Reputation Games with Finite Automata

Mehmet Ekmekci [*]and Andrea Wilson[†]

October 23, 2007

Abstract

This paper studies reputation effects in a 2-player repeated moral hazard game. A long-lived player, Player 1, would benefit if he could commit to playing a particular action which is strictly dominated in the stage game. His opponent, who may be either long-lived or myopic, believes there is a small probability that player 1 is a commitment type, and each period observes only a noisy signal about player 1's action. We depart from the standard literature by assuming that player 2 has finite memory: he is restricted to use a finite automaton, both to carry out his own strategy, and to update his beliefs about player 1's strategy. We show that this restriction enables player 1 to permanently maintain a reputation as a commitment type (in contrast to Cripps, Mailath, Samuelson's result for unbounded players, which showed that under imperfect monitoring, reputation effects are only temporary). However this relies on player 2 having a sufficiently large memory, and there are also equilibria in which player 1 does not build a reputation. The final section of the paper shows that if *both* players are patient and memory-constrained, in that they find less complex strategies less costly, then there is a lower bound on player 1's payoff which converges to his commitment payoff as monitoring becomes perfect.

1 Introduction

This paper studies reputation effects in a repeated moral hazard game with imperfect monitoring. For example, consider the following stage game:

		Player 2	
		L	R
Player 1	G	$(1,1)$	$(-1,0)$
	B	$(2,-1)$	$(0,0)$

[*] Northwestern MEDS; m-ekmekci@kellogg.northwestern.edu

[†] New York University, Economics Department; andrea.wilson@nyu.edu

Here G is a strictly dominated action for player 1, but both players would benefit if he could commit to playing G with probability at least $\frac{1}{2}$.

Fudenberg and Levine's (1989) original reputation result looked at the perturbation of this game, in which there is a small probability that player 1 is a commitment type who always plays G. They showed that if player 1 is sufficiently patient, then his expected payoff against a myopic opponent must be arbitrarily close to 1 (the commitment payoff) in any NE of the repeated game with incomplete information. Their 1992 paper shows that this result is robust to imperfect monitoring in an ex ante sense: for fixed but sufficiently high $\delta_1 < 1$, player 1's average payoff calculated at the beginning of the game is very close to 1, even when his actions are imperfectly observed by player 2.

More recently, Cripps, Mailath, and Samuelson (2004) showed that this reputation result is nevertheless a short-run phenomenon when monitoring is imperfect: player 2 eventually learns player 1's true type with probability 1, and hence play eventually converges almost surely to an equilibrium of the game with complete information. The intuition is that once player 1 successfully builds a reputation as a commitment type, player 2's optimal strategy must become almost unresponsive to new signals about player 1's actions. Once player 1 expects continuation play to be almost independent of his action choice, he strictly prefers to play his dominant action, B: hence, he must eventually deviate from the commitment strategy frequently enough that, under some identification assumptions, player 2 will be able to learn player 1's true type with probability 1.Therefore, in any equilibrium, player 2 eventually learns that he is facing a normal type and reputation effects collapse.

Ekmekci (2006) showed how it is possible to restore reputation effects by restricting the information observed by player 2. He studied "rating systems": rather than seeing the entire history of signals, the short-run players are informed only about the average frequency with which player 1 chose the good action. There are a finite number of possible ratings, which are updated and published by an external agency.

In this paper, we study whether it is possible for player 1 to develop a permanent reputation when his opponent has finite memory. Following Wilson (2005), we model this by restricting player 2 to strategies which can be implemented by a finite automaton: player 2 observes each signal about player 1's play, but cannot remember the entire sequence, and must instead use his automaton to optimally keep track of information. As in Ekmekci (2006), this implies a finite number of possible "ratings"; the difference is that player 2 designs the rating grid himself, and optimally updates it as he observes new information. More precisely: each state in the automaton can be identified with a belief about player 1's strategy, and about the history to date. A strategy for player 2 specifies an action for each state in the automaton, together with a transition rule, which specifies how the "rating" is updated in response to each new signal about player 1's strategy.

We study the long-run (stationary) equilibria of repeated moral hazard games in which (i) player 1 is a simple (Stackelberg) commitment type with probability $\pi > 0$; (ii) player 1's actions are imperfectly observed; (iii) player 2 uses a finite automaton to implement his strategy and update his beliefs, and can increase the number of states in his automaton at a cost. This is similar to the model in Aumann and Sorin (1989); the difference is that they restricted attention to deterministic automata and considered pure common interest games.

Our first main result, Proposition 2, shows that permanent reputations are possible with bounded memory: if player 2 has sufficiently many memory states, then there is an equilibrium in which, after every history, player 1's expected continuation payoff is equal to his *maximal commitment payoff* - that is, the payoff he could obtain by publicly committing to his favorite strategy (even if this differs from the strategy of the commitment type: in the example above, the maximal commitment payoff is $\frac{3}{2}$). Moreover, the equilibrium holds for any discount factor δ_2 for player 2. Unfortunately, this is only a "possibility result", there are also many equilibria in which player 1 earns a lower payoff. The final section of the paper looks for conditions which guarantee a high average payoff for player 1. Proposition 3 assumes that complexity costs are positive for *both* players; so player 1 is also constrained to choose a strategy which can be implemented by a finite-state automaton, and incurs a complexity cost which is increasing in the size of his chosen automaton. We show with the constraint on player 1, a reputation result obtains in the limit as monitoring becomes almost perfect: if player 1 is sufficiently patient, and complexity costs are positive but sufficiently small for both players, then in any equilibrium, and with any discount factor for player 2, the lower bound on player 1's average payoff approaches his commitment payoff as signal noise vanishes to zero. This reputation result does not hold when player 1 is unbounded: that is, if complexity is costly only for player 2, then there are equilibria in which player 1's average payoff is close to zero even when monitoring is almost perfect. We show additionally in Proposition 4 that for the model with complexity costs, almost-perfect monitoring implies existence of a pure-strategy equilibrium.

2 Model

Two players interact repeatedly in a moral hazard game, with (expected) stage game payoffs as shown below:

		Player 2	
		L	R
Player 1	G	$(u_L - e, 1)$	$(-e, 0)$
	B	$(u_L, -1)$	$(0, 0)$

with $u_L - e > 0$. The unique Nash equilibrium of the stage game is (B,R), as B is a dominant strategy for player 1 (playing the "good action" G incurs an effort cost e regardless of player 2's action), while player 2 will choose action R unless he expects his opponent to play G with probability at least $\frac{1}{2}$. However, since the gain u_L in player 1's payoff when his opponent plays L exceeds the effort cost e, both players would benefit if player 1 could commit to playing G with probability at least $\frac{1}{2}$. For future reference, define player 1's *maximal commitment payoff* as the average payoff $(u_L - \frac{1}{2}e)$ he would obtain under his optimal commitment strategy, which is to play G w.p. $\frac{1}{2}$ every period.

Both players are long-lived, and discount the future at rates δ_1, δ_2. It is assumed throughout the paper that player 1 wishes to maximize the limit, as $\delta_1 \to 1$, of his δ_1-discounted payoff.

To allow for reputation effects, we assume that there are two possible types for player 1, $\{\mathcal{N}, \mathcal{C}\}$. Type $\mathcal{N}$, the "normal type", is a standard strategic agent with payoffs as described above. Type $\mathcal{C}$ is the "commitment" type:

Assumption 1: With probability $\pi > 0$, Player 1 is a simple commitment type who plays G with probability 1 after every history.

This is the reputation game studied in Fudenberg and Levine (1989), who established that for δ_2 near 0, and δ_1 below but sufficiently close to 1, Player 1 earns almost his commitment payoff $(u_L - e)$ in any Nash equilibrium of the repeated game.

2.1 Imperfect Monitoring

Each period, player i's action is observed correctly with probability ρ_i, and incorrectly with probability $1 - \rho_i$, where $\rho_i \in \left(\frac{1}{2}, 1\right)$. The signal is public, and we assume up to Section 4 that player 2's actions are observed correctly with probability 1. Let $Y \equiv \{gl, gr, bl, br\}$ denote the set of possible signal realizations;[1] where, for example, the signal gl is observed with probability

[1] We include player 2's actions in the public signal for convenience, and as section 4 will assume a small amount of noise for both players.

ρ_1 if players 1,2 played (G,L), and with probability $1-\rho_1$ after (B,L).

One interpretation of this model is that payoffs depend stochastically on action choices (so that the payoffs in the stage game are expected payoffs), and that realized payoffs are the publicly observed signals.

Note that this structure satisfies the typical identification assumption: with sufficiently many observations, Player 2 would be able to identify any fixed stage game strategy of Player 1.

By methods from Abreu-Pearce-Stachetti (1990): if Player 2 is myopic, this information structure reduces Player 1's average equilibrium payoff in the complete-information repeated game to at most $(u_L - e) - e\left(\frac{1-\rho}{2\rho-1}\right)$ (for the range where this is positive).

In their (2004) paper, Cripps-Mailath-Samuelson showed that reputation is a short-run phenomenon under imperfect monitoring. Their result implies that in any Nash equilibrium of the game with $\pi > 0$, and for any $\rho \in (\frac{1}{2}, 1)$, Player 2 eventually learns Player 1's true type with probability 1, and hence play eventually converges almost surely to an equilibrium of the repeated game with complete information ($\pi = 0$).

In this paper, we study the sustainability of reputation effects when Player 2's memory is finite. We follow Wilson (2005) in defining a finite-memory strategy as one which can be implemented by a finite-state, non-deterministic automaton, but here allow the states to be chosen at some cost:

Assumption: Player i uses a (possibly non-deterministic) automaton to carry out his strategy.

The cost of an N-state automaton is $c_i \cdot N$.

2.2 Strategies and Equilibrium

A behavior strategy for an unbounded player 1 ($c_1 = 0$) is a map $\gamma_1 : \cup_{t=0}^{\infty} H_1^t \to \Delta(\{G, B\})$, where H_1^t is the set of t-period private histories for player 1:

$$H_1^t = \{(a_1^0, y^0), (a_1^1, y^1), ..., (a_1^t, y^t)\}$$

where a_1^t is player 1's realized action choice in period t, and $y^t \in \{gl, gr, bl, br\}$ is the signal realization in period t (which includes player 2's action).

We model Player 2 as an N_2-state, stationary, non-deterministic automaton.[2] A strategy for Player 2 is then a triplet $\gamma_2 = (i^0, \sigma, d)$, where:

[2] With the exception of Proposition 2, a strategy for player 2 includes also choosing the number of states in his automaton, given the cost c_2. Proposition 2 holds also for this model, but the present version of the paper assumes for that result that N_2 is given exogenously.

- i^0 is the initial memory state
- $\sigma : \mathcal{N}_2 \times Y \to \Delta(\{1, 2, ..., N_2\})$ is the transition rule, specifying how the memory state is updated after a new piece of information $y \in Y$. For $i, j \in \mathcal{N}$ and $y \in Y$, we will sometimes let $\sigma^y_{i,j} \equiv \sigma(i, y)(j)$ denote the probability of a transition $i \to j$ after signal realization $y \in Y$.
- $d : \mathcal{N}_2 \to [0, 1]$ is the action rule, where $d(i)$ denotes the probability of choosing action L in memory state i

Note that player 2's automaton strategy is required to be *stationary*: every time he is in state $i \in \{1, 2, ..., N_2\}$, he uses the same action and transition rule. The interpretation is that player 2's memory state represents all of the information available to him; he can use this information, and understanding of the rule σ, to make inferences about the history, but cannot recall exactly which history he has observed.

