

By Step 3, for any $h \in E(T, K, \frac{\varepsilon}{2}) \cap (\cap_{\omega \in W} L^\omega(K_\omega, \frac{\varepsilon}{2|W|}))$ all ω are in an ergodic set $\Theta^*_\omega(h) = \Theta^{i(\omega)}_\omega(h)$ and all ergodic states $\theta \in \cup_\omega \Theta^*_\omega$ have been visited more than K times by time T. By Step 4, either $L^\omega_T(h) < \frac{\varepsilon}{2|W|}$ or $|p^\omega_t(h) - 1| < \frac{\varepsilon}{2|W|} < \xi$ all but K times. However, by the definition of ξ in Step 5, either $L^\omega_T(h) < \frac{\varepsilon}{2|W|}$ or for any ω with $L^\omega_T(h) > \frac{\varepsilon}{2|W|}$ for all $\theta \in \Theta^*_\omega(h)$, $f_{a^s_1}(q) = 1$. That is, either $L^\omega_T(h) < \frac{\varepsilon}{2|W|}$ or $\omega \in S(h_T)$. So all $\omega \in W$ with $\mu(\omega|h_T) > \frac{\varepsilon}{2|W|}$ are in $S(h_T)$. Hence $\mu(W \cap \Omega_-(h_T)|h_T) < \frac{\varepsilon}{2}$ for any $h \in E(T, K, \frac{\varepsilon}{2}) \cap (\cap_{\omega \in W} L^\omega(K_\omega, \frac{\varepsilon}{2|W|}))$ delivering the result. □

Proof of Lemma 3. Let $\bar{E}_{[0,T(\sigma,\mu,z')-1)}$ denote the complement of event $E_{[0,T(\sigma,\mu,z')-1)}$. Player 1 has not deviated from the Stackelberg strategy in any period $t < T(\sigma, \mu, z', h)$ in history h if and only if $h \in \bar{E}_{[0,T(\sigma,\mu,z')-1)}$ and

$$\mu(S|h^{T(\sigma,\mu,z',h)-1}) = \frac{\mu(S) \Pr_{(\sigma_1(S),\sigma_2)} \bar{E}_{[0,T(\sigma,\mu,z')-1)}}{\sum_{\omega \in \Omega} \mu(\omega) \Pr_{(\sigma_1(\omega),\sigma_2)} \bar{E}_{[0,T(\sigma,\mu,z')-1)}}$$

$\mu(S|h^{T(\sigma,\mu,z',h)-1}) < z'$, and $\Pr_{(\sigma_1(S),\sigma_2)} \bar{E}_{[0,T(\sigma,\mu,z')-1)} = 1$, by definition. Consequently,

$$\sum_{\omega \in \Omega} \mu(\omega) \Pr_{(\sigma_1(\omega),\sigma_2)} \bar{E}_{[0,T(\sigma,\mu,z')-1)} > \frac{\mu(S)}{z'}.$$

Which implies that

$$\mu(N) \Pr_{(\sigma_1(N),\sigma_2)} E_{[0,T(\sigma,\mu,z')-1)} \leq \sum_{\omega \in \Omega} \mu(\omega) \Pr_{(\sigma_1(\omega),\sigma_2)} E_{[0,T(\sigma,\mu,z')-1)} < 1 - \frac{\mu(S)}{z'}.$$

□

Appendix B. A reputation result with review strategies: omitted proofs.

Proof of Lemma 7. There exists a $\delta^* < 1$ such that

$$\sum_{t=0}^{J_\epsilon - 1} \delta^t g_2(a_1^s, a_2) + \sum_{t=J_\epsilon}^{J_\epsilon + JP - 1} \delta^t g_2(a_1^p, a_2') < -lJ_\epsilon(P+1) \tag{7}$$

