A Note on Reputation Effects with Finite Memory

Andrea Wilson[*]and Mehmet Ekmekci[†]

December 5, 2006

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 finite memory 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.

Extremely preliminary and incomplete, please do not circulate

[*]Correspondence address: Harvard University, Economics Department; Cambridge, MA 02138. Email: wilson3@fas.harvard.edu.

[†]Correspondence address: Northwestern MEDS; email m-ekmekci@kellogg.northwestern.edu

1 Introduction

This paper studies reputation games with imperfect monitoring, when the uninformed player has a bounded memory.

We study a repeated moral hazard game, as in the following example:

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

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 $\frac{1}{2}$. We will also assume imperfect monitoring: each period, player 2 observes only a noisy public signal of player 1's action.

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.

In 1992, they showed that this result is robust to the addition of noise, in an ex ante sense: for fixed but sufficiently high $\delta_1 < 1$, player 1's average payoff as calculated at the beginning of the game is arbitrarily 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 incomplete 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. This destroys player 1's incentives to play G, so he will have an incentive to deviate from the commitment strategy. This means that in the long run, the distribution of signals will statistically identify player 1's type: hence, player 2 eventually learns that he is facing a normal type, at which point 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 about the average frequency with which player 1 chose the good action. There are only 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 use an optimal finite automaton strategy. Each period, he does observe the signal about player 1; however, he cannot recall the entire observed history, 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 he updates the rating in response to a new signal about player 1's action.

We study the long-run, steady-state equilibria of repeated moral hazard games in which (i)player 1 is a simple commitment type with probability $\pi > 0$; (ii)player 1's actions are imperfectly observed; (ii)player 2 has finite memory. Our main result is that if player 2's memory is finite but sufficiently large, then it is possible for player 1 to permanently maintain a reputation: there is an equilibrium in which player 1 expects to earn his maximal commitment payoff after any history, and player 2's automaton strategy is a best response independently of his discount factor δ_2. Unfortunately this is only a "possibility" result, there are also many equilibria in which player 1's expected payoff is significantly lower.

2 Model

We consider a simple reputation game in which two players interact repeatedly to play the following moral hazard stage game:

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

where $a - c > 0$. Thus the unique Nash equilibrium of the stage game is (B, R), as B is a dominant strategy for P1 (playing the "good action" G incurs a cost c regardless of P2's action), while player 2 chooses action L if and only if he expects G to be chosen with probability at least $\frac{1}{2}$. However, the assumption $a - c > 0$ implies that both players would benefit if P1 could commit to playing G with probability at least $\frac{1}{2}$.

It will be assumed throughout the paper that P1 is a long-run patient player, while player 2 discounts the future at rate $\delta_2 \in [0,1]$.

To allow for reputation effects, we make the following assumption:

Assumption 1: With probability $\pi > 0$, P1 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 must earn arbitrarily close to his commitment payoff $(a-c)$ in any Nash equilibrium of the repeated game.

2.1 Imperfect Monitoring

Each period, both players observe Player 2's action, but only a noisy public signal of player 1's action in the stage game. We restrict to a simple binary symmetric information structure: letting $Y \equiv \{gL, gR, bL, bR\}$ denote the set of signal realizations, the signal y is conditionally iid according to

$$\Pr\{gL|GL\} = \Pr\{gR|GR\} = \Pr\{bL|BL\} = \Pr\{BR|BR\} = \rho \in \left(\frac{1}{2}, 1\right)$$
$$\Pr\{bL|GL\} = \Pr\{bR|GR\} = \Pr\{gL|BL\} = \Pr\{gR|BR\} = 1-\rho$$

Thus player 1's action is observed correctly (g if he plays G, b if he plays B) with probability $\rho \in \left(\frac{1}{2}, 1\right)$. 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 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 $(a-c) - c\left(\frac{1-\rho}{2\rho-1}\right)$.

