Learning under Bounded Memory:
Impermanent Reputations, but Unlimited Memory*

DANIEL MONTE†
Simon Fraser University
Department of Economics
8888 University Drive
Burnaby, B.C. Canada
V5A 1S6
Email: daniel_monte@sfu.ca
Phone: (778)-782-4501

February 11, 2009

Abstract

A recent result in reputation games is that after an arbitrarily long history, any equilibrium of the continuation game must be an equilibrium of the complete information game. We show that for a particular class of games, this result assumes that the uninformed player has infinite memory. In fact, we show that if the game is sufficiently noisy, a bounded memory player may never be able to learn anything at all. Our result implies that bounded memory can be an explanation to long-term relationships, even in the extreme case of parties with zero-sum preferences.

JEL classification: C72, D82, D83

Keywords: Bounded memory, reputation games, learning.

*This is a substantially revised version of the second chapter of my dissertation at Yale University. I am very grateful to Dirk Bergemann, Stephen Morris and Ben Polak for their support throughout this project. I also thank John Geanakoplos, Pei-yu Lo, George Mailath, and Joel Watson for comments and conversations.

†E-mail: daniel_monte@sfu.ca.

1 Introduction

An important recent result in reputation games is that after an arbitrarily long history, any equilibrium of the continuation game must be an equilibrium of the complete information game. A player may benefit from a reputation in the short-run, but this strategic use of reputation will eventually wash off. Many recent papers have shown that this convergence result holds with an impressive generality, robust to different monitoring technologies and different underlying games. Particularly important papers include Benabou and Laroque (1992), and Cripps, Mailath and Samuelson (2004).[1] This leaves open the question of how to explain reputation in long-term relationships.

We study a particular class of reputation games and show that this learning result assumes infinite memory on the uninformed player. Thus, our result shows that bounded memory can explain long-run reputation incentives. Our setting is a two-player infinitely repeated game with one-sided incomplete information and perfect monitoring. One uninformed player faces a player that, with some exogenous probability, is committed to a specified mixed strategy.

Memory is modeled as a set of states. If the player has unlimited memory, then each memory state is associated to a different history. When the player's memory is finite, his strategy is to choose an action rule, which is a map from memory states to the set of actions, but also a transition rule from state to state. At the beginning of every stage game the only information that the player has about the history of the game is his current memory state. He can then compute a best response based on the beliefs about the actual history at that point, knowing that he is forgetful across periods. We say that a memory is finite if the set of beliefs that the player can hold in equilibrium is finite. We do, however, impose sequential rationality on the equilibrium strategies. This rules out the possibility of commitment ex-ante; in particular it rules out the automata models.[2]

We think of memory constraints as a categorization procedure. A bounded memory agent can categorize the world in only a finite number of ways. This captures the fact that in reality we may not be able to distinguish very similar histories. Our view is that people categorize histories and form coarse impressions, instead of precise beliefs.

We consider the case of public beliefs in this paper. This is done for technical reasons that will become clear later in the paper. This assumption can be motivated in at least three different ways. First, a slightly modified model in which the uninformed player takes an observable continuous

[1] Kalai and Lehrer (1993), Aumann and Maschler (1995), Jackson and Kalai (1999), Sorin (1999), and Cripps, Mailath and Samuelson (2007) also prove important results on learning and reputations, but are less related to the underlying model of our paper.

[2] In Monte (2007) we discuss the difference between bounded memory and finite automaton.

action before the agent acts together with a single peaked utility function would give us the same results as in this paper– without the assumption of observable memory states. Second, one can think of this as literally being public beliefs, such as institutions that publish credit ratings. Finally, in a psychological approach, we think of this model as one in which agents are not able to dissimulate their emotions.[3]

Our main result is shown in proposition 2 in the text. We compute an upper bound on the belief spread of the bounded memory player as a function of his memory size and the exogenous mixed strategy of the commitment type. If this mixed strategy is sufficiently noisy, albeit still informative, we show that the bounded memory player will never be able to learn his opponent's true type. In fact, learning may not be possible at all if the noise is large relative to the memory size.