for all $\delta > \delta^*$ For public history $h^{t+J_\epsilon-1} = \{a_1^0, y^0, a_1^1, y^1, ..., a_1^{t+J_\epsilon-1}, y^{t+J_\epsilon-1}\}$, let $i(^{t+J_\epsilon-1}) = 1$, if $\sum_{j=t}^{t+J_\epsilon-1} \delta^{j-t} g_1(a_1^j, y^j) < 1 - \eta$ and period t is the start of a review stage; and $i(h^t) = 0$, otherwise. If $i(^{t+J_\epsilon-1}) = 1$, then player 1 receives at least zero in period t through period $t + J_\epsilon + J_\epsilon P - 1$. Consequently, $U_1(\sigma_1(D_\epsilon), \sigma_2, \delta) \geq 1 - \eta - J_\epsilon(1+P)(1-\delta)\mathbb{E}_{(\sigma_1(D_\epsilon),\sigma_2)}[\sum_{t=0}^{\infty} \delta^t i(h^t)]$. By construction $\eta < \epsilon$ and so $(1-\delta)\mathbb{E}_{(\sigma_1(D_\epsilon),\sigma_2)}[\sum_{t=0}^{\infty} \delta^t i(h^t)] \geq r/J_\epsilon(1+P)$. If $i(h^{t+J_\epsilon-1}) = 1$, then player 2 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 $\delta > \delta^*$ by equation (7). In any block where player 1 receives at least $1 - \eta$, player 2 receives at most $\rho\eta < \rho\epsilon$. Consequently, $U_2(\sigma_1(D_\epsilon), \sigma_2) \leq \epsilon\rho - J_\epsilon(1+P)l(1-\delta)\mathbb{E}_{(\sigma_1(D_\epsilon),\sigma_2)}[\sum_{t=0}^{\infty} \delta^t i(h^t)] \leq \epsilon\rho - lr$, if $\delta > \delta^*$. □

B.1. **Proof of Theorem 3.**

Lemma 8. *Posit Assumption 1 and Assumption 3. Pick any NE σ of $\Gamma^\infty(\mu, \delta)$ and period t public history h^t. Let $\mu'(\cdot) = \mu(\cdot|h^t, \sigma_1(D_\epsilon, h^t), y)$ for any y. Suppose that $a_1 \in supp(\sigma_1(N, h^t))$ and $a_1 \neq \sigma_1(D_\epsilon)$, i.e., player 1 deviates from $\sigma_1(D_\epsilon)$ with positive probability. Then $U_1(\sigma, \delta|h^t, a_1, \sigma_2(h^t)) \geq 1 - \epsilon - R(\mu', \delta) - (1 + J_\epsilon(1+P))M(1-\delta)$. Consequently,*

$$|U_2(\sigma_1(N), \sigma_2|h^t, a_1, a_2)| \leq \frac{\rho}{\underline{\pi}}(R(\mu', \delta) + \epsilon + (1 + J_\epsilon(1+P))M(1-\delta))$$

Proof. Player 1's payoff from playing any $a_1 \neq a_1^s$ is at most

$$(1-\delta)M + U_1(\sigma|h^t, a_1, \sigma_2(h^t)) = (1-\delta)M + \sum_{y \in Y} \pi_y(\sigma_2(h^t))U_1(\sigma|h^t, a_1, y).$$

Player 1 can instead play according to $\sigma_1(D_\epsilon)$. This may include finishing a J_ϵ period review phase and then playing a PJ_ϵ period punishment phase. If player 1 instead plays $\sigma_1(D_\epsilon)$ during the $J_\epsilon + PJ_\epsilon$ periods, then he receives at least zero in these periods and beliefs are updated to $\mu'(\cdot)$. His continuation payoff, afterwards is at least $1-\epsilon-R(\mu',\delta)$. Consequently, for any $a_1 \in supp(\sigma_1(h^t))$,

$$(1-\delta)M + U_1(\sigma|h^t, a_1, \sigma_2(h^t)) \geq 1-\epsilon-R(\mu',\delta) - J_\epsilon(P+1)M(1-\delta)$$
$$U_1(\sigma|h^t, a_1, \sigma_2(h^t)) \geq 1-\epsilon-R(\mu',\delta) - (1+J_\epsilon(P+1))M(1-\delta).$$

The bound on player 2's payoff follows from an identical argument as in Lemma 2. □