A strategy for a bounded player 1 ($c_1 > 0$) is a choice on the number of automaton states N_1, together with a strategy triplet γ_1 as defined for player 2.

We will focus explicitly on equilibrium steady states, and restrict attention to irreducible automata (which is implied by equilibrium for $c_1, c_2 > 0$, provided that players put a sufficiently high weight on the future).

More precisely: let $f^{\mathcal{C}}(i)$ denote the steady-state probability that player 2 is in memory state $i \in \{1, ..., N_2\}$, conditional on player 1 being type $\mathcal{C}$; and let $f^{\mathcal{N}}(i)$ denote the steady-state probability of state i, conditional on player 1 being type $\mathcal{N}$ and playing γ_1. Then when $\rho_2 = 1$, and suppressing the subscript on ρ_1, $f^{\mathcal{C}}$ is the solution to the following $N_2 \times N_2$ system of equations:

$$\forall i \in \mathcal{N}_2 : f^{\mathcal{C}}(i) = \sum_{j \in \mathcal{N}_2} f^{\mathcal{C}}(j) \left[d(j) \left(\rho\sigma^{gl}_{j,i} + (1-\rho)\sigma^{bl}_{j,i} \right) + (1 - d(j)) \left(\rho\sigma^{gr}_{j,i} + (1-\rho)\sigma^{br}_{j,i} \right) \right]$$

The distribution conditional on type $\mathcal{N}$ is the solution to a similar system of equations, provided that these probabilities exist (again implied if both players follow automaton strategies):

$$\forall i \in \mathcal{N}_2 : f^{\mathcal{N}}(i) = \sum_{j \in \mathcal{N}_2} f^{\mathcal{N}}(j) \left[d(j) \left(p_j\sigma^{gl}_{j,i} + (1-p_j)\sigma^{bl}_{j,i} \right) + (1 - d(j)) \left(p_j\sigma^{gr}_{j,i} + (1-p_j)\sigma^{br}_{j,i} \right) \right]$$

where p_i is the probability (long-run frequency) of a g-signal when Player 2 is in state i, conditional on type $\mathcal{N}$.

As an example for how these are calculated: let γ_1 be the strategy for player 1 (type $\mathcal{N}$) which specifies playing G w.p. 1 if the last signal was gl or gr, and B w.p. 1 after bl, br; and let γ_2 be the strategy for player 2 on $\mathcal{N}_2 \equiv \{1, 2\}$ which specifies the following transition rule:

- in state 1: stay in state 1 after bl, br; move to state 2 with probability σ after gl, gr
- in state 2: stay in state 2 after gl, gr; move to state 1 with probability 1 after bl, br

To calculate player 2's beliefs: in state 2, he expects a type $\mathcal{N}$ opponent to play G w.p. 1 (as he can infer from being in state 2 that the last signal was gl or gr, at which point player 1 will play G under γ_1), and hence $p_2 = \rho$. To calculate p_1 : let $f^{\mathcal{N}}(G1), f^{\mathcal{N}}(B1)$ denote the long-run frequencies with which a type $\mathcal{N}$ opponent plays G,B when player 2 is in state 1. These solve

$$f^{\mathcal{N}}(G1)\,(1-\rho(1-\sigma)) = f^{\mathcal{N}}(B1)(1-\rho)(1-\sigma)$$

(To see this: by definition, $f^{\mathcal{N}}(G1)$ is the steady-state probability of the "pair-state" G1, in which the normal type of player 1 plays G, and player 2 is in memory state 1. The probability of a transition into this pair-state G1 is $\rho(1-\sigma)$ from G1, and $(1-\rho)(1-\sigma)$ from B1 (as they will move from G1 to G1 iff they observe signal gl or gr and player 2 then stays in state 1; given that the normal type is playing G with probability 1 in state G1, this happens with probability $\rho(1-\sigma^g_{1,2})$). There is zero probability of a transition into G1 from the remaining two pair-states in this example, G2 and B2, as both of these states go to B1 (player 1 plays B and player 2 is in state 1) after signals bl, br, and to G2 after gl, gr).

So in state 1, Player 2 expects a normal type of opponent to play G with probability

$$\alpha_1 \equiv \frac{f^{\mathcal{N}}(G1)\cdot(1) + f^{\mathcal{N}}(B1)\cdot(0)}{f^{\mathcal{N}}(G1) + f^{\mathcal{N}}(B1)} = \frac{(1-\rho)(1-\sigma)}{2(1-\rho)+\sigma(2\rho-1)}$$

This implies steady-state probabilities $f^{\mathcal{C}}(2) = \frac{\rho\sigma}{1-\rho+\rho\sigma}, f^{\mathcal{N}}(2) = \frac{p_1\sigma}{1-p_1+p_1\sigma}$, and $f^s(1) = 1 - f^s(2)$ for $s \in \{\mathcal{C}, \mathcal{N}\}$. Finally, in memory state $i \in \{1, 2\}$, Player 2 believes that he is facing a commitment type with probability $\Pr\{\mathcal{C}|i\} = \dfrac{f^{\mathcal{C}}(i)}{\pi f^{\mathcal{C}}(i) + (1-\pi) f^{\mathcal{N}}(i)}$, and expects to observe a g-signal w.p. $\Pr\{\mathcal{C}|i\}\cdot\rho + (1-\Pr\{\mathcal{C}|i\})\cdot p_i$.

Player 1's problem is similar for $c_1 > 0$. If $c_1 = 0$, his problem is standard: define

$$E^{(\gamma_1,\gamma_2)}\left[(1-\delta_1)\sum_{t=\tau}^{\infty}\delta_1^{t-\tau}u_1(a_1^t, a_2^t) \mid \mathcal{H}_1^t\right]$$

as his expected continuation payoff conditional on $\mathcal{H}_1^t$, where $\{\mathcal{H}_1^t\}_{t=1}^{\infty}$ is the filtration on $(A_1 \times Y)^{\infty}$ induced by private histories for player 1, $u_1(\cdot)$ is player 1's stage game payoff function, and expectations are taken with respect to the probability distribution over $(A_1 \times Y)^{\infty}$ induced by (γ_1, γ_2); note that player 1 does not directly observe player 2's memory state. If P1 is unbounded ($c_1 = 0$), say that γ_1 is a *best response to* γ_2 if $V^1(\gamma_1, \gamma_2) \geq V^1(\gamma_1', \gamma_2)$ for all behavior strategies γ_1', where

$$V^1(\gamma_1, \gamma_2) \equiv E^{(\gamma_1,\gamma_2)}\lim_{\delta_1\to 1}\left[(1-\delta_1)\sum_{t=0}^{\infty}\delta_1^t u_1(a_1^t, a_2^t)\right]$$

For Player 2: say that (N_2, γ_2) is a *best response to* γ_1 if:

1. Given the action rule $d_2 : (N_2, i_2^0, \sigma_2)$ maximizes player 2's ex ante average expected payoff,
$$\sum_{i \in \{1,2,..,N_2\}} d_2(i) \left[\pi f^{\mathcal{C}}(i) + (1-\pi) f^{\mathcal{N}}(i) \left(2\alpha_i - 1\right)\right] - c_2 N_2$$
where α_i is the long-run frequency with which type n plays G when player 2 is in memory state i (the corresponding probability of a g-signal is $p_i = 1 - \rho + (2\rho - 1)\alpha_i$), and $d_2(i)$ is the probability that player 2 plays L in memory state i

2. In each state $i \in \{1, 2, .., N_2\}$, $d_2(i)$ maximizes player i's δ_i-discounted expected continuation payoff, given the state-i beliefs about player 1's strategy, and using the continuation payoffs induced by (γ_1, γ_2).

3. Among all rules satisfying conditions 1 and 2, (N_2, γ_2) maximizes player 2's expected payoff.[3]

Definition: An equilibrium of the game with incomplete information is a pair (γ_1^*, γ_2^*) such that γ_i^* is a best response to γ_{-i}^*, for $i \in \{1, 2\}$.

Some comments on this definition: the second condition requires that player 2 always choose an action which is optimal, given his beliefs. For a myopic player (δ_2 near zero), this means playing L whenever he expects his opponent to play G with (total) probability at least $\frac{1}{2}$. For larger values of δ_2, optimality may imply playing L even when the opponent is almost certain to play B, if doing so induces the normal type of player 1 to play a more attractive (expected) continuation strategy.

The first condition says that, given the action rule, the memory (size and transition rule) is chosen to maximize the expected *undiscounted* average payoff. (To guarantee existence of equilibrium, we in fact need to assume that the memory rule maximizes the expectation of
$$(1-\delta) \sum_{t=0}^{\infty} \delta^t \sum_{i \in \mathcal{N}_2} d_2(i) \left[\pi g_i^{t,c} + (1-\pi) g_i^{t,n} (2\alpha_i^t - 1)\right] - c_2 N_2$$
for some $\delta < 1$, where $g_i^{t,c}, g_i^{t,n}$ are the probabilities of being in state $i \in \mathcal{N}_2$ in period t, conditional on type c, n; and α_i^t is the expected probability that the normal type of player 1

[3] This specification implies that the memory transition rule is "hard-wired" at the start (and need not be interim optimal), while actions must maximize the continuation payoff at the time they are chosen. Condition 3 precludes the choice of a strategy such as: play R in all memory states, and follow the memory transition rule which chooses each state with probability $\frac{1}{N_2}$ after every history.

will play G if player 2 is in memory state i in period t; in the limit as $\delta \to 1$, this converges to the expression given above.)[4]

This is the obvious candidate for optimality when player 2 indeed wishes to maximize the limit, as $\delta_2 \to 1$, of his δ_2-discounted average payoff. It is less obviously the right concept when δ_2 is significantly below 1, as it implicitly assumes that a different discount rate is used when designing the memory, than at the time when actions are chosen. One alternative would be to instead look for memory rules which maximize the δ_2-discounted average payoff; the problem with this rule is that the definition of an optimal memory would become meaningless as $\delta_2 \to 0$. Our formulation implies that an agent with $\delta_2 = 0$ will choose myopically optimal action choices, but will store information in a way which is optimal in the long run, and which in particular implies Bayesian behavior in the limit as memory costs (c_2) go to zero.

3 Sustainable Reputations

With finite memory and noisy signals, it is impossible for player 2 to become convinced that he is facing a particular type of player 1: for any $i \in \{1, 2, ..., N_2\}$, and any strategy pair (γ_1, γ_2), the probability that player 2 assigns to the commitment type in state i, $\Pr\{\mathcal{C}|i\}$, is bounded away from both zero and one.

The fact that $\Pr\{\mathcal{C}|i\}$ is bounded above 0 implies that permanent reputations may be possible: in contrast to the Cripps-Mailath-Samuelson result with unbounded players, it is no longer true that an infinite number of deviations by Player 1 from the commitment strategy will lead Player 2 to statistically identify him as the normal type. Hence, provided that π, N_2 are not too low, it is possible to construct an automaton which is optimal for player 2, yet provides player 1 with incentives to play G often enough to maintain a reputation.