The intuition for our result is that with bounded memory the agent can hold only a finite number of beliefs in equilibrium. And, these beliefs cannot be too far apart from each other, or else the sequential rationality constraints would not be satisfied.

Other authors have worked on alternative explanations for permanent reputations. In a game where types are continuously changing, permanent reputation can be sustained as shown by Holmstrom (1999), and Mailath and Samuelson (2001). In a related study, Bar-Isaac (2004) showed that a model of reputation in teams can endogenously introduce this type uncertainty and thus sustain reputation.

Ekmecki (2005) showed that if the memory of the uninformed player is restricted (in the form of a finite set of ratings) then there exists a rating system (set of ratings and a transition rule) that can explain permanent reputation. The main difference between our paper and Ekmecki's is that here memory is endogenous. It is part of the uninformed player's strategy and has to satisfy sequential rationality constraints. In Ekmecki (2005) the memory process is exogenous: designed by a third party.

We proceed as follows. Section 2 describes the model. In section 3 we solve the model for the full memory case. In section 4 we define memory and describe the game when the uninformed player is restricted to a bounded memory. Section 5 is the main part of the paper where we show that under bounded memory reputation will always be sustained. We conclude in section 6.

[3]There are no reasons to believe that our results would not extend to the case of private beliefs. However, the case of private beliefs greatly complicates the analysis, so we leave it for future work.

2 Reputation Game

We study a two-player infinitely repeated game with incomplete information. One player is informed about a choice of nature and the other one is uninformed. Before the first stage game, nature draws one of two possible types for the informed player: either a commitment type (c) or a normal type (n). The commitment type is chosen with probability ρ whereas the normal type is the true type with probability $1-\rho$. The uninformed player is not aware of nature's choice.

The commitment type is playing a given mixed strategy known by both players. The normal type maximizes his payoff. The payoffs of the stage game are shown in figure 1. Players discount their repeated game payoff by a discount factor $\delta < 1$. We define $\delta = (1-\eta)$, where η is an exogenous probability that the game ends.[4]

		Informed	
		Head	Tail
Uninformed	Head	$1, -1$	$-1, 1$
	Tail	$-1, 1$	$1, -1$

Fig. 1: Matching Pennies

One interpretation for this game is the following. A policy maker is uninformed about his adviser's motives. With some exogenous probability ρ, this adviser is playing a known action, which can be thought of as giving the correct advice about some issue. This loyal adviser makes mistakes with a fixed probability (committed to a mixed strategy). The adviser may, instead, be (with probability $1-\rho$) a player with opposite preferences to the policy maker.

The action space is $\{Head, Tail\}$. A public history in this game is denoted by h and is a sequence of actions played by both players. The set of all public histories is $\mathcal{H}$. The commitment type is playing a behavioral strategy with fixed probability:

$$\bar{q} \equiv \Pr(Head|h) > \frac{1}{2},$$

for any history $h \in \mathcal{H}$, where $\bar{q}$ is common knowledge in the game. A behavioral strategy for the normal type is

$$q : \mathcal{H} \to \Delta(\{Head, Tail\}),$$

where with some abuse of notation we denote $q(h)$ the probability that the normal type will play action $Head$ given a public history h.

[4] This is done to ensure that beliefs are consistent, as will be discussed on section (4.2).

We have restricted our attention to the repeated matching pennies game, however there are no reasons to think that our results would not hold in a general game. There are, however, two main reasons for considering this particular game. First, the question of learning under bounded memory is more clear cut here. In a repeated game with incomplete information, a player with a bounded number of states faces two constraints: bounded complexity on implementing a strategy and bounded ability on updating beliefs about the actual type. The literature on automata has focused on the first issue, while we focus on the second.[5] In the setting of this paper, the bounded memory player is uninformed about the type of his opponent. Moreover, the complete information game has a unique equilibrium in the repeated game and, thus, the complexity of implementing a strategy is simple. Therefore, the issue is on updating beliefs and learning.