Proof of Theorem 3. Choose equilibrium σ such that $r(\sigma_2,\delta) \geq \bar{R}(\xi, z_n, \delta, \phi) - \xi$. Let $T(h) = T(\sigma, \mu, z_{n-1}, h)$. We look at the event that player 1 is the normal type and deviates from $\sigma_1(D_\epsilon)$ for the first time at $t < T-1$ (event $E_{[0,T-1)}$), the event that player 1 is the normal type and deviates from $\sigma_1(D_\epsilon)$ for the first time at $t \geq T-1$ (event $E_{[T-1,\infty)}$), the event that player 1 is type D_ϵ, and the event that player 1 is any other type. Player 2 can receive at most $\rho\epsilon + (1-\delta)MJ_\epsilon$ until the time that player 1 deviates from $\sigma_1(D_\epsilon)$. Using Lemma 7 and 8 and applying an argument similar to Lemma 4 implies that

(8)
$$U_2(\sigma,\delta) \leq \frac{\rho}{\underline{\pi}}(q(\delta,\phi,n,\xi)\bar{R}(\xi,z_n,\delta,\phi) + \epsilon K^{n-1} + 2\epsilon + 2(1-\delta)M(1+J_\epsilon(1+P))$$
$$+2\rho\epsilon + (1-\delta)MJ_\epsilon + (z_n(\delta,\phi)-\xi)(\rho\epsilon - l(\bar{R}(\xi,z_n,\delta,\phi)-\xi)) + M\phi$$

where $q(\delta,\phi,n,\xi) = 1-(z_n(\delta,\phi)-\xi)/z_{n-1}(\delta,\phi)$.

Pick any NE σ of $\Gamma^\infty(\mu,\delta)$. Let σ_2^* denote a strategy that plays a_1^b after any period k public history h^k, if there is no deviation from $\sigma_1(D_\epsilon)$ in h^k, and coincides with NE strategy σ_2 if player 1 has deviated from $\sigma_1(D_\epsilon)$ in h^k. Let strategy profile $\sigma^* = (\{\sigma_1(\omega)\}_{\omega\in\Omega}, \sigma_2^*)$. Let $T(h) = T(\sigma,\mu,z_{n-1},h)$. We again look at the event that player 1 is the normal type and deviates from $\sigma_1(D_\epsilon)$ for the first time at $t < T-1$ (event $E_{[0,T-1)}$), the event that player 1 is the normal type and deviates from $\sigma_1(D_\epsilon)$ for the first time at $t \geq T-1$ (event $E_{[T-1,\infty)}$), the event that player 1 is type D_ϵ, and the event that player 1 is any other type. Player 2's payoff until player 1 deviates from $\sigma_1(D_\epsilon)$ is at least $-2\rho\epsilon$. Consequently, an argument along the lines of Lemma 5 implies that

$$U_2(\sigma^*,\delta) \geq -\frac{\rho}{\underline{\pi}}(q(\delta,\phi,n,\xi)\bar{R}(\xi,z_n,\delta,\phi) + K^{n-1}\epsilon + 2\epsilon + 2(1-\delta)M(1+J_\epsilon(1+P))$$
$$-2\rho\epsilon - z_n(\delta,\phi)\rho\epsilon - M\phi$$

□

B.2. **Proof of Theorem 4.** A commitment type $\omega \in \Omega$ plays according to strategy $\sigma_1^T(\omega)$ in the finitely repeated game $\Gamma^T(\delta, \mu)$. The strategy $\sigma_1^T(\omega)$ is the projection of the infinitely repeated game strategy $\sigma_1(\omega)$. For $\xi > 0$, let $NE_\xi(\Gamma^T(\mu, \delta))$ denote the set of ξ Bayes-Nash equilibria of the finitely repeated game $\Gamma^T(\mu, \delta)$.