The difficulty in doing this is that a finite number of memory states implies also that player $2's$ maximal posterior (on the probability of a commitment type) is bounded below 1, for any strategy pair (γ_1, γ_2). Proposition 1 states that if player 2's memory cost c_2 is sufficiently high, relative to the prior π and informativess of the signal, then he will never find it worthwhile to learn enough information for player 1 to benefit from reputation-building: (i) there is an equilibrium in which player 1 earns average payoff zero ((B,R) is played every period) if and only if c_2 is above a cutoff $\overline{c}_2(\rho, \pi)$; (ii) if $\delta_2 = 0$ (player 2 chooses myopically optimal action

[4] For example: if the normal type of player 1 plays B w.p. 1 every period, independently of the history, then an optimal memory for player 2 will involve at least one transition probability which is proportional to $\sqrt{1-\delta}$. There is a discontinuity in the payoff at the point where this transition probability hits zero, which implies that if the normal type of player 1 plays this strategy, an optimal memory does not exist for player 2 at $\delta = 1$.

choices), then this cutoff $\overline{c}_2(\rho, \pi)$ also implies that there is no equilibrium in which player 1 earns more than $(u_L - e) - e\left(\frac{1-\rho}{2\rho-1}\right)$, his maximum payoff for the complete-information game. The cutoff $\overline{c}_2(\rho, \pi)$ is strictly increasing in ρ : a more informative signal increases the value of memory states for player 2, hence it will require a higher cost for him to choose not to learn anything. For a similar reason, the cutoff is increasing in the prior π.

Proposition 1: For any $\rho \in \left(\frac{1}{2}, 1\right), \pi \in \left(0, \frac{1}{2}\right)$, and $c_1 \geq 0$, there is a cutoff $\overline{c}(\rho, \pi)$, strictly increasing in both parameters, such that:

i. If $c_2 \geq \overline{c}(\rho, \pi)$, then there is an equilibrium in which player 1 earns an average payoff v^1 if and only if $v^1 \in [0, u_L - e - \frac{e(1-\rho)}{2\rho-1}]$.

ii. If $c_1 > 0$ and $c_2 < \overline{c}(\rho, \pi)$, then in any equilibrium, player 2's average expected payoff is bounded above zero. Moreover: for any $\Delta > 0$, there exist $\underline{c_2}, \rho^*$ s.t. if $c_2 < \underline{c_2}$ or $\rho > \rho^*$, then P2's average payoff satisfies $v^2 \geq \pi - \Delta$.

The proof is in the appendix. We first calculate the cutoff (which turns out to be $\overline{c}(\rho, \pi)$) such that if the normal type of player 1 plays B w.p. 1 in every period, then it is a best response for player 2 to play R w.p. 1 in every period if and only if $c_2 \geq \overline{c}(\rho, \pi)$; against this strategy, it is clearly optimal for the normal type of player 1 to play B every period, yielding an equilibrium with average payoff 0. We then use this result to show that if $c_1 > 0$ (so that player 1 also follows an automaton strategy), then player 2's equilibrium payoff must be bounded above zero whenever his memory cost is below the cutoff. For the remaining assertion: we show that if $c_2 > \overline{c}(\rho, \pi)$, then in any equilibrium player 2 must believe that his opponent is the normal type with probability greater than $\frac{1}{2}$; for $\delta_2 = 0$ this means that he will only play L when he expects the normal type to play G; this does not give player 1 enough incentive to play G, as there is no reward phase where he gets to "ride off his reputation".

3.1 Upper Payoff Bounds

Proposition 1 showed that having a bounded-memory opponent can have a negative effect: it is impossible for player 1 to benefit (in the long run) from the uncertainty about his type if there is too much of a constraint on his opponent's memory. This section describes a positive result: Proposition 2 states that if there is a constraint on player 2's memory, but this constraint is sufficiently small, then there is an equilibrium in which an unbounded player 1 earns (on average) his maximal commitment payoff after every history:

Proposition 2: For any $\pi > 0, \rho > \frac{3}{4}$,[5] and $\varepsilon > 0$: there exists N_2^* such that whenever $N_2 \geq N_2^*$, there is an equilibrium in which player 1's average expected payoff is at least $u_L - \frac{1}{2}e - \varepsilon$ after every history.

Note that this result relies on player 2 having a finite but sufficiently large number of automaton states, and on the signal not being too noisy. Player 1 may be either unbounded ($c_1 = 0$), or a bounded-memory player with a sufficiently small memory cost; and the equilibrium holds for all discount factors δ_2 for player 2.

3.1.1 Sketch of proof of Proposition 2

The idea is to construct an automaton strategy for player 2 which has a temporary "punishment phase" (reached after a large number of b-signals), a temporary "reward phase" (reached after a large number of g-signals), and transitions in between which are responsive enough to make player 1 willing to play G, but indifferent after most histories.

To make this automaton optimal for player 2: we further specify an transition/action rule which does *almost* as well as possible conditional on player 1 being a commitment type,[6] and which is close to being the most informative automaton against a stationary strategy by player 1. For the few transitions which are suboptimal in terms of learning (but required to provide the desired incentives for player 1), we make them optimal by choosing a strategy for player 1 which essentially "uses up" player 2's states: rather than using them to learn, he uses them to catch particular signal sequences, so that his strategy links up in an optimal way with that of the normal type of player 1 (ie, he plays L when he expects the normal type to play G). Finally, we choose an automaton with deterministic transitions, so that player 1 always has a probability-1 belief on player 2's automaton state; this means that he can condition his strategy on player 2's state, and detect (and in principle punish) most deviations.

Consider first the following automaton for player 2:

- In any state $i \notin \{1, N_2 - 1, N_2\}$: Player 2 plays L w.p. 1, moves up $i \to i + 1$ w.p. 1 after a g-signal (gl or gr), and down $i \to i - 1$ w.p. 1 after a b-signal.

[5] The condition on ρ is stronger than necessary, but required for the specific algorithm below. This result also holds if P1 is restricted to use an automaton strategy, provided that he has at least $N_2^* + 1$ memory states.

[6] This is required when ρ is close to 1, or when player 2 has a large number of available memory states: by Proposition 1, either condition implies that if player 1 follows a stationary strategy, then there is an automaton strategy for player 2 which guarantees an expected payoff arbitrarily close to π. So, to make player 2 follow a strategy which does poorly conditional on a commitment type, we need to reward him conditional on the normal type; clearly this is impossible in an equilibrium where player 1 earns on average $\frac{3}{2}$.

- In state N_2 : player 2 plays G w.p. $1-\frac{e\left(\frac{1}{2}-\frac{1-\rho}{2\rho-1}\right)}{2u_L}$, stays after b, and moves to state N_2-3 after g

- In state N_2-1 : player 2 plays G w.p. 1, moves to state N_2 after b, and moves to state N_2-3 after g

- In state 1: Player 2 plays L w.p. $\frac{1}{u_L}\left(u_L-\frac{1}{2}e-e\frac{(1-\rho)}{2\rho-1}\right)$, stays in state 1 after a b-signal, and moves to state 3 after a g-signal.

To verify that this works for player 1: define V_i^1 as player 1's expected continuation payoff, conditional on knowing that player 2 is currently in state $i\in\{1,2,...,N_2\}$. With the automaton constructed above, and supposing that B is always an optimal action for player 1, these satisfy:

$$i\notin\{1,N_2-1,N_2\}:V_i^1=(1-\delta)u_L+\delta(1-\rho)V_{i+1}^1+\delta\rho V_{i-1}^1$$
$$\Rightarrow\rho\lim_{\delta\to1}\frac{V_i^1-V_{i-1}^1}{1-\delta}=u_L-\lim_{\delta\to1}V_i^1+(1-\rho)\lim_{\delta\to1}\frac{V_{i+1}^1-V_i^1}{1-\delta}$$

Using a similar calculation for the corner states:

$$\lim_{\delta\to1}\frac{V_3^1-V_1^1}{1-\delta}=\left[\lim_{\delta\to1}V^1-\left(u_L-\frac{1}{2}e\right)\right]+\frac{e}{2\rho-1}$$
$$\lim_{\delta\to1}\frac{V_{N_2}^1-V_{N_2-3}^1}{1-\delta}=\left[\frac{\left(u_L-\frac{1}{2}e\right)-\lim_{\delta\to1}V^1}{1-\rho}\right]+\frac{e}{2\rho-1}$$

Solving, we find that player 1 earns his maximal commitment payoff, i.e. $\lim_{\delta\to1}V^1=u_L-\frac{1}{2}e$, if and only if in every state $i\in\{1,..,N_2-2\}$, we have $(2\rho-1)\lim_{\delta\to1}\frac{V_{i+1}^1-V_{i-1}^1}{1-\delta}=e$; this is precisely the condition for player 1 to be indifferent between playing G,B when player 2 is in state i (except for state 1, where the condition is $\frac{V_3^1-V_1^1}{1-\delta}=\frac{e}{2\rho-1}$), as the LHS is the increase in the continuation payoff from playing G, while the RHS is the cost. In states N_2-1,N_2, player 1 strictly prefers to play B.

Now consider the following strategy for player 1: choose an N_2-state automaton strategy and follow the same transitions as prescribed above for player 2. As long as no deviations are detected by player 2, play G w.p. α_i in state i (the α_i's to be determined). If a deviation is detected (ie, player 2 plays R in a state $i\notin\{1,N_2-1,N_2\}$), switch permanently to a strategy which plays B w.p. 1 after every history. Note that by construction, any such strategy is optimal for player 1 provided that $\alpha_{N_2-1}=\alpha_{N_2}=0$.[7]

[7] It is clearly optimal if player 1 is either unbounded, or an automaton with at least N_2 (exogenously specified) states. If player 1 chooses his automaton size at a cost $c_1>0$, this is not an equilibrium, as there is no state in which player 1 has a strict incentive to play G. (Therefore, with $c_1>0$, he should deviate to the least costly best response, namely a one-state automaton that always plays B).

Now, for player 2: observe that he plays L w.p. 1 in every state except for $1, N_2$: this implies that his payoff conditional on the commitment type is very close to 1, and goes to 1 as $\rho \to 1$ (since both of states $1, N_2$ are only reachable after b-signals, which have zero probability in the limit as $\rho \to 1$ conditional on a commitment type). Moreover, except for the corner states, he moves to a state with a higher posterior (probability of the commitment type) after evidence of the commitment type (g-signals), and to a state with a lower posterior after evidence of the normal type; as shown in Wilson (2005), this is the optimal automaton for learning player 1's type, assuming a stationary strategy by player 1.