The second reason for choosing this game is that we want to emphasize that an infinitely renewable reputation can happen even in a world in which parties have completely opposite interests. In a general two-player game (not zero-sum) the different types may be pooling in the same equilibrium of the repeated game, in which case the question of learning looses its byte. The zero-sum nature of our game will ensure that the types will not play the same equilibrium continuation strategy in any subgame. In other words, the normal type will not mimic the commitment type forever.

3 Full Memory

In this section we show that when both players have full memory, types are revealed asymptotically. The normal type of the informed player has a current incentive for misleading the uninformed player. However, he might find it profitable to 'pretend' to be the commitment type, and explore the benefits of a higher reputation in a future period.

The uninformed player can condition his action on the entire public history of the game, so his behavioral strategy is

$$a : \mathcal{H} \rightarrow \Delta\left(\{Head, Tail\}\right),$$

where we denote $a(h)$ to be the probability that the uninformed player will play action $Head$ given a public history $h \in \mathcal{H}$. The strategy spaces for the uninformed and the informed player are denoted by Σ_a and Σ_q, respectively. Players are maximizing their repeated game expected utility, which can be written as:

$$U_i(a, q) = \sum_{t=1}^{\infty} \delta^{t-1} u_i\left(a^t, q^t\right),$$

[5] See, for example, Neyman (1985), Rubinstein (1986), Abreu and Rubinstein (1988) and Kalai and Stanford (1988) for repeated games with automata.

where player i can be the normal type or the uninformed player, $\left(a^t, q^t\right)$ is the induced action profile, and $u_i(\cdot)$ is the expected stage game payoff.

Under full memory, the uninformed player updates his beliefs using Bayes' Rule. Thus, given a strategy profile (a, q), the uninformed player's posteriors are obtained as follows. The belief of the uninformed player that his opponent is the commitment type, after some history $h \in \mathcal{H}$, is denoted as

$$\rho(h) := \Pr(c|h).$$

If the public history is h and informed player plays $Head$, the uninformed player's belief in the beginning of the following stage game is:

$$\rho_{t+1}^{H,h} = \frac{\rho_t \bar{q}}{\rho_t \bar{q} + (1-\rho_t)\, q(h)}, \tag{1a}$$

while if the informed player plays $Tail$, the uninformed player's belief is:

$$\rho_{t+1}^{T,h} = \frac{\rho_t (1-\bar{q})}{\rho_t (1-\bar{q}) + (1-\rho_t)(1-q(h))}. \tag{1b}$$

We will focus on Markovian equilibria only, since they suffice for our purposes. We define a Markov Perfect Equilibrium under full memory to be a Perfect Bayesian Equilibrium in which the strategies of both players are conditioned on the current posterior of the uninformed player only.

Definition 1 (Markov Perfect Equilibrium under Full Memory) *A strategy profile (a, q) is a Markov Perfect Equilibrium if after any history $h \in \mathcal{H}$:*

$$\begin{aligned} U_i(a, q|h) &\geq U_i\left(a', q|h\right), \ for\ \forall a' \in \Sigma_a, \\ U_i(a, q|h) &\geq U_i\left(a, q'|h\right), \ for\ \forall q' \in \Sigma_q. \end{aligned}$$

$$\begin{aligned} a(h) &= a\left(h'\right) \text{ for any } h, h' \in \mathcal{H} \text{ in which } \rho(h) = \rho\left(h'\right), \\ q(h) &= q\left(h'\right) \text{ for any } h, h' \in \mathcal{H} \text{ in which } \rho(h) = \rho\left(h'\right). \end{aligned}$$

And beliefs are computed according to (1a) and (1b).

In this game Markov Perfect Equilibrium does exist and is unique. Moreover, the repeated game expected payoff of the normal type is non-decreasing in his reputation level and the uninformed player's belief is such that $\pi(h) = \rho\bar{q} + (1-\rho)\, q(h) > \frac{1}{2}$, for any history $h \in \mathcal{H}$, (Benabou and Laroque (1992, Theorem 1)).