Fudenberg-Levine (1992) showed that for any $\pi > 0$, their 1989 reputation result is robust to the addition of noise, in the following sense: if δ_1 is below 1, then in any equilibrium of the repeated game, P1's ex ante expected payoff must be arbitrarily close to $(a-c)$ for δ_1 sufficiently close to 1.

More recently, Cripps-Mailath-Samuelson (2004) argued that reputation is nevertheless a short-run phenomenon: in any Nash equilibrium of the game with $\pi > 0$ and imperfect monitoring, 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 particular, this implies that in the long run, player 1's expected continuation payoff falls to at most $(a - c) - \frac{1-\rho}{2\rho-1}$.

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. The main result is that, as in Ekmekci (2006), this finite memory restriction allows for sustainable reputation effects: there are equilibria in which Player 1 permanently maintains a reputation for playing a strategy which would not be credible in the complete-information game. Unfortunately this is only a "possibility" result, there are also equilibria in which Player 1's average payoff is significantly below his preferred commitment payoff.

2.2 Strategies and Equilibrium

A behavior strategy for player 1 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 is the signal realization in period t (which includes player 2's action).

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

- i^0 is the initial memory state
- $\sigma : \mathcal{N} \times Y \to \Delta(\mathcal{N})$ is the transition rule, specifying how the memory $\mathcal{N} = \{1, 2, ..., N\}$ is updated after a new piece of information $y \in Y$. For $i, j \in \mathcal{N}$ and $y \in Y$, let $\sigma_{i,j}^y \equiv \sigma(i, y)(j)$ denote the probability of a transition $i \to j$ after signal realization $y \in Y$.
- $d : \mathcal{N} \to \{L, R\}$ specifies an action choice, as a function of the current memory state $i \in \mathcal{N}$.

Note that player 2's automaton strategy is required to be *stationary*: every time he is in state $i \in \mathcal{N}$, he uses the same action and transition rule. The interpretation is that player 2's memory state $i \in \mathcal{N}$ 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. We will also restrict attention to irreducible automata, and explicitly focus on equilibrium steady states.

More precisely: let $\{n, c\}$ denote the two possible types for Player 1, where n is the normal type who plays strategy γ_1, and c is the commitment type who plays G with probability 1 every period. Let f_i^c, f_i^n denote the steady-state probabilities that player 2 is in memory state i, conditional on P1 being type c, n (respectively). It is straightforward to determine the distribution conditional on type c : namely, f^c the solution to the following $N \times N$ system of equations:

$$\forall j \in \mathcal{N} : f_j^c = \sum_{\{j \in \mathcal{N} | d(j)=R\}} f_i^C \left[\rho \sigma_{i,j}^{gR} + (1-\rho)\sigma_{i,j}^{bR}\right] + \sum_{\{j \in \mathcal{N} | d(j)=L\}} f_j^C \left[\rho \sigma_{i,j}^{gL} + (1-\rho)\sigma_{i,j}^{bL}\right]$$

The distribution conditional on type n is the solution to a similar system of equations:

$$f_j^c = \sum_{\{i \in \mathcal{N} | d(i)=R\}} f_i^C \left[p_i \sigma_{i,j}^{gR} + (1-p_i)\sigma_{i,j}^{bR}\right] + \sum_{\{i \in \mathcal{N} | d(i)=L\}} f_i^C \left[p_i \sigma_{i,j}^{gL} + (1-p_i)\sigma_{i,j}^{bL}\right]$$

where p_i is the probability (long-run frequency) of a g-signal when Player 2 is in state i, conditional on type n. Then in memory state i, Player 2 believes that he is facing a commitment type with probability

$$\pi_i \equiv \frac{\pi f_i^c}{\pi f_i^c + (1-\pi) f_i^n}$$

and he expects to observe a g-signal with probability $\pi_i \rho + (1-\pi_i) p_i$.