Recall that $D_{\epsilon/n}$ is defined recursively and is the type that plays a strategy that coincides with $D_{\epsilon/(n-1)}$ up to time T_{n-1} and then plays a review strategy with accuracy ϵ/n. Theorem 3 implies that if $\mu \in interior\Delta(\Omega \cup \{D_{\epsilon/n}\})$, then there exists $\delta(D_{\epsilon/n})$ such that $U_1(\sigma, \delta) > 1 - 3\epsilon/2n - \max\{\epsilon/n, \mu(\Omega_-)\}K(\mu(D_{\epsilon/n}))$ for any $\delta \geq \delta(D_{\epsilon/n})$ and any $\sigma \in NE(\Gamma^\infty(\mu, \delta))$. We choose $\delta(D_{\epsilon/n}) > \max\{\epsilon/n, \delta_{\epsilon/n}\}$, where $\delta_{\epsilon/n}$ is defined as in Lemma 7, and large enough to ensure that the first T_{n-1} periods have payoff consequence of at most $\epsilon/2n$.

Lemma 9. *Suppose that $[\underline{\delta}, \bar{\delta}] \subset [\delta(D_{\epsilon/n}), 1)$. For every $\xi > 0$ there exists a $T^*(\xi, [\underline{\delta}, \bar{\delta}])$ such that for all $\mu \in interior\Delta(\Omega \cup \{D_{\epsilon/n}\})$, all $\delta \in [\underline{\delta}, \bar{\delta}]$, all $T \geq T^*(\xi, [\underline{\delta}, \bar{\delta}])$ and all $\sigma^T \in NE_\xi(\Gamma^T(\mu, \delta))$, $U_1(\sigma^T, \delta) \geq 1 - 3\epsilon/2n - \max\{\epsilon/n, \mu(\Omega_-)\}K(\mu(D_{\epsilon/n})) - \xi$.*

Proof. Fudenberg and Levine (1983, 1986) prove the following theorem: suppose that $\sigma^T \in NE_{\xi(T)}(\Gamma^T(\mu, \delta))$, $\lim_{T\to\infty} \xi(T) = 0$ and $\lim_{T\to\infty} \sigma^T = \sigma$, then $\sigma \in NE(\Gamma^\infty(\mu, \delta))$. The lemma is an immediate consequence of this theorem. If the lemma was not true then we could pick a $\xi > 0$ and, for each T, we could pick a $\delta^T \in [\underline{\delta}, \bar{\delta}]$, $\mu^T \in interior\Delta(\Omega \cup \{D_{\epsilon/n}\})$, and a $\sigma^T \in NE_\xi(\Gamma^T(\mu^T, \delta^T))$ such that $U_1(\sigma^T, \delta^T) < 1 - 3\epsilon/2n - \max\{\epsilon/n, \mu(\Omega_-)\}K(\mu(D_{\epsilon/n})) - \xi$. Taking the limits $\delta^T \to \delta \in [\underline{\delta}, \bar{\delta}]$, $\mu^T \to \mu \in \Delta(\Omega \cup \{D_{\epsilon/n}\})$, and $\sigma^T \to \sigma \in NE(\Gamma^\infty(\delta, \mu))$. If $\lim_T \mu^T(D_{\epsilon/n}) = \mu(D_{\epsilon/n}) = 0$, then $\lim_T U_1(\sigma^T, \delta^T) = U_1(\sigma, \delta) = -\infty$ which is a contradiction. If $\mu(D_{\epsilon/n}) > 0$, then $U_1(\sigma, \delta) \leq 1 - 3\epsilon/2n - \max\{\epsilon/n, \mu(\Omega_-)\}K(\mu(D_{\epsilon/n})) - \xi$ contradicting Theorem 3. □

Type D_∞ first plays a review strategy with accuracy ϵ for T_1 periods, then plays a review strategy with accuracy $\epsilon/2$ for T_2 periods, and then plays a review strategy with accuracy ϵ/n for T_n periods, and so on. The development below uses Lemma 9 to show that $T_1, T_2, ..., T_n, ...$ can be picked appropriately.