The reputation $\{\rho_t\}_{t\in\mathbb{N}}$ is a Markov process (as shown above in (1a) and (1b)). It depends on the sequence of actions, and on the actual type of the informed player. From the uninformed player's point of view the reputation of the opponent follows a *martingale*:

$$\begin{aligned}\mathbb{E}\left[\rho_{t+1}|\rho_t=\rho\right] &= \pi(\rho)\,\rho_{t+1}^{H}+\rho\,(1-\bar{q})+(1-\pi(\rho))\,\rho_{t+1}^{T}\\ &= \rho_t.\end{aligned}$$

However, conditional on the actual type of the opponent, the reputation evolves differently. When the actual type is a commitment type, we have that:

$$\mathbb{E}\left[\rho_{t+1}|\rho,c\right]=\bar{q}\rho_{t+1}^{H}+(1-\bar{q})\,\rho_{t+1}^{T}.$$

We use equations (1a), (1b) and also the fact that:

$$\frac{\bar{q}^2}{\pi(\rho)}+\frac{(1-\bar{q})^2}{1-\pi(\rho)}>1.$$

This implies that conditional on the actual type being commitment, the reputation of this type will evolve according to:

$$\mathbb{E}\left[\rho_{t+1}|c,\rho\right]=\rho\left\{\frac{\bar{q}^2}{\pi(\rho)}+\frac{(1-\bar{q})^2}{1-\pi(\rho)}\right\}>\rho. \tag{2}$$

The reputation tends to increase every period and is a *strict submartingale*.

We show that the evolution of beliefs on the informed player's type given that the actual type is a normal type follows, instead, a supermartingale. Before we can show this result, we first prove the following lemma.

Lemma 1 *In any Markov Perfect Equilibrium, the normal type plays Head less often than the commitment type:*

$$q(h)<\bar{q},\quad \forall h\in\mathcal{H}.$$

Proof. Suppose that there exists a history $h\in\mathcal{H}$ in which the reputation level is $\rho(h)$ and such that $q(h)>\bar{q}$. This implies that $\pi(h)>\bar{q}>\frac{1}{2}$. Also it must be that the updating is given by:

$$p^{H,h}=\frac{\rho(h)\,\bar{q}}{\pi(h)}<\rho(h),$$

whereas

$$p^{T,h}=\frac{\rho(h)\,(1-\bar{q})}{(1-\pi(h))}>\rho(h).$$

Thus, playing Tail gives the normal type a better current payoff, but also induces a higher reputation level, which implies a contradiction. ■

Conditional on the type of the informed player being normal, the reputation of this normal type will evolve according to:

$$\mathbb{E}\left[\rho_{t+1}|\rho, n\right] = q\left(\rho\right)\rho_{t+1}^{H} + \left(1 - q\left(\rho\right)\right)\rho_{t+1}^{T}.$$

Using equations (1a) and (1b) we can write this expected increase in the belief on the type being commitment given that it is actually the normal type is given by:

$$\mathbb{E}\left[\rho_{t+1}|\rho, n\right] = \rho\left\{\frac{\bar{q}q\left(\rho\right)}{\pi\left(\rho\right)} + \frac{\left(1-\bar{q}\right)\left(1-q\left(\rho\right)\right)}{1-\pi\left(\rho\right)}\right\} < \rho, \tag{3}$$

thus the belief ρ_t follows a *supermartingale* when the actual type is normal.

We use the Martingale Convergence Theorem together with (2) and (3) to show that the beliefs will converge to the correct ones.

Proposition 1 ***(Learning Under Full Memory)***
There is complete learning in this game:

$$\lim_{t\to\infty} \mathbb{E}\left[\rho_t|c\right] = 1,$$

$$\lim_{t\to\infty} \mathbb{E}\left[\rho_t|n\right] = 0.$$

For the formal proof see Benabou and Laroque (1992, p953-954). The intuition for the proof is the following. We know from (2) that $\mathbb{E}\left[\rho_{t+1}|\rho, c\right] \geq \rho$ with strict inequality if $\rho \in (0,1)$. From the martingale convergence theorem, $lim_{t\to\infty}\mathbb{E}\left[\rho_{t+1}|\rho, c\right] \to x_\infty$ for some random variable $x_\infty \in [0,1]$ with distribution $d\mu_\infty$. Then, for all period t, it must be true that:

$$\mathbb{E}\left[\rho_{t+1}|c\right] = \int_0^1 \mathbb{E}\left[\rho_{t+1}|\rho, c\right] d\mu\left(\rho\right).$$