To make the action choices myopically optimal for player 2 in states $\{2, ..., N_2 - 2\}$ (which will imply action optimality at any δ_2), it is sufficient to specify that player 1 plays G with probability $\alpha_i \geq \frac{1}{2}$ in these states (which he is willing and able to do, as by construction he always knows player 2's state, and is indifferent between playing G,B in these states). To make the action choices optimal in the "corner states", we further choose a strategy for player 1 such that if player 2 follows the above automaton, then in state N_2, the probability of a commitment type is exactly $\frac{1}{2}$ (as player 2 is supposed to randomize his action choice in this state, while he believes that a normal type of player 1 is playing B with probability 1). This then implies that he strictly prefers to play G in state $N_2 - 1$, where the probability of a commitment type is slightly higher than $\frac{1}{2}$. This is possible iff N_2 is sufficiently high, as we need a large enough number of states for player 2's posterior to increase to $\frac{1}{2}$, when the normal type of player 1 is playing G on average with probability $\frac{1}{2}$. If N_2 increases above this point, we can reduce the informativeness of the signals by moving player 1's strategy closer to that of a commitment type (ie, by increasing the $\alpha_i's$ closer to 1); this does not affect player 1's payoff, because while it reduces his payoff in the interior states, it also increases the fraction of time spent in states $N_2, N_2 - 1$ (where player 1 rides off his reputation and earns strictly above the commitment payoff). We also need to choose the $\alpha_i's$ such that in state 1, where player 2 has the lowest posterior, he expects player 1 to play G with total probability $\frac{1}{2}$. It is shown more precisely in the appendix that this is always possible for N_2 finite but sufficiently high.

Finally, to show that the corner transitions are optimal: a deviation to a different N_2-state automaton would be detected if it resulted in player 2 playing R after some history where he is supposed to be in state $i \notin \{1, N_2\}$; such a deviation is costly in the period where it occurs (as both types of player 1 are playing G with probability greater than $\frac{1}{2}$ after this history), and also triggers a permanent switch by player 1 to a strategy of playing B w.p. 1 after every history. We show in the appendix that player 2's expected payoff from conforming to the prescribed strategy is at least as high as his maximum expected payoff when the normal type of player 1 plays B w.p. 1 (ie, his maximum payoff from deviating, if he counts on triggering the permanent "punishment phase" by player 1 and optimizes against this). Therefore, it is not optimal for player 2 to deviate to any automaton which alters the signal sequences that take him to states $1, N_2$; and by construction, he is then indifferent between playing R,L in these states, hence willing to randomize as prescribed.

4 Reputation Results for Patient Players

This section studies whether private information ($\pi > 0$) implies any lower bound on player 1's equilibrium payoff, for the case in which both players are long-lived and patient.

We study this case since the automaton set-up is most natural for long-run players. When both players are unbounded, reputation has no effect in moral hazard games with δ_2 near 1: Chan (2000) proved a folk theorem under perfect montoring. (In other related papers: Aumann-Sorin (1989) obtained a reputation result for pure common interest games, restricting to deterministic automata; and Cripps-Dekel-Pesendorfer (2003) obtained a reputation result for equally patient players in games of strictly conflicting interests). We find that when δ_2 is near 1, a bound on player 2's memory is not sufficient to obtain a reputation result. It is, however, possible to obtain an attractive lower bound on player 1's payoff when *both* players are bounded:

Proposition 3: Let $\pi >> 0$, $c_1, c_2 > 0$, and $\rho_1, \rho_2 < 1$. For any $\Delta > 0$, there exists ρ^*, c^* such that whenever $\rho_1 > \rho^*$ and $c_2 < c^*$, player 1's average expected payoff is at least $u_L - e - \Delta$ in any equilibrium. Moreover, this holds for any $\delta_2 \in [0, 1)$.

Recall that $(u_L - e)$ is player 1's commitment payoff, attained when (G,L) is played w.p. 1 every period. Proposition 3 therefore establishes a reputation result for almost-perfect monitoring, stating that the commitment payoff becomes a lower bound on player 1's equilibrium payoff in the limit as signal noise goes to zero.

4.1 Sketch of Proof

Consider the example from the introduction, with expected stage game payoffs

	L	R
G	$(1, 1)$	$(-1, 0)$
B	$(2, -1)$	$(0, 0)$

For this example, we wish to show that in any equilibrium, player 1's average payoff must go to 1 as ρ (the probability of a correct signal about player 1) goes to 1.

We know from Proposition 1 that P2's average equilibrium payoff must go to at least π as $\rho \to 1$. To see why, consider the following 2-state automaton strategy: play R in state 1, L in state 2, move $1 \to 2$ with probability $\sqrt{1-\rho}$ after a g-signal, move $2 \to 1$ after a b-signal (and with the residual probabilities stay in the current state). With this automaton, the lowest payoff that player 2 can possibly earn against a normal type of opponent is attained if player 1 (somehow) always plays B when he is in state 2 (giving player 2 a payoff of -1 there), and always plays G when he is in state 2 (to maximize the fraction of periods in which player 2 earns a low payoff). In this case, player 2's average payoff against a normal type of opponent is

$$-\frac{\Pr\{1 \to 2 \mid \mathcal{N}\}}{\Pr\{1 \to 2 \mid \mathcal{N}\} + \Pr\{2 \to 1 \mid \mathcal{N}\}} = -\frac{\rho\sqrt{1-\rho}}{\rho\sqrt{1-\rho}+\rho} \to 0 \text{ as } \rho \to 1$$

Against a commitment type, his average payoff is

$$\frac{\Pr\{1 \to 2 \mid \mathcal{C}\}}{\Pr\{1 \to 2 \mid \mathcal{C}\} + \Pr\{2 \to 1 \mid \mathcal{C}\}} = \frac{\rho\sqrt{1-\rho}}{\rho\sqrt{1-\rho}+1-\rho} \to 1 \text{ as } \rho \to 1$$

So no matter what strategy player 1 (of type $\mathcal{N}$) chooses, this automaton for player 2 guarantees him an average limit payoff of at least π.

Now, to get an intuition for the reputation result, suppose that player 2 is restricted to use a 3-state *deterministic* automaton, and that he plays R,R,L in states 1,2,3. Let V_i^1 be player 1's average expected continuation payoff conditional on knowing that player 2 is in state i, wlog assume $V_1^1 \leq V_2^1$, and note that we must have $V_3^1 > V_2^1$.

Suppose first that player 2 leaves state 3 after a g-signal. Then his expected payoff against a commitment type stays bounded below 1 as $\rho \to 1$ (as he sometimes switches from playing L to R even if he almost always observes g-signals). Then since we showed above that his average equilibrium payoff cannot stay bounded below π, it must be that his payoff against a normal type of opponent stays bounded above zero as $\rho \to 1$. Since player 2 is following a deterministic automaton strategy, this requires that the normal type of player 1 have an incentive to play G when he believes that player 2 is in state 3. Clearly this requires that player 2 move to a better (for player 1) state after g than b; and since we assumed that player 2 leaves player 1's favorite state after a g-signal, the only remaining pure-strategy possibility is that player 2 goes to state 2 after g, to state 1 after b, and that $\lim_{\delta\to 1} \frac{V_2^1 - V_1^1}{1-\delta} \geq \frac{1}{2\rho-1}$. If P2 stays in state 1 after b, and therefore (to earn a positive payoff) leaves state 1 after g, then this inequality cannot be strict, otherwise player 1 should play G w.p. 1 when he believes Player 2 is in state 1, so playing R there cannot be optimal for player 2 (note, this part relies on $c_1 > 0$ and $\rho_2 < 1$).

So we have

$$\begin{aligned} V_1^1 &= (1-\delta)(0) + \delta(1-\rho)V_1^1(g) + \delta\rho V_1^1 \\ \Rightarrow \ \frac{1}{2\rho-1} &=_{(\text{need})} \lim_{\delta\to 1} \frac{V_1^1(g) - V_1^1}{1-\delta} = \frac{\lim_{\delta\to 1} V_1^1}{1-\rho} \end{aligned}$$

But this means that $V_1^1 \to 0$ as $\rho \to 1$, which implies that player 2's average payoff against a normal type of opponent also goes to zero as $\rho \to 1$, a contradiction (to 3rd sentence of this paragraph). Therefore he must leave state 1 after a b-signal; then we have

$$\begin{aligned} V_1^1 &\geq (1-\delta)(0) + \delta\rho V_2^1 + \delta(1-\rho)V_1^1 \\ \Rightarrow \ \frac{1}{2\rho-1} &\leq_{(\text{need})} \lim_{\delta\to 1} \frac{V_2^1 - V_1^1}{1-\delta} \leq \frac{\lim_{\delta\to 1} V_1^1}{\rho} \end{aligned}$$

But this implies that player 1's average payoff goes to the commitment payoff 1 as $\rho \to 1$, as desired.

The remaining possibility is that player 2 does stay in state 3 after a g-signal. In this case, let player 1 choose a strategy in which he always plays G when player 2 is in state 3, and in the remaining states plays whichever action earns a higher continuation payoff. With this strategy, state 3 is reached from both states 1,2 with a probability that stays bounded above zero as $\rho \to 1$ (unless states 1,2 are absorbing, in which case player 2 earns payoff zero against both types, which again cannot be optimal as $\rho \to 1$). However, states 1,2 are reached from state 3 with probability at most $1-\rho$ (the probability of a b-signal in state 3 given that player 1 is playing G). Therefore the probability of state 3 (where (G,L) is played) goes to 1 as $\rho \to 1$, and therefore player 1's payoff must go to the commitment payoff.

This completes the proof if player 2 is restricted to play a deterministic 3-state automaton strategy; the argument is relatively straightforward to generalize if we instead wish to restrict Player 2 to strategies which can be implemented by deterministic automata of arbitrary (exogenously specified) size.

When randomized transitions are permitted, the argument is that in any equilibrium, player 2's average payoff against a commitment type must go to 1 as $\rho \to 1$: if not, then player 2 should add one extra state which plays L, and jump there after a g-signal with a probability proportional to $\sqrt{1-\rho}$ from some state in which, with positive probability, the normal type of opponent is believed to be playing B. The deviation then has a negligible effect as $\rho \to 1$ on player 2's expected payoff against a normal type of opponent (this relies on $c_1 > 0$, as discussed in the appendix), but guarantees a payoff of almost 1 against a commitment type, as the long-run probability of being in this extra state goes to 1 as $\rho \to 1$ against an opponent who is always playing G. Therefore, the deviation is profitable (from a strategy in which the payoff against a commitment type is bounded below 1), provided that the cost of this one extra state is not too high. Once we show that player 2's payoff against a commitment type goes to 1 as $\rho \to 1$, it is immediate that player 1's payoff cannot stay bounded below 1, as he can simply choose mimic the commitment type.

5 Equilibrium Existence (Issues)

Proposition 3 stated only that if we can find a sequence of equilibria along which the signal noise vanishes to zero, then along this sequence of equilibria, player 1's average payoff must go to the commitment payoff.

The result below establishes existence of such a sequence:

Proposition 4: For any $c_1, c_2, \pi > 0$, there exists ρ^* such that $\rho_1 \in (\rho^*, 1)$ implies existence of a pure strategy equilibrium with $N_1 = N_2 = 3$.

Proof:

Let both players use the following 3-state automaton strategy: move $1 \to 2$ regardless of the signal realization, $2 \to 3$ regardless of the signal realization, stay in state 3 after a *gl* or *gr*, and move $3 \to 1$ after *bl* or *br*; play (B,R) in states 1 and 2, (G,L) in state 3.