Player 1's problem is standard: define

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

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. Say that γ_1 is a *best response to* γ_2 if for all behavior strategies γ_1' :

$$E^{(\gamma_1, \gamma_2)} \left[(1-\delta) \sum_{t=\tau}^{\infty} \delta^{t-\tau} u_1(a_1^t, a_2^t) \right] \geq E^{(\gamma_1', \gamma_2)} \left[(1-\delta) \sum_{t=\tau}^{\infty} \delta^{t-\tau} u_1(a_1^t, a_2^t) \right]$$

For Player 2: say that γ_2 is a *best response to* γ_1 if:

1. Given the action rule $d : (i^0, \sigma)$ maximizes player 2's average expected payoff,

$$\sum_{\{i \in \mathcal{N} | d(i) = L\}} [\pi f_i^c + (1-\pi) f_i^n (\alpha_i(1) + (1-\alpha_i)(-1))]$$

   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 $p_i = 1 - \rho + (2\rho - 1)\alpha_i$).

2. In each memory state i : $d(i)$ maximizes player i's expected continuation payoff, given the state-i beliefs about player 1's strategy, and using the continuation payoffs induced by (γ_1, γ_2).

This is equivalent to saying that player 2 has no incentives for one-shot deviations: given the beliefs and continuation payoffs implied by his strategy, there is no memory state in which he wishes to deviate from his prescribed action and transition rules.

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 $i \in \{1, 2\}$.

3 Sustainable Reputations

With finite memory and noisy signals, it is impossible for player 2 to be convinced that he is facing a particular type of player 1: for all $i \in \mathcal{N}$, and for any strategy pair (γ_1, γ_2), $\Pr\{C|i\}$ is bounded away from both 0 and 1.

The fact that $\Pr\{C|i\}$ 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 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 is the upper bound $\max_\gamma \max_{i \in \mathcal{N}} \Pr\{C|i\}$ on player 2's beliefs, which depends on both the ex ante prior π and on the number of memory states. Proposition 1 says that if π, N are too low, then there are no reputation effects: player 2 can never become convinced enough of the commitment type for player 1 to benefit from reputation-building:

Proposition 1: If $\sqrt{\frac{\pi}{1-\pi}} < \left(\frac{1-\rho}{\rho}\right)^{N-1}$, then there are no reputation effects:

i. If δ_2 is sufficiently close to zero, then in any NE: $\lim_{\delta\to 1} E\left[(1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}u_1(a_1^t,a_2^t)\right] \leq (a-c)-c\left(\frac{1-\rho}{2\rho-1}\right)$

ii. For any δ_2, there is a NE of the game with incomplete information in which Player 1's expected payoff is 0.

Recall that in the game with complete information ($\pi = 0$), the set of NE payoffs for player 1 is given by $[0,(a-c)-c\left(\frac{1-\rho}{2\rho-1}\right)]$ (whenever the upper bound exceeds 0). Thus Proposition 1 establishes that if π is too low, relative to N and the informativeness of the signal, then the uncertainty about player 1's type has no equilibrium effect: part (i) says that player 1's payoff cannot exceed the upper bound of the set of complete-information NE payoffs, and part (ii) says that there is an equilibrium in which player 1 expects to earn the worst possible payoff of the complete-information game, 0.

The proof is in the appendix. Note that, contrary to reputation games with unbounded players, the ex ante prior π does matter. For the range given in Proposition 1, it is impossible for player 2 to believe that he is facing a commitment type with probability greater than $\frac{1}{2}$: the proof simply shows that for any automaton, and any player 1 strategy γ_1, player 2 will always believe his opponent is most likely the normal type. At the other extreme, a very high initial prior $\sqrt{\frac{\pi}{1-\pi}} > \left(\frac{\rho}{1-\rho}\right)^{N-1}$ would imply that for any strategy pair (γ_1,γ_2), player 2 always believes that his opponent is most likely a commitment type: at this point the uniquely optimal strategy for player 2 is to play L after all histories, regardless of γ_1; and so player 1 earns expected payoff a. (He plays B every period, but nevertheless playelr 2 is convinced that he's facing a commitment type).