Taking the limit as $t \to \infty$, means that we are taking the expectation for the reputation of the player conditional on the true type being the commitment type. This expectation depends on the initial reputation, which is given by some known distribution $\mu\left(\rho\right)$. As we take the limit for $t \to \infty$ we want to know the expectation of future reputation when the distribution itself converged to μ_∞. From (2) we know that $\mathbb{E}\left[\rho_{t+1}|\rho, c\right] > \rho$. Thus, for $\rho \in (0,1)$ it must be that the reputation should converge to either zero or one. It should be intuitively clear, though, that there can be no mass at 0. Thus, the reputation must converge to one. The same reasoning is true for the case of a normal type.

4 Bounded Memory

4.1 Memory and Strategies

We now study the case in which the uninformed player has bounded memory. We make no restrictions on the normal type's memory. He can recall the exact public history of the game. The uninformed player's memory is defined as a finite set of states $\mathcal{M} = \{s_1, s_2, ..., s_n\}$. The strategy of the bounded memory player is to choose a map from states to action, which we call the action rule,

$$a : \mathcal{M} \to \Delta \{Head, Tail\}.$$

Also, the bounded memory player chooses a transition from state to state

$$\varphi : \mathcal{M} \times \{Head, Tail\}^2 \to \Delta(\mathcal{M}),$$

which determines how he updates beliefs. Finally, he decides on an initial state $\varphi_0 \in \Delta(\mathcal{M})$, which is decided before he enters the first stage game. We denote $\varphi_H(i, j)$ as the probability of moving from state i to state j given that the opponent has played $Head$, regardless of his own action.[6]

As we discussed in the introduction, it will be assumed that the memory state of the uninformed player is observed by the normal type of player at every point in time. Thus, a history h of the game with a bounded memory uninformed player includes the sequence of action profiles but also the sequence of memory states of the uninformed player: $h_\tau = \{(a_t, q_t, s_t)\}_{t=1}^{\tau}$. The set of all histories of this modified game is denoted $\mathcal{H}^*$. Notice that histories are now private, only the informed player can recall the history $h \in \mathcal{H}^*$.

4.2 Beliefs

We view memory as a conscious process. A bounded memory player knows that he is forgetful. At every memory state he will hold a distribution of beliefs over the set of histories. This implies that there is an equilibrium reputation level associated to every memory state. Given that this is an infinitely repeated game, the set of possible histories in a particular memory state may be unbounded. We will *assume* that the beliefs of the bounded memory player are computed using relative frequencies, following Piccione and Rubinstein (1997).

Given a strategy profile $\sigma = (\varphi, a, q)$, the memory states form a partition of the possible histories. Let $\mu(h|s_i, \sigma)$ denote the belief of the uninformed player in state s_i that the correct history is h,

[6] The player ignores his own action in the transition rule because it does not reveal anything about his opponent's type. This is due to the simultaneous nature of the stage game. We discuss this further in section 4.3.

given the strategy profile σ. As usual, at any information set the beliefs about all histories must sum up to one

$$\sum_{h \in s_i} \mu\left(h|s_i, \sigma\right) = 1.$$

Let $f\left(h|\sigma\right)$ be the probability that a particular play of the game passes through the history h given the strategy profile σ. For each history h and memory state s_i, let the uninformed player's belief be given by the relative frequency as defined below.

Definition 2 ***(Consistency)***
A strategy profile σ is consistent with the beliefs μ if, for every memory state s_i and for every history $h \in s_i$, we have that the beliefs are computed as follows:

$$\mu\left(h|s_i, \sigma\right) = \frac{f\left(h|\sigma\right)}{\sum_{h' \in s_i} f\left(h'|\sigma\right)}. \tag{5}$$

At the beginning of a stage game, given some memory state s_i, the uninformed player's *prior* belief that his opponent is a commitment type is denoted by:

$$\rho_i \equiv \Pr\left(c|s_i, \sigma\right) = \sum_{h \in s_i} \mu\left(h|s_i, \sigma, c\right). \tag{6}$$