With this automaton, long-run frequencies do not depend on player 1's type, and are given by (for $s \in \{\mathcal{C}, \mathcal{N}\}$)

$$f^s(1) = f^s(2) = \frac{1-\rho}{1+2(1-\rho)},\ f^s(3) = \frac{1}{1+2(1-\rho)}$$

Clearly playing B is optimal for player 1 in states 1 and 2 (as continuation play is independent of the signal realization). To verify that playing G is optimal in state 3, observe that doing so implies

$$\begin{aligned} V_3^1 &= 1-\delta+\delta\rho V_3^1+\delta(1-\rho)V_1^1 \\ \Rightarrow \lim_{\delta\to 1}\frac{V_3^1-V_1^1}{1-\delta} &= \frac{1-\lim_{\delta\to 1}V_3^1}{1-\rho} = \frac{1-\frac{1}{1+2(1-\rho)}}{1-\rho} = \frac{2}{[1+2(1-\rho)]} \end{aligned}$$

This exceeds $\frac{1}{2\rho-1}$ for $\rho > \frac{5}{6}$, implying that playing G is indeed optimal in this range (as the gain in his continuation payoff from playing G rather than B in state 3, $(2\rho-1)\lim_{\delta\to 1}\frac{V_3^1-V_1^1}{1-\delta}$, exceeds the cost 1). Therefore if Player 2 is indeed expected to follow this automaton strategy, then player 1 is playing an unconstrained best response.

For player 2: if his opponent is in fact a normal type, then player 2 is playing the unconstrained best response (as he plays R when player 1 plays B, L when player 1 plays G, and his actions have no effect on player 1's continuation strategy). If his opponent is a commitment type, then player 1 is losing $\frac{2(1-\rho)}{1+2(1-\rho)}$ payoff units relative to maximum possible payoff, 1. It is straightforward to show that, provided that π is not too large, no other 3-state automaton can increase player 2's total expected payoff. Therefore, he needs at least one extra state to profitably deviate, implying that there are no profitable deviations when

$$c_2 > \frac{2\pi(1-\rho)}{1+2(1-\rho)}$$

For fixed c_2, we can always choose ρ high enough to satisfy this inequality. ■

[Will discuss some issues arising at this point in the seminar].

6 Conclusion

This paper studies a simple repeated moral hazard game in which one player has a potential incentive to develop a reputation as a commitment type, who always plays a fixed action (which is strictly dominated in the stage game). We departed from the literature by assuming that players 1,2 use finite automata to implement their strategies. This implies three main results, which differ from the existing literature:

By Proposition 1, the magnitude of the prior (probability that player 1 is a commitment type) matters: if it is too low, then the bound on player 2's memory may prevent him from ever believing that he is facing a commitment type, which eliminates player 1's incentives to mimic the commitment type.

By Proposition 2: in contrast to the Cripps-Mailath-Samuelson result, which showed that it is impossible for a player to maintain a reputation in the long run under imperfect monitoring, we find that for any bound on player 2's memory which is positive but not too large, there is an equilibrium in which Player 1 earns his maximal commitment payoff after every history.

Finally, Proposition 3 shows that when both players are patient and have bounded memory, there is a lower bound on player 1's payoff: and in particular, as the signal becomes perfectly informative, he earns at least his "simple commitment payoff" (from mimicking the commitment type) in any equilibrium. This differs from the standard literature, which shows that for unbounded equally patient players, there are no reputation effects in moral hazard games.

A Appendix

A.1 Preliminary Result

Proposition 0 below is the main ingredient for Proposition 1: we show here that if player 1 is unconstrained, then there is an equilibrium in which he plays a constant strategy if and only if P2's memory cost is sufficiently high. (It is clear that such an equilibrium exists if P2's cost is very high: in particular, if c_2 is so high that he cannot earn a positive payoff, then he should just choose a one-state automaton which always plays R, in which case P1's best response is to always play B. Here we show the other side: that for any cost below this, player 2 must have at least one state in which he plays L, and that this rules out an equilibrium in which player 1 plays a constant strategy.)

Proposition 0: For any $\rho_1 = \rho \in \left(\frac{1}{2}, 1\right)$ and $\pi > 0$, there is a cutoff $\overline{c}_2(\rho)$ such that: If $c_1 = 0$, then an equilibrium with $N_1 = N_2 = 1$ exists if and only if $c_2 > \overline{c}_2(\rho)$. If $c_2 < \overline{c}_2(\rho)$, then P2's expected payoff must be strictly positive in any equilibrium. Moreover: if $c_1 > 0$, then for any $\Delta > 0$, there exists $\underline{c_2}(\rho, \Delta) < \overline{c}_2(\rho)$ s.t. whenever $c_2 < \underline{c}_2(\rho, \Delta)$, P2's average expected payoff is at least $\pi - \Delta$.

Proof: We first calculate Player 2's best response to any player 1 strategy with $N_1 = 1$ (ie, both types of player 1 play a constant strategy), and show that this payoff is strictly decreasing in c_2, going to π as $c_2 \to 0$. This will then imply the second part of the statement in Proposition 0: that if $c_1 > 0$, hence P1 follows an automaton strategy, then P2 must earn a payoff arbitrarily close to π if his memory cost c_2 is sufficiently small. For the first part of the Proposition, we show that (i)if $N_1 = 1$, then there is a cutoff $\overline{c}_2(\rho)$ s.t. if c_2 is higher than this, P2's best response is a one-state automaton which always plays R, in which case it is indeed optimal for P1 to follow a one-state automaton strategy (namely, play B every period); (ii)if $N_1 = 1$ and c_2 is less than this cutoff, then P2's best response implies that $N_1 = 1$ is in fact not optimal for P1 at $c_1 = 0$. This implies the desired result, that there is an equilibrium with $N_1 = N_2 = 1$ (namely, P1 always plays B, and P2 always plays R - unless the prior is very high) iff P2's memory cost is sufficiently high.

So: first assume that $N_1 = 1$, and let α be the probability that the normal type of P1 plays G. Clearly, this can only happen at equilibrium if $\alpha < \frac{1}{2}$ (as if $\alpha \geq \frac{1}{2}$ and $\pi > 0$, then the normal type of P2 always expects G to be played with probability greater than $\frac{1}{2}$; and since $N_1 = 1$ implies that his action choice does not affect his continuation payoff, he should then use the one-state automaton strategy of playing L w.p. 1 in every period; but against this automaton, P1's best response is in fact to play his dominant action B every period, contradicting optimality of $\alpha \geq \frac{1}{2}$).

So, assume $\alpha < \frac{1}{2}$; and let $p_\alpha \equiv 1 - \rho + (2\rho - 1)\alpha$ be the associated probability of a g-signal, conditional on the normal type of P1. Suppose that P2 chooses an automaton with $N_2 \geq 2$ states: by Wilson (200?), the optimal N_2-state automaton against an iid opponent (with two possible types) satisfies the following conditions: (i)the action rule is deterministic: without loss of generality set $d(1) = 0, d(N_2) = 1$ (recall, $d(j)$ is the probability that P2 plays L in state j); (ii)ordering the states s.t. $f_j^{\mathcal{C}}/f_j^{\mathcal{N}}$ is increasing in j (where f_j^s is the long-run frequency of state j conditional on type $s \in \{\mathcal{C}, \mathcal{N}\}$), the transition rule specifies: for $j \in \{2, 3, .., N_2 - 1\}$, move $j \to j+1$ with positive probability (bounded above zero for all δ) after a g-signal, and $j \to j-1$ with positive probability after a b-signal, in both cases staying in state j with any residual probability; in state 1 (N_2), stay after a b-signal (g-signal), and for δ sufficiently close to 1, move $1 \to 2$ after g (and $N_2 \to N_2 - 1$ after b) with a probability proportional to $\sqrt{1-\delta}$.

Letting $\sigma \equiv \frac{\Pi_{j \neq N_2} \sigma^g_{j,j+1}}{\Pi_{j \neq 1} \sigma^b_{j,j-1}}$, this implies that conditional on the commitment type, the relative long-run frequency of state N_2 to state 1 is $\frac{f^{\mathcal{C}}(N_2)}{f^{\mathcal{C}}(1)} = \sigma \left(\frac{\rho}{1-\rho}\right)^{N_2-1}$; conditional on the the normal type of P1, the expression $\dfrac{f^{\mathcal{N}}(N_2)}{f^{\mathcal{N}}(1)}$ is identical, just replacing ρ (the probability of a g-signal conditional on type $\mathcal{C}$) with p_α. As $\delta \to 1$, the stated transition rule implies that the long-run frequency of state $j \notin \{1, N_2\}$ goes to zero: so, using $\lim_{\delta \to 1} (f^s(N_2) + f^s(1)) = 1$ and the above ratios, this implies that P2's limiting (as $\delta \to 1$) payoff is:

$$\pi \cdot \frac{\sigma}{\sigma + \left(\frac{1-\rho}{\rho}\right)^{N_2-1}} - (1-\pi) \cdot \frac{\sigma(1-2\alpha)}{\sigma + \left(\frac{1-p_\alpha}{p_\alpha}\right)^{N_2-1}}$$

The FOC for σ is

$$\frac{\pi}{(1-\pi)(1-2\alpha)} = \left(\frac{1-p_\alpha}{p_\alpha}\right)^{N_2-1} \left(\frac{\rho}{1-\rho}\right)^{N_2-1} \left(\frac{\sigma + \left(\frac{1-\rho}{\rho}\right)^{N_2-1}}{\sigma + \left(\frac{1-p_\alpha}{p_\alpha}\right)^{N_2-1}}\right)^2$$

Note that here σ is a ratio of probabilities, hence can take on any value in $[0, \infty)$; therefore, it is possible to choose σ to satisfy the above expression iff

$$\left(\frac{p_\alpha}{1-p_\alpha}\right)^{N_2-1} \left(\frac{1-\rho}{\rho}\right)^{N_2-1} < \frac{\pi}{(1-\pi)(1-2\alpha)} < \left(\frac{1-p_\alpha}{p_\alpha}\right)^{N_2-1} \left(\frac{\rho}{1-\rho}\right)^{N_2-1}$$

(If the LHS inequality is violated, then it is better to choose a one-state automaton which plays R w.p. 1; and if the RHS inequality is violated, then it is better to choose a one-state automaton which plays L w.p. 1). For the range in between, the optimal σ earns an expected average payoff of

$$\pi \cdot \frac{\left(1 - \sqrt{\frac{1-\pi}{\pi}(1-2\alpha)\left(\frac{p_\alpha(1-\rho)}{\rho(1-p_\alpha)}\right)^{N_2-1}}\right)^2}{1 - \left(\frac{p_\alpha(1-\rho)}{\rho(1-p_\alpha)}\right)^{N_2-1}}$$

Since $\rho > \frac{1}{2} \Leftrightarrow \frac{1-\rho}{\rho} < 1$, and $\alpha < \frac{1}{2} \Leftrightarrow \frac{p_\alpha}{1-p_\alpha} < 1$, this expression goes to π as $N_2 \to \infty$, and is strictly positive whenever the first inequality above is satisfied. Moreover, it is straightforward to check that the expression is strictly increasing and concave in N_2. Therefore, if the cost of

N_2 states is $c_2 N_2$, then there will be a uniquely optimal choice of N_2; and for c_2 sufficiently close to zero, this choice earns a payoff which is strictly positive, and can be made arbitrarily close to π.