For the range in between, there are multiple equilibria. Proposition 2 describes the upper bound on Player 1's payoff: we show that if the signal is not too noisy, and if player 2's memory is finite but sufficiently large, then there is a NE in which player 1 earns the maximum possible commitment payoff, $a-\frac{1}{2}c$: (this is what he would obtain by publicly committing to play G with probability $\frac{1}{2}$ every period):

Proposition 2: For any $\pi > 0$,$\rho > \frac{1}{2}+\frac{1}{4}\frac{c}{a}$, and $\varepsilon > 0$, there exists $\overline{\delta_1}$,$\overline{N}$ such that whenever $\delta_1 \geq \overline{\delta_1}$ and $N \geq \overline{N}$ (but finite): there is a NE (γ_1^*,γ_2^*) such that:

i. Player 1's expected payoff satisfies $E\left[(1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}u_1(a_1^t,a_2^t)\right] \geq a-\frac{1}{2}c-\varepsilon$

ii. Player 2's strategy γ_2^* is optimal for any $\delta_2 \in [0,1]$

3.1 Sketch of proof:

Consider the following strategy γ_2 for player 2:

- In any state $i \neq 1, N$: Player 2 plays L w.p. 1, moves up $i \to i+1$ w.p. 1 after a g-signal, and down $i \to i-1$ w.p. 1 after a b-signal.
- In state N : Player 2 plays L w.p. $\frac{1}{a}\left(a - \frac{1}{2}c + \frac{1}{2}c\left(\frac{1-\rho}{2\rho-1}\right)\right)$, stays in state N after a b-signal, jumps down to state $N-2$ after a g-signal.
- In state 1: Player 2 plays L w.p. $\frac{1}{a}\left(a - \frac{1}{2}c - \frac{c(1-\rho)}{2\rho-1}\right)$, stays in state 1 after a b-signal, and moves to state 3 after a g-signal. (Note: this probability is positive for $\rho > \frac{1}{2} + \frac{1}{4}\frac{c}{a}$).

Note that all transitions in this automaton are deterministic, so as long as there are no deviations, player 1 knows player 2's memory state. Next, define V_i^1 as player 1's expected continuation payoff, conditional on (correctly believing that) player 2 being in state $i \in \mathcal{N}$. Now consider the following stationary strategy for player 1:

- As long as no deviations detected by player 2: play G with probability α_i when he is believed to be in state $i \in \mathcal{N}$
- if any deviation is detected (player 2 plays R when expected to play L, or vice versa), switch permanently to the strategy of playing B with probability 1 after any history.

We will show here that if player 2 follows the above strategy γ_2, then (i)player 1 strictly prefers to play B when player 2 is in state N, but is indifferent between playing G,B in all other states; (ii)any such strategy $(\alpha_i)_{i\in\mathcal{N}}$ with $\alpha_N = 0$ earns expected payoff $a - \frac{1}{2}c$. This implies that the specified strategy for player 1 is indeed a best response to γ_2, and earns the maximal commitment payoff $a - \frac{1}{2}c$. We show in the appendix that it is also possible to choose the α_i's such that γ_2 is a best response for player 2, for any $\delta_2 \in [0, 1]$.

So, to prove optimality for player 1: if player 2 follows γ_2, then player 1's expected continuation payoffs satisfy:

$$\lim_{\delta\to 1}\left(\frac{V_N^1 - V_{N-1}^1}{1-\delta}\right) = \frac{\left(a - \frac{1}{2}c + \frac{1}{2}c\left(\frac{1-\rho}{2\rho-1}\right)\right) - c\alpha_N - \lim_{\delta\to 1} V_N}{p_N} \tag{1}$$

$$\lim_{\delta\to 1}\left(\frac{V_{i+1}^1 - V_i^1}{1-\delta}\right) = \frac{(1-p_i)}{p_i}\lim_{\delta\to 1}\left(\frac{V_i - V_{i-1}}{1-\delta}\right) - \frac{(a - c\alpha_i - \lim_{\delta\to 1} V_i)}{p_i} \tag{3}$$

$$\lim_{\delta\to 1}\left(\frac{V_3^1 - V_1^1}{1-\delta}\right) = \frac{\lim_{\delta\to 1} V_1^1 + \alpha_1 - \left(a - \frac{1}{2}c - \frac{c(1-\rho)}{2\rho-1}\right)}{\delta p_1} \tag{3}$$