At the beginning of every stage game, we denote $\pi_i \equiv \Pr\left(Head|s_i, \sigma\right)$ as the probability that the informed player will play $Head$ in that stage game, given the current memory state s_i. After observing the action played by the informed player, the bounded memory player updates his belief on his opponent's type. We denote this *posterior* belief as $p_i^H \equiv \Pr\left(c|Head, s_i, \sigma\right)$ if the last action was $Head$ and $p_i^H \equiv \Pr\left(c|Head, s_i, \sigma\right)$ if the last action was $Tail$. Since the player is not forgetful within the period, but only across periods, at the end of the stage the player updates his beliefs using Bayes' rule. These posteriors are computed as:

$$\begin{aligned} p_i^H &= \frac{\rho_i \bar{q}}{\pi_i}, \\ p_i^T &= \frac{\rho_i\left(1-\bar{q}\right)}{1-\pi_i}. \end{aligned} \tag{7}$$

4.3 Sequentially Rational Equilibrium

In games with forgetfulness, an agent may visit the same information set at different points in time. Since, as usual, an agent cannot distinguish different histories in the same information set, if the player deviates at some point in time, he will not remember it, unless his memory explicitly accounts for deviations. The beliefs that the player holds at all information sets are the ones induced by the strategy profile σ.

We view memory as a conscious process. The player makes conscious decisions on how to think about his opponent. However, the solution concept for games with bounded memory is not the sequential equilibrium. Sequential equilibrium implies that at any information set, the continuation strategy is optimal for the player given his opponent's strategy and given his beliefs at that information set. In a game with bounded memory, the continuation strategy at an information set need not be optimal. The player is not able to revise his entire strategy, since he does not remember actions, or 'revised plans'.

Our concept of sequential rationality with bounded memory involves optimal actions and transitions given the beliefs induced by the strategy profile *and* taking as given the player's *own* behavior in future nodes. This concept was used by Piccione and Rubinstein (1997), Wilson (2003) and Monte (2007).[7]

We will say that the pair (σ, μ), is *sequentially rational* if two conditions hold. First, for the informed player we must have that his strategy is a best response for him given a memory rule (φ, a), after every history h. Second, for the uninformed player there are no incentives to deviate from the specified memory rule (φ, a) at any time t given the strategy q and assuming that at all other period $t' \neq t$, he was (will be) using the same strategy (φ, a).

For the first condition, we can write:

$$U_n(\varphi, a, q|h) \geq U_n(\varphi, a, q'|h) \qquad \forall q' \in \Sigma_q^*, \forall h \in \mathcal{H}^*. \tag{8}$$

Where $U_n(\varphi, a, q)$ is the expected repeated game payoff for the normal type of the informed player, given the strategy profile (φ, a, q).

To define the second condition formally, we need extra notation. For every strategy profile $\sigma = (\varphi, a, q)$ each memory state has an associated expected continuation payoff. We denote v_i^k as the expected continuation payoff for the bounded memory player at memory state s_i, given that the actual type of the informed player is $k \in \{c, n\}$. We can write these payoffs as a sum of two terms. The first term of v_i^k corresponds to the expected payoff in the stage game given that the memory state is s_i. The second term corresponds to the expected continuation payoff after the first stage game at memory state s_i, which depends on the v_i^k of all states and on the transition rule φ.

[7] Piccione and Rubinstein (1997) denoted this concept as modified multiself consistent. Wilson (2003) and Monte (2007) refer to it by incentive compatibility.

Formally we write that given a commitment type, the uninformed player's expected continuation payoff at memory state s_i is given by:

$$v_i^c = (2a_i - 1)(2\bar{q} - 1) + \delta \left(\bar{q} \sum_{j \in \mathcal{M}} \varphi_H(i,j) v_j^c + (1 - \bar{q}) \sum_{j \in \mathcal{M}} \varphi_T(i,j) v_j^c \right).$$

The expected continuation payoff given the normal type, v_i^n, must account for the entire probability distribution over the set of histories $\mathcal{H}^*$. The reason for this is that the informed player may be using a non-stationary strategy. The following result will allow us to greatly simplify the notation.