Given a precision level ϵ pick interval $[\underline{\delta}_1, \bar{\delta}_1]$ such that $\delta_\epsilon < \underline{\delta}_1$ and $\delta_{\epsilon/2} < \bar{\delta}_1$ where δ_ϵ and $\delta_{\epsilon/2}$ are as defined in Lemma 7. Lemma 9 implies that there exists an integer

T^* such that for all $\delta \in [\underline{\delta}_1, \bar{\delta}_1]$, all $T \geq T^*$, all $\mu \in interior\Delta(\Omega \cup \{D_\epsilon\})$ and any $\sigma \in NE_{\epsilon/2}(\Gamma^T(\mu, \delta))$

$$U_1(\sigma_1^T, \sigma_2^T, \delta) > 1 - \epsilon - K(\mu(D_\epsilon)) \max\{\epsilon, \mu(\Omega_-)\} - 1/2\epsilon.$$

We pick T_1 so that $T_1 \geq T^*$ and $\bar{\delta}_1^{T_1} < \epsilon/2$. Consequently, for all $\delta \in [\underline{\delta}_1, \bar{\delta}_1]$, all $\mu \in interior\Delta(\Omega \cup \{D_\infty\})$ and any σ that is a NE profile of $\Gamma^\infty(\mu, \delta)$

$$U_1(\sigma, \delta) > 1 - \epsilon - K(\mu(D_\infty)) \max\{\epsilon, \mu(\Omega_-)\} - 1/2\epsilon - 1/2\epsilon$$
$$U_1(\sigma, \delta) > 1 - 2\epsilon - K(\mu(D_\infty)) \max\{\epsilon, \mu(\Omega_-)\}.$$

Notice, these statements hold regardless of what type D_∞ plays after period T_1.

Let $D_{\epsilon/2}$ be the type that first play a review strategy with accuracy ϵ for T_1 periods and then plays a review strategy with accuracy $\epsilon/2$. Pick $\underline{\delta}_2$ such that $\underline{\delta}_2 > \bar{\delta}_1$ and $1 - \underline{\delta}_2^{T_1} < (\epsilon/2)1/3$, i.e., the payoff impact of the first T_1 periods is less than $(\epsilon/2)1/3$ for discount factors that exceed $\underline{\delta}_2$. Also, pick $\bar{\delta}_2 > \underline{\delta}_2$ and $\bar{\delta}_2 > \delta_{\epsilon/3}$. Lemma 9 implies that there exists an integer T^{**} such that:

(*i*) For all $\delta \in [\underline{\delta}_1, \bar{\delta}_2]$, all $T \geq T^{**} + T_1$, all $\mu \in interior\Delta(\Omega \cup \{D_{\epsilon/2}\})$ and any $\sigma^T \in NE_{\epsilon/2}(\Gamma^T(\mu, \delta))$

$$U_1(\sigma^T, \delta) > 1 - \epsilon - K(\mu(D_{\epsilon/2})) \max\{\epsilon, \mu(\Omega_-)\} - \epsilon/2;$$

(*ii*) For all $\delta \in [\underline{\delta}_2, \bar{\delta}_2]$, all $T \geq T^{**} + T_1$, all $\mu \in interior\Delta(\Omega \cup \{D_{\epsilon/2}\})$ and all $\sigma^T \in NE_{\epsilon/6}(\Gamma^T(\mu, \delta))$

$$U_1(\sigma^T, \delta) > 1 - \epsilon/2 - K(\mu(D_{\epsilon/2})) \max\{\epsilon/2, \mu(\Omega_-)\} - 2\epsilon/6.$$

We pick T_2 so that $T_2 \geq T^{**}$ and $\bar{\delta}_2^{T_2} < \epsilon/6$. Consequently, for all $\delta \in [\underline{\delta}_1, \bar{\delta}_2]$, all $\mu \in interior\Delta(\Omega \cup \{D_\infty\})$ and all σ_2^T that is part of a NE profile of $\Gamma^\infty(\mu, \delta)$

$$U_1(\sigma, \delta) > 1 - 2\epsilon - K(\mu(D_\infty)) \max\{\epsilon, \mu(\Omega_-)\}.$$

Also, for all $\delta \in [\underline{\delta}_2, \bar{\delta}_2]$, all $\mu \in interior\Delta(\Omega \cup \{D_\infty\})$ and all σ that is a NE profile of $\Gamma^\infty(\mu, \delta)$

$$U_1(\sigma, \delta) > 1 - \epsilon - K(\mu(D_\infty)) \max\{\epsilon/2, \mu(\Omega_-)\}.$$