(For example, the first equation is obtained by rearranging, and taking limits as $\delta \to 1$, the following equation:

$$V_N^1 = (1-\delta)\left[a - \frac{1}{2}c + \frac{1}{2}c\left(\frac{1-\rho}{2\rho-1}\right) - c\alpha_N\right] + \delta p_N V_{N-1}^1 + \delta(1-p_N)V_N^1$$

The term in square brackets is the expected current-period payoff, and then after a g-signal (probability p_N) player 2 moves to state $N-1$, after a b-signal he stays in state N).

If we set $\alpha_N = 0$, then solving these equations yields, for all $i \in \mathcal{N}$:

$$\lim_{\delta\to 1} V_i^n = a - \frac{1}{2}c$$
$$\lim_{\delta\to 1} \frac{V_{i+1} - V_i}{1-\delta} = \frac{1}{2}\frac{c}{2\rho-1}$$

The first line implies that player 1 indeed earns his maximal commitment payoff $a - \frac{1}{2}c$ with this strategy profile.

To verify that player 1's strategy is optimal: it is clear that he must play B with probability 1 in state N, since playing G is both costly now, and yields a worse expected continuation payoff (it increases the probability of a g-signal, after which player 2 moves from state N to state $N-1$, where $V_{N-1}^1 < V_N^1$). The rest of his strategy is optimal if we can show that he is indifferent between playing G,B in all states $i \neq N$. So consider first any state $i \neq 1, N$: here, player 2 moves to state $i+1$ after a g-signal, and to state $i-1$ after a b-signal. Player 1 is then indifferent between playing G,B if the expected gain in his continuation payoff from playing G, $\delta(2\rho-1)\left(V_{i+1}^1 - V_{i-1}^1\right)$, exceeds the cost $(1-\delta)c$; this holds in the limit as $\delta \to 1$, as we have above that

$$\lim_{\delta\to 1} \frac{V_{i+1}^1 - V_{i-1}^1}{1-\delta} = \frac{c}{2\rho-1}$$

Similarly, he is indifferent between playing G,B in state 1, since player 2 stays in 1 after a b-signal, moves to 3 after a g-signal, and our automaton yields $\lim_{\delta\to 1} \frac{V_3 - V_1}{1-\delta} = \frac{c}{2\rho-1}$.

So, any strategy for player 1 which sets $\alpha_N = 0$ is a best response to γ_2, and earns the maximal commitment payoff $a - \frac{1}{2}c$. It remains to prove that we can choose $(p_i)_{i\neq N}$ such that γ_2 is a best response for player 2; this is shown in the appendix.

4 Unwritten Unappealing Results

Still working on lower bounds when both players are patient, that is δ_1, δ_2 both near 1. In the seminar, I will sketch proofs of the following possible theorems, looking for opinions on which are worth writing up:

- Assuming that public randomization is not allowed: if $\frac{\pi}{1-\pi}$ exceeds the bound in Proposition 1, then in any NE of the game with incomplete information,

$$\lim_{\delta \to 1} V^1 \geq c \cdot \frac{1-\rho}{2\rho - 1}$$

(reasonable bound for noisy signals, useless as $\rho \to 1$)

- If player 1 is also an automaton, with the same # states as player 2: his minimum payoff climbs fairly quickly from the above bound. If player 1 unbounded (but still no public randomization), can construct equilibria with the above payoff for any $\frac{\pi}{1-\pi} \leq 1$

- If we allow public randomization: no lower bound, get a folk theorem

5 Conclusion

to be written....