First, remember that histories are private in the game with bounded memory: only the informed player recalls the history h. Thus, the uninformed player cannot condition his action on the history, but only on the current memory state. This implies that in any equilibrium of this game, the expected continuation payoff of the normal type must depend only on the current memory state. Suppose that this is not so and that there are two histories $h, h' \in s_i \cap \mathcal{H}^*$ such that: $U_n(h) > U_n(h')$ given some equilibrium strategy $\sigma^* = (\varphi^*, a^*, q^*)$. In this case, the informed player can play a modified strategy q^{**} such that q^{**} is identical to q^* for every history different than h' or any other history h'' in which h' is a subhistory: $h' \subset h''$. Following history h', the modified strategy q^{**} plays the same continuation strategy as $q^*(h)$. Informally, this means that the informed player can always pretend to be in a different history. Therefore, we can write:

$$v_i^n = -U_n(s_i)$$

The important thing for us is to realize that v_i^n depends only on the current memory state. Thus, when deciding on which memory state to move to, the bounded memory player will 'evaluate' the memory states in a stationary way.

We define a sequentially rational strategy for the bounded memory player as follows.

Definition 3 ***(Sequential Rationality: Memory Rule)***
If a strategy (φ, a) is a sequentially rational strategy given the strategy of the informed player q, then it satisfies the following conditions. For $\forall s_i, s_j, s_{j'} \in \mathcal{M} \quad \forall k \in \{Head, Tail\}$:

$$\varphi_k(i,j) > 0 \Rightarrow p_i^k v_j^c + \left(1 - p_i^k\right) v_j^n \geq p_i^k v_{j'}^c + \left(1 - p_i^k\right) v_{j'}^n, \tag{IC1}$$

$$a_i^* = \arg\max_{a \in [0,1]} (2a - 1)\{\pi_i - (1 - \pi_i)\}. \tag{IC2}$$

The first condition, says that when taking the decision of to which memory state to move, the bounded memory player chooses the optimal state, with its associated expected payoff, given his beliefs about the opponent's type.

The second condition, (IC2), says that if $\sigma = (\varphi, a, q)$ is a sequentially rational equilibrium, then the action rule implies taking the myopic best action every stage game. Suppose $a \neq a^*$ where a^* is the optimal myopic action. If the bounded memory player deviates and play a^*, and then transition to state s_j, for example, the informed player will play a best response to this deviation. The fact that the uninformed player is forgetful implies that whenever he reaches state s_j he will assume that on-equilibrium path actions were taken. Now suppose that the informed player's best response to this deviations is more profitable than ignoring the deviation and playing the continuation strategy that he would otherwise play. Then, he could have used this best response strategy even if no deviation had taken place.

Under the multi-self interpretation, a strategy is sequentially rational if no interim self wants to deviate from the equilibrium strategy assuming that all future selves are following it, and all past selves have been following the equilibrium strategy as well. The bounded memory player can deviate from his equilibrium strategy, but he cannot revise his entire strategy. In other words, he cannot trigger a sequence of deviations.

Definition 4 ***(Sequentially Rational Equilibrium)***
The strategy profile $\sigma = (\varphi, a, q)$ is a sequentially rational equilibrium *if there exists a belief μ such that the pair (μ, σ) is* consistent *and* sequentially rational.

In this game, the question of existence is not an issue. In fact, there are typically multiple equilibria in games with imperfect recall.[8] In this game, for example, there is always an equilibrium in which all memory states induce the same belief. The action taken is the same regardless of the state and the transition rule is identical regardless of the action profile in the stage game. In this case, the bounded memory player plays a sequence of one shot games and never learns anything (as if he had no memory at all). In the next section we compute bounds on the posterior that must hold for *any* equilibrium of the game, including the most informative equilibria in which learning is maximized.

[8] See Piccione and Rubinstein (1997) and Aumann et al. (1997).

5 Learning with Bounded Memory

In a game with full memory, we say that an agent learns if his beliefs approach the correct probability distribution of states of the world. We showed that the uninformed player's beliefs about the actual type of the informed player either converges to one (if the informed player is a commitment type) or to zero (if he is a normal type).