To complete the proof: we showed that whenever P1 follows a constant strategy, P2 can earn an expected payoff arbitrarily close to π if his memory cost c_2 is sufficiently small. The second statement in Proposition 0 then follows from the fact that if P2 can earn a payoff arbitrarily close to π when P1 plays a constant strategy, then he can also do so against any automaton strategy. Then finally, to prove the first statement: it is straightforward to show that if P2 indeed follows an automaton strategy as described above for $N_2 \geq 2$, then in fact a constant strategy is not optimal for P1 (provided that c_1 is close to zero). Therefore, a strategy with $N_1 = 1$ can only be optimal for P1 if we also have $N_2 = 1$. In this case, P2's play is independent of P1's strategy, therefore optimality for P1 requires playing his dominant action B every period. So, an equilibrium with $N_1 = N_2 = 1$ exists iff, at $\alpha = 0$, it is optimal for P2 to choose a one-state automaton. This holds iff P2's expected payoff from choosing N_2 states (as calculated above at $\alpha = 0$) is maximized by $N_2 = 1$: that is, whenever

$$\pi \cdot \frac{\left(1 - \sqrt{\frac{1-\pi}{\pi}} \left(\frac{1-\rho}{\rho}\right)^{N_2 - 1}\right)^2}{1 - \left(\frac{1-\rho}{\rho}\right)^{2(N_2-1)}} - c_2 \left(N_2 - 1\right) < \max\{0, 2\pi - 1\} \; \forall N_2$$

As noted above, the first term is increasing and concave in N_2, which further implies that the payoff is strictly decreasing in c_2; hence, $\overline{c}_2(\rho)$ is value of c_2 for which the above expression holds with equality. ■

A.2 Proof of Proposition 1

This is for N_2 exogenous, and shows that player 1 benefits from his reputation iff N_2 is sufficiently large, relative to ρ, π. The remaining part of the proposition - that player 2's payoff must go to $\pi > 0$ as either $\rho \to 1$ or $N_2 \to \infty$ - is implied by Proposition 0.

First, to show that there is a NE of the game with incomplete information in which player 1's expected payoff is 0: Let γ_1 be the behavior strategy according to which P1 plays B with probability 1 after every history. Then Player 2's problem is to choose a transition rule to maximize

$$\sum_{\{i \in \mathcal{N} | d(i) = L\}} \left[\pi f^{\mathcal{C}}(i) - (1-\pi) f^{\mathcal{N}}(i)\right]$$

together with a decision rule satisfying $d(i) = L$ iff $\Pr\{\mathcal{C}|i\} \geq \Pr\{\mathcal{N}|i\}$. This is identical to the problem studied in Wilson (200?), which established that if $\sqrt{\frac{\pi}{1-\pi}} < \left(\frac{1-\rho}{\rho}\right)^{N_2 - 1}$, then the upper bound on Player 2's expected payoff is 0, attained by an automaton which plays R in all memory states. Given this automaton, it is indeed a best response for player 1 to play B after any history.

Next, to show that δ_2 near 0 implies that player 1's average discounted payoff cannot exceed $(u_L - e) - e\frac{1-\rho}{2\rho-1}$ (the maximum NE payoff in the game with complete information), we first calculate bounds on player 2's beliefs about the type of player 1. Fix strategies (γ_1, γ_2), and order the states $\{1, 2, ..., N_2\}$ such that the induced beliefs $\frac{\pi}{1-\pi}\frac{f^{\mathcal{C}}(i)}{f^{\mathcal{N}}(i)}$are weakly increasing in i. Define $\tau^s_{i,j}$ as the total probability of an $i \rightarrow j$ transition conditional on type $s \in \{\mathcal{C}, N\}$: eg for a state i with $d(i) = L$, $\tau^c_{i,j} = \rho\sigma^{gL}_{i,j} + (1-\rho)\sigma^{bL}_{i,j}$. Rearranging the steady-state equations $f^s(i) = \sum_j f^s(j)\tau^s_{j,i}$, we have for all $i \in \{1, ..., N_2\}$:

$$\frac{\sum_{j\leq i-1} f^{\mathcal{C}}(j)\tau^{\mathcal{C}}_{j,i}}{\sum_{j\leq i-1} f^{\mathcal{N}}(i)\tau^{n}_{j,i}} = \frac{f^{\mathcal{C}}(i)\sum_{j\neq i}\tau^{\mathcal{C}}_{i,j} - \sum_{j\geq i+1} f^{\mathcal{C}}(j)\tau^{\mathcal{C}}_{j,i}}{f^{\mathcal{N}}(i)\sum_{j\neq i}\tau^{N}_{i,j} - \sum_{j\geq i+1} f^{\mathcal{N}}(j)\tau^{N}_{j,i}}$$

Note also that $\frac{1-\rho}{\rho} \leq \frac{\tau^{\mathcal{C}}_{i,j}}{\tau^{N}_{i,j}} \leq \frac{\rho}{1-\rho}$ for all i, j. Our ordering of the states implies that the LHS above is at most $\frac{f^{\mathcal{C}}(N_2-1)}{f^{\mathcal{N}}(N_2-1)}\frac{\rho}{1-\rho}$, while the RHS is at least $\frac{f^{\mathcal{C}}(N_2)}{f^{\mathcal{N}}(N_2)}\frac{1-\rho}{\rho}$: hence, we have $\frac{f^{\mathcal{C}}(N_2)}{f^{\mathcal{N}}(N_2)}\frac{1-\rho}{\rho} \leq \frac{\rho}{1-\rho}\frac{f^{\mathcal{C}}(N_2-1)}{f^{\mathcal{N}}(N_2-1)}$. Substituting this into the equation for $N_2 - 1$: the LHS above is at most $\frac{f^{\mathcal{C}}(N_2-2)}{f^{\mathcal{N}}(N_2-2)}\frac{\rho}{1-\rho}$, while the RHS is at least $\frac{f^{\mathcal{C}}(N_2-1)}{f^{\mathcal{N}}(N_2-1)}\frac{1-\rho}{\rho}$; hence, we have $\frac{f^{\mathcal{C}}(N_2-1)}{f^{\mathcal{N}}(N_2-1)}\frac{1-\rho}{\rho} \leq \frac{f^{\mathcal{C}}(N_2-2)}{f^{\mathcal{N}}(N_2-2)}\frac{\rho}{1-\rho}$. Iterating this argument:

$$\frac{f^{\mathcal{C}}(N_2)}{f^{\mathcal{N}}(N_2)} \leq \left(\frac{\rho}{1-\rho}\right)^2 \frac{f^{\mathcal{C}}(N_2-1)}{f^{\mathcal{N}}(N_2-1)} \leq ... \leq \frac{f^{\mathcal{C}}(1)}{f^{\mathcal{N}}(1)}\left(\frac{\rho}{1-\rho}\right)^{2(N_2-1)}$$

Moreover, the ordering of the states implies $\frac{f^{\mathcal{C}}(1)}{f^{\mathcal{N}}(1)} \leq 1$. (It is not possible that all states are reached more frequently conditional on $\mathcal{C}$ than $\mathcal{N}$). Hence, for any strategy pair (γ_1, γ_2), we have:

$$\max_{i\in\mathcal{N}} \frac{\Pr\{\mathcal{C}|i\}}{\Pr\{\mathcal{N}|i\}} = \frac{\pi}{1-\pi}\frac{f^{\mathcal{C}}_{N_2}}{f^{N}_{N_2}} \leq \frac{\pi}{1-\pi}\left(\frac{\rho}{1-\rho}\right)^{2(N_2-1)}$$

If the bound in the proposition holds, then this is below 1.

Finally, let $i^* \in \mathcal{N}$ be the memory state in which player 1's continuation payoff is highest. Define $V^1(i^*) \equiv E[(1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}u_1(a^t_1, a^t_2) \mid i^*]$ as player 1's expected continuation payoff, conditional on knowing i^*. If $d(i^*) = R$ then we are done, as this implies that player 1's expected payoff is at most 0. So, assume that $d(i^*) = L$. Since we showed above that player 2 always believes he is more likely to be facing a normal type, $d(i^*) = L$ can only be optimal for a myopic player 2 if he expects the normal type to play G with sufficiently high probability in i^*. For player 1 to want to play G in i^*, we need

$$\delta(2\rho-1)\left[\sigma^g_{i^*,i^*} - \sigma^b_{i^*,i^*}\right]\frac{V^1(i^*) - \max_{j\neq i^*} V^1(j)}{1-\delta} \geq e \qquad (0)$$

To show that this is impossible when player 1's payoff is too high: let α_i^* denote the probability that player 1 plays G in i^*, and $p_i^* = 1-\rho+(2\rho-1)\alpha_i^*$ the implied probability of a g-signal in state i^*. Then we have,

$$V^1(i^*) \leq (1-\delta)(u_L - e\alpha_{i^*}) + \delta V^1(i^*) - \delta\left(p_{i^*}(1-\sigma^g_{i^*,i^*}) + (1-p_{i^*})(1-\sigma^b_{i^*,i^*})\right)\left(V^1(i^*) - \max_{j\neq i^*} V^1(j)\right)$$

which rearranges to

$$\lim_{\delta\to 1}\frac{V^1(i^*) - \max_{j\neq i^*} V^1(j)}{1-\delta} \leq \frac{u_L - e\alpha_{i^*} - \lim_{\delta\to 1} V^1(i^*)}{\delta\left(p_{i^*}(1-\sigma^g_{i^*,i^*}) + (1-p_{i^*})(1-\sigma^b_{i^*,i^*})\right)}$$

So for (0) to hold, we need

$$\begin{aligned}\left[\sigma^g_{i^*,i^*} - \sigma^b_{i^*,i^*}\right]\left[u_L - e\alpha_{i^*} - \lim_{\delta\to 1} V^1(i^*)\right] &\geq e\left[\frac{\left(p_{i^*}(1-\sigma^g_{i^*,i^*}) + (1-p_{i^*})(1-\sigma^b_{i^*,i^*})\right)}{2\rho-1}\right] \\ &= e\left[\frac{(1-\rho)(1-\sigma^g_{i^*,i^*}) + \rho(1-\sigma^b_{i^*,i^*})}{2\rho-1} + \alpha_i^*\left(\sigma^b_{i^*,i^*} - \sigma^g_{i^*,i^*}\right)\right]\end{aligned}$$

So,

$$\left[1-\sigma^b_{i^*,i^*}\right]\left[u_L - \lim_{\delta\to 1} V^1(i^*) - e\frac{\rho}{2\rho-1}\right] > (1-\sigma^g_{i^*,i^*})\left[u_L - \lim_{\delta\to 1} V^1(i^*) + e\frac{1-\rho}{2\rho-1}\right]$$

The RHS is non-negative (the payoff cannot possibly exceed u_L), so this requires

$$\lim_{\delta\to 1} V^1(i^*) \leq u_L - e\frac{\rho}{2\rho-1} = (u_L - e) - e\left(\frac{1-\rho}{2\rho-1}\right)$$

as desired. (This proof is incomplete because the last part of the argument relies on player 1 knowing when player 2 is in state i^*.) ■