A Appendix

A.1 Proof of Proposition 1:

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

$$\sum_{\{i \in \mathcal{N} | d(i) = L\}} [\pi f_i^c - (1-\pi) f_i^n]$$

together with a decision rule satisfying $d(i) = L$ iff $\Pr\{C|i\} \geq \Pr\{N|i\}$. This is identical to the problem studied in Wilson (2005), which established that if $\sqrt{\frac{\pi}{1-\pi}} < \left(\frac{1-\rho}{\rho}\right)^{N-1}$, 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 $(a-c) - \frac{1-\rho}{2\rho-1}$ (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 in $\mathcal{N}$ such that the induced beliefs $\frac{\pi}{1-\pi}\frac{f_i^c}{f_i^n}$ are weakly increasing in i. Define $\tau_{i,j}^s$ as the total probability of an $i \to j$ transition conditional on type $s \in \{c, n\}$: 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 $i \in \mathcal{N}$:

$$\frac{\sum_{j \leq i-1} f^c_j \tau^c_{j,i}}{\sum_{j \leq i-1} f^n_j \tau^n_{j,i}} = \frac{f^c_i \sum_{j \neq i} \tau^c_{i,j} - \sum_{j \geq i+1} f^c_j \tau^c_{j,i}}{f^n_i \sum_{j \neq i} \tau^n_{i,j} - \sum_{j \geq i+1} f^n_j \tau^n_{j,i}}$$

Note also that $\frac{1-\rho}{\rho} \leq \frac{\tau^c_{i,j}}{\tau^n_{i,j}} \leq \frac{\rho}{1-\rho}$ for all i, j. For $i = N$, our ordering of the states implies that the LHS above is at most $\frac{f^c_{N-1}}{f^n_{N-1}} \frac{\rho}{1-\rho}$, while the RHS is at least $\frac{f^c_N}{f^n_N} \frac{1-\rho}{\rho}$: hence, we have $\frac{f^c_N}{f^n_N} \frac{1-\rho}{\rho} \leq \frac{\rho}{1-\rho} \frac{f^c_{N-1}}{f^n_{N-1}}$. Substituting this into the equation for $N-1$: the LHS above is at most $\frac{f^c_{N-2}}{f^n_{N-2}} \frac{\rho}{1-\rho}$, while the RHS is at least $\frac{f^c_{N-1}}{f^n_{N-1}} \frac{1-\rho}{\rho}$; hence, we have $\frac{f^c_{N-1}}{f^n_{N-1}} \frac{1-\rho}{\rho} \leq \frac{f^c_{N-2}}{f^n_{N-2}} \frac{\rho}{1-\rho}$. Iterating this argument:

$$\frac{f^c_N}{f^n_N} \leq \left(\frac{\rho}{1-\rho}\right)^2 \frac{f^c_{N-1}}{f^n_{N-1}} \leq \ldots \leq \frac{f^c_1}{f^n_1} \left(\frac{\rho}{1-\rho}\right)^{2(N-1)}$$

Moreover, the ordering of the states implies $\frac{f^c_1}{f^n_1} \leq 1$. (It is not possible that all states in $\mathcal{N}$ are reached more frequently conditional on c than n). Hence, for any strategy pair (γ_1, γ_2), we have:

$$\max_{i \in \mathcal{N}} \frac{\Pr\{c|i\}}{\Pr\{n|i\}} = \frac{\pi}{1-\pi} \frac{f^c_N}{f^n_N} \leq \frac{\pi}{1-\pi} \left(\frac{\rho}{1-\rho}\right)^{2(N-1)}$$