We can precede exactly as we did for T_1 and T_2 to pick the remaining $T_3, ..., T_n, ...$ to ensure that for all $\delta \in [\underline{\delta}_{n-1}, \bar{\delta}_n]$, all $\mu \in interior\Delta(\Omega \cup \{D_\infty\})$ and all NE profiles σ of $\Gamma^\infty(\mu, \delta)$

$$U_1(\sigma, \delta) > 1 - \frac{2\epsilon}{n-1} - K(\mu(D_\infty)) \max\{\frac{\epsilon}{n-1}, \mu(\Omega_-)\}.$$

Also, for all $\delta \in [\underline{\delta}_n, \bar{\delta}_n]$, all $\mu \in interior\Delta(\Omega \cup \{D_\infty\})$ and all NE profiles σ of $\Gamma^\infty(\mu, \delta)$

$$U_1(\sigma, \delta) > 1 - \frac{2\epsilon}{n} - K(\mu(D_\infty)) \max\{\epsilon/n, \mu(\Omega_-)\}.$$

.

This implies that for any $\mu \in interior\Delta(\Omega \cup \{D_\infty\})$ and any NE profile σ of $\Gamma^\infty(\mu, \delta)$

$$\lim_{\delta \to} U_1(\sigma, \delta) \geq 1 - K(\mu(D_\infty))\mu(\Omega_-).$$

This implies that $U_1(\sigma, \delta)$ can be made arbitrarily close to one if δ is large and $\mu(\Omega_-)$. Suppose that type D_∞ plays each pure action of player 1 at the start of each review stage before playing a_1^s for the required number of periods. This play will not affect any of the construction so far. However, a learning result analogous to Lemma 1 implies that for any ϕ there exists a T such that $\Pr_{(\sigma_1(D_\infty),\sigma_2)}\{h : \frac{\mu(\Omega_-|h_T)}{\mu(D_\infty|h_T)} < \phi\} > 1 - \phi$, for any strategy σ_2 of player 2. In particular, if player 1 plays according to D_∞ we can reject the hypothesis, with any required level of accuracy ϕ, by some time $T(\phi)$, that the strategy player 1 is playing has fewer states than the most complicated finite automaton in the finite set Ω_-. Consequently, we can get arbitrarily close to one by first manipulating player 2's beliefs and then using the payoff bound.

References

Aoyagi, M. (1996): "Reputation and dynamic Stackelberg leadership in infinitely repeated games," *Journal of Economic Theory*, 71(2), 378–393.

Atakan, A., and M. Ekmekci (2008): "Reputation with Long Run Players," *Working Paper*.

Billingsley, P. (1995): *Probability and Measure*. John Wiley and Sons, New York, NY.

Celantani, 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., and J. Thomas (1997): "Reputation and perfection in repeated common interest games," *Games and Economic Behavior*, 18, 141–158.

Evans, R., and J. Thomas (1997): "Reputation and Experimentation in Repeated Games With Two Long-Run Players," *Econometrica*, 65(5), 1153–1173.

Fudenberg, D., and D. Levine (1983): "Subgame-perfect equilibria of finite- and infinite-horizon games.," *Journal of Economic Theory*, 31(2), 251–268.

——— (1986): "Limit games and limit equilibria.," *Journal of Economic Theory*, 38(2), 261–279.

Fudenberg, D., and D. Levine (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.

Kreps, D., and R. Wilson (1982): "Reputation and Imperfect Information," *Journal of Economic Theory*, 27, 253–279.

Mailath, G., and L. Samuelson (2006): *Repeated games and reputations: long-run relationships.* Oxford University Press, USA.

Milgrom, P., and J. Roberts (1982): "Predation, Reputation, and Entry Deterrence," *Journal of Economic Theory*, 27, 280–312.

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

——— (1985): "Repeated principal-agent games with discounting," *Econometrica: Journal of the Econometric Society*, pp. 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.

MEDS, Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208

E-mail address: a-atakan@kellogg.northwestern.edu

E-mail address: m-ekmekci@kellogg.northwestern.edu