A.3 Proof of Proposition 2

A.3.1 Steady-State Probabilities

It will be useful to calculate the steady-state distribution over $\{1,2,...,N_2\}$, conditional on player 1's type and on (γ_1,γ_2). Define p_i^s as the probability of a g-signal when player 2 is in state $i \in \{1,..,N_2\}$, conditional on player 1's type $s \in \{\mathcal{C},\mathcal{N}\}$. Denote by $f_i^s \equiv f^s(i)$ the steady-state probability of i conditional on player 1 being type s. If player 2 follows the transition rule specified by γ_2, then f^s is the solution to the following system of equations:

$$\begin{aligned}&i = N : f^s_{N_2}p^s_{N_2} = f^s_{N_2-1}p^s_{N_2-1}\\ &4 \leq i \leq N_2 - 1 : f_i^s = f^s_{i-1}p^s_{i-1} + f^s_{i+1}(1-p^s_{i+1})\\ &i = 3 : f_3^s = f_1^s p_1^s + f_2^s p_2^s + f_4^s(1-p_4^s)\\ &i = 2 : f_2^s = f_3^s(1-p_3^s)\\ &i = 1 : f_1^s p_1^s = f_2^s(1-p_2^s)\end{aligned}$$

(For example: each period, player 1 is in state 1 if either he was already here the previous period and observed a b-signal, or if he was in state 2 and observed a b-signal: hence, $f_1^s = f_1^s(1-p_1^s) + f_2^s(1-p_2^s)$. Solving this system recurvisely yields:

$$3 \leq i \leq N-1 : f_i^s = \prod_{j=i}^{N_2-2} \frac{1-p_{j+1}^s}{p_j^s} \cdot \frac{p_{N_2}^s}{p_{N_2-1}^s} f_{N_2}^s$$

$$f_2^s = (1-p_3^s)f_3^s$$

$$f_1^s = \frac{(1-p_2^s)}{p_1^s}(1-p_3^s)f_3^s$$

Write all of these in terms of $f_{N_2}^s$ and use the fact that probabilities sum to 1, to solve for f^s. Under the commitment strategy, $p_i^{\mathcal{C}} = \rho \ \forall i$, implying

$$f_{N_2}^C = \left[\frac{2\rho-1}{3\rho-1-\left(\frac{1-\rho}{\rho}\right)^{N_2-2}}\right]; \quad f_1^C = \frac{1}{\rho}\left(\frac{1-\rho}{\rho}\right)^{N_2-3}\left[\frac{2\rho-1}{3\rho-1-\left(\frac{1-\rho}{\rho}\right)^{N_2-2}}\right] \tag{1}$$

For the normal type: recall that as long as there are no deviations, player 1 always knows player 2's state i; he must play B in state N_2, so $p_{N_2}^{\mathcal{N}} = (1-\rho)$, but is willing to choose any probabilties p_i^n in the remaining states. For future reference: if he sets $p_i^{\mathcal{N}} = \frac{1}{2} \ \forall i \neq 1, N_2-1, N_2$, then we obtain

$$f_{N_2}^{\mathcal{N}} = \frac{1}{\left[1+\frac{1-\rho}{p_{N-1}^{\mathcal{N}}}\left(1+2(1-p_{N_2-1}^{\mathcal{N}})(N_2-4)\left(1+\frac{1}{2}+\frac{1}{4p_1^{\mathcal{N}}}\right)\right)\right]}; \tag{2}$$

$$f_1^{\mathcal{N}} = \frac{\frac{(1-\rho)}{2p_1^{\mathcal{N}}}\left(\frac{1-p_{N-1}^{\mathcal{N}}}{p_{N-1}^{\mathcal{N}}}\right)}{\left[1+\frac{1-\rho}{p_{N_2-1}^{\mathcal{N}}}\left(1+2(1-p_{N_2-1}^{n})(N_2-4)\left(1+\frac{1}{2}+\frac{1}{4p_1^{\mathcal{N}}}\right)\right)\right]}$$

A.3.2 Optimality for Player 2

We need to choose $(p_i^{\mathcal{N}})_{i\in\{1,2,\dots,N_2\}}$ such that:

- $\Pr\{\mathcal{C}|N_2\} = \frac{1}{2}$: this implies that player 2 is indifferent between playing L, R (hence willing to randomize) in state N_2, given that he expects the normal type to play B here

- Player 2 is indifferent between L, R conditional on state 1 (to be willing to randomize in state 1)

- Player 2 is indifferent between L, R conditional on observing a b-signal in state 2 (This is required for optimality of the 2 → 1 transition, $\sigma_{21}^b = 1$: note that transitions out of states 1,2 are identical (go to 3 after a g-signal, 1 after a b-signal), so moving from 2 to 1 only affects the probability of playing L in the subsequent period)

If these conditions are satisfied, and $p_i^{\mathcal{N}} \geq \frac{1}{2}\ \forall i \neq 1, N_2$, then player 2's strategy is optimal for any δ_2. To see this: in all states $i \neq 1, N_2$, he is supposed to play L : this is a myopic best response to γ_1 (since normal P1 plays G here with probability at least $\frac{1}{2}$), and a strict myopic best response to the commitment type's strategy. The above conditions guarantee that player 2's action choice is also a myopic best response at all other information sets (states 1,$N_2, N_2 - 1$, and when first moving into state 1). Therefore, any one-shot deviation in the action can only reduce the current-period payoff, and may trigger the permanent punishment phase by Player 1: A one-shot deviation in the transition only matters if it changes the signal sequences that take player 2 to states 1,N_2: and in this case, again the result is that he will play R (with positive probability) when supposed to play L with probability 1, triggering a permanent switch by player 1 to the strategy of always playing B. So, there are no incentives for one-shot deviations. It is also straightforward to show that player 2's expected payoff in this equilibrium exceeds his payoff from optimizing against the belief that type n always plays B (ie, count on triggering a deviation and design the corresponding optimal automaton).

To show that it is possible to satisfy the above conditions for N_2 sufficiently high: the first condition, action indifference in state N_2, requires

$$\frac{\Pr\{\mathcal{C}|N_2\}}{\Pr\{\mathcal{N}|N_2\}} \equiv \frac{\pi}{1-\pi}\frac{f_{N_2}^{\mathcal{C}}}{f_{N_2}^{\mathcal{N}}} = 1 \tag{3}$$

For the second and third conditions (action indifference in 1, and after observing a b-signal in state 2) to hold, player 1 must play G with a slightly lower probability after the first b-signal in state 2, than after two or more consecutive b-signals starting in state 2. (The probability of a commitment type is lower in the latter case, so we need to increase the probability that the normal type plays G to keep player 2 indifferent). More precisely: after every history in which player 1 (correctly) believes that player 2 is in state 2, let him play G with probability α_1^0 after the first b-signal, and with probability α_1^1 after each subsequent consecutive b-signal. Also define p_1^0, p_1^1 as the probabilities of a g-signal induced by α_1^0, α_1^1 (ie $\alpha_1^0 = 1-\rho+(2\rho-1)\alpha_1^0$). Then the long-run frequency with which player 1 is type $\mathcal{N}$ and plays G when player 2 is in state 1 is given by:

$$(1-\pi)f_1^{\mathcal{N}}\alpha_1 \equiv (1-\pi)f_2^{\mathcal{N}}(1-p_2)\left[\alpha_1^0 + (1-p_1^0)\alpha^1 + (1-p_1^0)(1-p_1^1)\alpha_1^1 + ...\right] = \frac{f_2^{\mathcal{N}}(1-p_2)}{p_1^1}\left[(1-\rho)\alpha_1^0 + \rho\alpha_1^1\right]$$

And the probability that player 2 is in state 1 conditional on type $\mathcal{N}$ is:

$$f_1^{\mathcal{N}} \equiv (f_2^{\mathcal{N}}(1-p_2)\left[1 + (1-p_1^0)\left(1 + (1-p_1^1) + (1-p_1^1)^2 + ...\right)\right] = \frac{f_2^{\mathcal{N}}(1-p_2)}{p_1^1}\left(1 + (2\rho-1)\left(\alpha_1^1 - \alpha_1^0\right)\right)$$

Conditional on being in state 1, player 2 is then indifferent between playing L, R iff the total probability that player 1 plays G is $\frac{1}{2}$:

$$\frac{\pi f_1^{\mathcal{C}} + (1-\pi) f_1^{\mathcal{N}} \alpha_1}{\pi f_1^{\mathcal{C}} + (1-\pi) f_1^{\mathcal{N}}} = \frac{1}{2} \Leftrightarrow \frac{\pi f_1^{\mathcal{C}}}{(1-\pi) f_2^{\mathcal{N}} (1-p_2)} = \frac{\left(1-\alpha_1^1-\alpha_1^0\right)}{1-\rho+(2\rho-1)\alpha_1^1}$$

Similarly, conditional on being in state 2 and observing a b-signal, he is indifferent between L, R if:

$$\frac{\pi f_1^{\mathcal{C}} + (1-\pi) f_2^{\mathcal{N}} (1-p_2) \alpha_1^0}{\pi f_1^{\mathcal{C}} + (1-\pi) f_2^{\mathcal{N}} (1-p_2)} = \frac{1}{2} \Leftrightarrow \frac{\pi f_1^{\mathcal{C}}}{(1-\pi) f_2^{\mathcal{N}} (1-p_2)} = (1-2\alpha_1^0)$$

Solving, we need:

$$\frac{\pi f_1^{\mathcal{C}}}{(1-\pi) f_2^{\mathcal{N}} (1-p_2)} = (1-2\alpha_1^0) \text{ and } \alpha_1^1 = \frac{1}{2} \tag{4}$$

So, for player 2's strategy to be optimal, it sufficies to choose (p_i) such that $p_i \geq \frac{1}{2}$ $\forall i \neq 1, N-1, N$, and equations (3),(4) are satisfied. For example, if he sets $p_i^{\mathcal{N}} = \frac{1}{2}$ $\forall i \neq 1, N$, then subsituting (1),(2) into (3),(4), we need:

$$\frac{\pi}{1-\pi} \frac{(2\rho-1)\left[1+2(1-\rho)\left(1+(N_2-4)\left(1+\frac{1}{2}+\frac{1}{4p_1^{\mathcal{N}}}\right)\right)\right]}{3\rho-1-\left(\frac{1-\rho}{\rho}\right)^{N_2-2}} = 1 \tag{3a}$$

$$\frac{\pi}{1-\pi} \frac{2(2\rho-1)}{\rho(1-\rho)} \left(\frac{1-\rho}{\rho}\right)^{N_2-3} \left[\frac{1+2(1-\rho)\left(1+(N_2-4)\left(1+\frac{1}{2}+\frac{1}{4p_1^{\mathcal{N}}}\right)\right)}{3\rho-1-\left(\frac{1-\rho}{\rho}\right)^{N_2-2}}\right] = 1-2\alpha_1^0 \tag{4a}$$

where $p_1^{\mathcal{N}}$ is the *average* probability of a g-signal conditional on state 1 and type $\mathcal{N}$: namely, the solution to $p_1^{\mathcal{N}} = \frac{f_2^{\mathcal{N}}(1-p_2)}{f_1^{\mathcal{N}}}$, which at $\alpha_1^1 = \frac{1}{2}$ is $p_1^{\mathcal{N}} = \frac{1}{1+2\rho-2(2\rho-1)\alpha_1^0}$. The LHS of (3a) goes to infinity as $N_2 \to \infty$, while the RHS of (4a) goes to 0; since we can choose α_1^0 arbitrarily, and are also free to increase any p_i^N for $i \neq N_2$ (note that at $p_i^n = \rho$ the commitment and normal strategies are identical, so the LHS and RHS of (3a),(4a) would be very close to the ex ante prior $\frac{\pi}{1-\pi}$), it is always possible to satisfy these equalities for N_2 sufficiently large.