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

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

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,

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

which rearranges to

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

So for (1) to hold, we need

$$\left[\sigma^g_{i^*,i^*} - \sigma^b_{i^*,i^*}\right]\left[a - c\alpha_{i^*} - \lim_{\delta\to 1} V^1(i^*)\right] \geq c\left[\frac{\left(p_{i^*}(1-\sigma^g_{i^*,i^*}) + (1-p_{i^*})(1-\sigma^b_{i^*,i^*})\right)}{2\rho - 1}\right]$$

$$= c\left[\frac{(1-\rho)(1-\sigma^g_{i^*,i^*}) + \rho(1-\sigma^b_{i^*,i^*})}{2\rho-1} + \alpha^*_i\left(\sigma^b_{i^*,i^*} - \sigma^g_{i^*,i^*}\right)\right]$$

So,

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

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

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

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.2 Proof of Proposition 2

A.2.1 Steady-State Probabilities

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

$$i = N : f^s_N p^s_N = f^s_{N-1} p^s_{N-1}$$

$$4 \leq i \leq N-1 : f^s_i = f^s_{i-1} p^s_{i-1} + f^s_{i+1}(1 - p^s_{i+1})$$

$$i = 3 : f^s_3 = f^s_1 p^s_1 + f^s_2 p^s_2 + f^s_4(1 - p^s_4)$$

$$i = 2 : f^s_2 = f^s_3(1 - p^s_3)$$

$$i = 1 : f^s_1 p^s_1 = f^s_2(1 - p^s_2)$$

(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, $f^s_1 =$

$f_1^s(1-p_1^s) + f_2^s(1-p_2^s)$. Solving this system recurvisely yields:

$$3 \le i \le N-1 : f_i^s = \prod_{j=i}^{N-2} \frac{1-p_{j+1}^s}{p_j^s} \cdot \frac{p_N^s}{p_{N-1}^s} f_N^s$$
$$f_2^s = (1-p_3^s) f_3^s$$
$$f_1^s = \frac{(1-p_2^s)}{p_1^s}(1-p_3^s) f_3^s$$

Write all of these in terms of f_N^s and use the fact that probabilities sum to 1, to solve for f^s. Under the commitment strategy, $p_i^c = \rho \ \forall i \in \mathcal{N}$, this implies:

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

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

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

A.2.2 Optimality for Player 2

We need to choose $(p_i^n)_{i \in \mathcal{N}}$ such that:

- $\Pr\{c|N\} = \frac{1}{2}$: this implies that player 2 is indifferent between playing L, R (hence willing to randomize) in state N, 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 $2 \to 1$ 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 $p_i^n \geq \frac{1}{2}$ $\forall i \neq 1, N$, then player 2's strategy is optimal for any δ_2. To see this: in all states $i \neq 1, N$, 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, 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: 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 sufficiently high: the first condition, action indifference in state N, requires

$$\frac{\Pr\{c|N\}}{\Pr\{n|N\}} \equiv \frac{\pi}{1-\pi} \frac{f_N^c}{f_N^n} = 1 \tag{3}$$

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 n and plays G when player 2 is in state 1 is given by:

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

And the probability that player 2 is in state 1 conditional on type n is:

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

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 $\frac{1}{2}$:

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

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

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

Solving, we need:

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

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 $p_i^n = \frac{1}{2}$ $\forall i \neq 1, N$, then subsituting (1),(2) into (3),(4), we need:

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

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

where p_1^n is the *average* probability of a g-signal conditional on state 1 and type n : namely, the solution to $p_1^n = \frac{f_2^n(1-p_2)}{f_1^n}$, which at $\alpha_1^1 = \frac{1}{2}$ is $p_1^n = \frac{1}{1+2\rho-2(2\rho-1)\alpha_1^0}$. The LHS of (3a) goes to infinity as $N \to \infty$, 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 $i \neq N$ (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 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 required is strictly decreasing in both ρ and $\frac{\pi}{1-\pi}$.

This completes the proof, as optimality for player 1 was shown in the text. ∎

References

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

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

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

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

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

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

[7] Ekmekci (2006), "Sustainable Reputations with Rating Systems", working paper

[8] Wilson, A (2005), "Bounded Memory and Biases in Information Processing", still under revision