For example: at $\rho = .95$ and $\frac{\pi}{1-\pi} = .05$, we need , equation (3a) needs $N = 205$ and α_1^0 very close to (slightly below) $\frac{1}{2}$. The minimal N_2 required is strictly decreasing in both ρ and $\frac{\pi}{1-\pi}$.

This completes the proof, as the remaining arguments were shown in the text. ■

A.4 Proof of Proposition 3

Lemma 1: If $c_1 > 0$ and $\rho_2 < 1$, then $\sigma(j, gl) = \sigma(j, gr)$ and $\sigma(j, bl) = \sigma(j, br)$ for any state $j \in \{1, 2, ..., N_1\}$ in which P1 has a probability-1 belief on either P2's current memory state or next action.[8]

[8] A signal realization is uninformative, for example, if player 1 has a probability-1 belief on P2's memory state and the action that he will play.

Proof: In any such state j, the signal about P2's action conveys no information. Therefore, since P1's expectation about P2's continuation play is independent of the signal realization, the set of unconstrained optimal continuation strategies for P1 is also independent. But since $\rho_2 < 1$, P1's equilibrium automaton must specify moving to a state that maximizes his continuation payoff after both signal realizations l, r. Then the best-response automaton with the fewest # states (as required for optimality with $c_1 > 0$) must set $\sigma(j, gl) = \sigma(j, gr)$ and $\sigma(j, bl) = \sigma(j, br)$. [add a step here] ■

Lemma 2: Fix $\rho_2 < 1, c_2 > 0$; take a sequence $\rho_{1,n} \to 1$, and let (σ_n) be a corresponding convergent sequence of equilibria. Let α_n be the average expected probability (across all histories) with which a normal type of P1 plays G: if $\alpha_n \to 1$, then also player 1's average payoff satisfies $V_n^1 \to 1$.

Proof: Suppose first that for some $\varepsilon > 0$, there is a state j in P2's automaton such that $f_n^{\mathcal{N}}(j) > \varepsilon$ for all n, and in which P2 believes that P1 expects him to play both L and R after a g-signal with probability at least ε. Then it must be that whenever P2 is in state j, he expects his opponent to play G in the continuation equilibrium following both gl and gr with an average probability that goes to 1 as $\rho_{1,n} \to 1$. (Otherwise, there would be a strictly positive probability of a history occuring after which P1 plays G with a probability that stays bounded below 1 as $\rho_{1,n} \to 1$, contradicting $\alpha_n \to 1$). But, $f_n^{\mathcal{N}}(j) > \varepsilon$ for all n and $\alpha_n \to 1$ imply also that in state j, P2 expects his opponent to play G in both the current period, and in the subsequent period following a g-signal, with a probability that goes to 1 as $n \to \infty$. This implies that when P2 observes a g-signal in state j, it is a strict myopic best response to play L in the subsequent period, while playing L rather than R will affect P1's continuation strategy (and hence P2's expected continuation payoff) by an amount which goes to zero as $\rho_{1,n} \to 1$. Therefore, if such a state j exists then P2 cannot be playing a best response, a contradiction to equilibrium.

Therefore, for an equilibrium with $\alpha_n \to 1$, it must be that in every state j of P2's automaton with $f_n(j) \nrightarrow 0$, P2 believes that his opponent expects him to either play L with a probability that goes to 1 as $\rho_{1,n} \to 1$, or R with a probability that goes to 1 as $\rho_{1,n} \to 1$. In the latter case, $\rho_2 < 1$ then implies that for n sufficiently high, P2's action choice conveys no information to P1; then by Lemma 1 and $c_1 > 0$, playing L vs R does cannot affect his continuation payoff, and hence he must choose a myopic best response. Since $f_n(j) \nrightarrow 0$, $\alpha_n \to 1$ implies that P1 is expected to play G with probability near 1, and hence this myopic best response is to play L. Hence, for any state with positive limit probability, P1 must expect P2 to play L with a probability that goes to 1 as $\rho_{1,n} \to 1$.

But this implies that the fraction of periods in which (G,L) is played goes to 1 along the sequence σ_n, and hence $V_n^1 \to 1$ as desired. ■

Lemma 3: Fix $\rho_2 < 1, c_2 > 0$, take a sequence $\rho_{1,n} \to 1$, and let (σ_n) be a corresponding convergent sequence of equilibria. Player 2's average expected payoff against a commitment type must go to 1 as $n \to \infty$.

Proof: Suppose, by contradiction, that we can construct such a sequence (σ_n) in which P2's payoff against a commitment type stays bounded below 1 as $\rho_{1,n} \to 1$. Then there must exist some $\varepsilon > 0$ such that for each equilibrium σ_n along the sequence, P2's automaton has some state j with the following features: (i) $f_n^{\mathcal{C}}(j) > \varepsilon$; (ii) if P2 observes a g-signal in state j, he expects the normal type of P1 to play B with probability at least ε in the subsequent period. (Otherwise, the average expected probability with which a normal type of P1 plays G must go to 1 as $\rho_{1,n} \to 1$; but then by Lemma 2, this implies $V_n^1 \to 1$, which requires that P2 play L with an average probability that goes to 1 as $\rho_{1,n} \to 1$, and hence his expected payoff against a commitment type of opponent must also go to 1, a contradiction).

For each equilibrium along the sequence, fix any such $j \in \{1, 2, ..., N_2\}$ (where N_2 is the size of the automaton P2 chooses under σ_n). Now, consider a deviation to the following (N_2+1)-state automaton: in states $1, 2, ..., N_2$, follow the behavior prescribed by σ_n, except for the following modification: in state j, with probability $\sqrt{1-\rho_{1,n}}$, instead jump to state N_2+1 following a g-signal. In state N_2+1, play L w.p. 1, stay after a g-signal, and after a b-signal follow the transition rule prescribed for $\sigma(j, g)$.[9]

Under this deviation: conditional on a normal type of opponent, the long-run probability of state N_2+1 satisfies

$$f_n^{\mathcal{N}}(N_2+1) \leq f_n^{\mathcal{N}}(N_2+1)\,[1-\rho_{1,n}+(2\rho_{1,n}-1)\varepsilon] + f_n^{\mathcal{N}}(j) \cdot p_j\sqrt{1-\rho_{1,n}}$$

$$\Rightarrow f_n^{\mathcal{N}}(N_2+1) \leq f_n^{\mathcal{N}}(j)\frac{p_j\sqrt{1-\rho_{1,n}}}{[\rho_{1,n}-(2\rho_{1,n}-1)\varepsilon]}$$

where p_j is the average probability with which P2 observes a g-signal in state j, conditional on a normal type of opponent. Since $\varepsilon > 0$, this implies that $f_n^{\mathcal{N}}(N_2+1) \to 0$ as $\rho_{1,n} \to 1$. Therefore, by construction, this change in P2's expected payoff resulting from this deviation goes to zero against a normal type of opponent as $\rho_{1,n} \to 1$.

However, conditional on a commitment type of opponent, we have

$$f_n^{\mathcal{C}}(N_2+1) \cdot (1-\rho_{1,n}) = f_n^{\mathcal{C}}(j) \cdot \rho_{1,n}\sqrt{1-\rho_{1,n}}$$

$$\Rightarrow f_j^C = \frac{\sqrt{1-\rho_{1,n}}}{\rho_{1,n}} \cdot f_n^{\mathcal{C}}(N_2+1)$$

[9] That is, for each $i \in \{1, 2, ..., N_2\}$, follow the transition rule $\sigma(i, b)$ with probability $\sigma(j, g)(i)$.

Since $f_n^{\mathcal{C}}(j) > \varepsilon$ for all n, this implies that the deviation creates a new state $N_2 + 1$ which is infinitely more likely than state j, which had strictly positive limit probability in the original equilibrium sequence. This implies that $f_n^{\mathcal{C}}(N_2 + 1) \to 1$ as $\rho_{1,n} \to 1$; and since P2 plays L in state $N_2 + 1$, this generates a payoff which goes to 1 against a commitment type of opponent as $\rho_{1,n} \to 1$.

So, we have constructed a deviation that changes P2's expected payoff by an amount which goes to zero as $\rho_{1,n} \to 1$ against a normal type of opponent, but by an amount which is bounded above zero against a commitment type of opponent (by the hypothesis that P2's expected payoff against type $\mathcal{C}$ stays bounded below 1 as $\rho_{1,n} \to 1$). Since $\pi >> 0$ and the deviation requires only one additional automaton state, the deviation is then strictly profitable for c_2 sufficiently small, contradicting the fact that (σ_n) is an equilibrium sequence. ■

Proof of Proposition 3:

Suppose, by contradiction, that for some sequence $\rho_{1,n} \to 1$ we can construct a corresponding convergent sequence of equilibria (σ_n) in which P1's average expected payoff stays bounded below 1 as $n \to \infty$.

By Lemma 3, player 2's average expected payoff against a commitment type of opponent goes to 1 as $n \to \infty$. For this, it must be that against an opponent who always plays G, the fraction of periods in which P2 plays L goes to 1 as $n \to \infty$. But then by fully mimicking the commitment type and playing G in every period, P1 could earn an average payoff which goes to 1 as $n \to \infty$. Therefore, he cannot be playing a best response if his average payoff stays bounded below 1, contradicting σ_n an equilibrium sequence. ■

References

[1] Abreu, Pearce, and Stachetti (1990), "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring", *Econometrica*, 58, 1041-1063

[2] Aumann, R. and S. Sorin (1989), "Cooperation and Bounded Recall", *Games and Economic Behavior*, 1, 5-39

[2] Celentani, Fudenberg, Levine, and Pesendorfer (1996), "Maintaining a Reputation against a Long-Lived Opponent", *Econometrica*, 64, 691-704

[4] Chan, J (2000), "On the Non-Existence of Reputation Effects in Two-Person Infinitely Repeated Games", working paper

[5] Cripps, Dekel, And Pesendorfer (2003), “Reputation with Equal Discounting in Repeated Games with Strictly Conflicting Interests”,working paper

[6] Cripps, Mailath, and Samuelson (2004), “Imperfect Monitoring and Impermanent Reputations”, *Econometrica*, 72, 407-432

[7] Cripps, M. and J. Thomas (1993), “Reputation and Perfection in Repeated Common Interest Games”, *Games and Economic Behavior*, 18, 141-158

[8] Fudenberg and Levine (1989), “Reputation and Equilibrium Selection in Games with a Patient Player”, *Econometrica*, 57, 759-771

[9] Fudenberg and Levine (1992), “Maintaining a Reputation when Strategies are Imperfectly Observed”, *Review of Economic Studies*, 59, 561-579

[10] Fudenberg, Levine, and Maskin (1994), “The Folk Theorem with Imperfect Public Information”, *Econometrica*, 62, 997-1040

[11] Ekmekci (2006), “Sustainable Reputations with Rating Systems”, working paper

[12] Wilson, A (200?), “Bounded Memory and Biases in Information Processing”,under review