A bounded memory player will only hold a finite number of beliefs in equilibrium. Thus, the maximum that an agent can learn is given by the extreme beliefs of this player: his highest belief and his lowest belief. We show that, under bounded memory, there is a limit on learning. We compute this bound, which is a function of the commitment type's strategy and the memory size, and show that the player may learn close to nothing if the game is sufficiently noisy.

We first show the result for irreducible memories. I.e., if a memory is such that the only ergodic set is the entire set of states, then types are never fully separated. We then show the proof for the case of a reducible memory.

The main intuition for the result is the following. The player has only a finite set of states, thus the number of posteriors after a particular action induced in equilibrium is finite. If a memory state is ever to be reached in equilibrium, the uninformed player must have had an incentive to do so. This implies that the beliefs in all memory states that are reached through the transition rule must not be "too far apart".

5.1 Irreducible memories

We will first consider only irreducible memories and leave the case of reducible memories for the following section. An irreducible memory is a set of states with no transient states (in which you leave with probability one and never return).

The proposition below is the main result of the paper. It shows that under irreducible memories there is an upper bound, function of the parameters, on the distribution of types over the memory states.

A given strategy profile σ will induce an expected continuation payoff for the uninformed player for every memory state. We label the memory states such that they are non-decreasing in the expected payoff when the true type of the informed player is c. I.e. $v_1^c \leq v_2^c \leq ... \leq v_n^c$.

Lemma 2 *Consider any sequentially rational equilibrium σ. Then, for all memory states that are reached with positive probability we must have that $v_1^n \geq v_2^n \geq ... \geq v_n^n$*

Proof. Consider two memory states $s_i, s_{i'} \in \mathcal{M}$ reached in equilibrium through σ. Suppose that $i' < i$, so that $v_i^c \geq v_{i'}^c$. If it were also true that $v_i^n > v_{i'}^n$, then for any belief $p_j^\alpha \in [0,1]$, for $\alpha \in \{Head, Tail\}$ and $\forall s_{i'} \in \mathcal{M}$, we must have that:

$$p_j^\alpha \left(v_i^c - v_{i'}^c\right) + \left(1 - p_j^\alpha\right)\left(v_i^c - v_{i'}^c\right) \geq 0,$$

which implies that state $s_{i'}$ would only be reached in equilibrium if $p_j^\alpha = 1$ for some state s_j and $v_i^c = v_{i'}^c$. However, if $p_j^\alpha = 1$, then $\rho_j = 1$ and the uninformed player is in a state where with probability one he faces a commitment type. This state is then absorbing, which contradicts irreducibility. ■

Lemma 3 *At least one memory state has a belief (weakly) lower than the original prior and at least one memory state must have a belief (weakly) higher than the prior: $\rho_i \leq \rho$ and $\rho_j \geq \rho$ for some $s_i, s_j \in \mathcal{M}$.*

Proof. Suppose that $\rho_i > \rho$, $\forall s_i \in \mathcal{M}$. Given a strategy profile $\sigma = (\varphi, a, q)$, the memory states form a partition of the set of all possible histories $\mathcal{H}$. Let $\mathcal{H}_c$ be the set of histories in which the type of the informed player is c. Similarly, $\mathcal{H}_n$ is the set of histories for which the type is n; hence, $\mathcal{H}_c \cup \mathcal{H}_n = \mathcal{H}$.

From (5) and (6) we have that:

$$\sum_{h \in s_i \cap \mathcal{H}_c} \left(\frac{f(h|\sigma)}{\sum_{h' \in s_i} f(h'|\sigma)} \right) > \rho,$$

which implies that:

$$\sum_{h \in s_i \cap \mathcal{H}_c} f(h|\sigma) > \rho \sum_{h' \in s_i} f\left(h'|\sigma\right).$$

This must hold for all s_i, thus:

$$\sum_{s_i \in \mathcal{M}} \left(\sum_{h \in s_i \cap \mathcal{H}_c} f(h|\sigma) \right) > \sum_{s_i \in \mathcal{M}} \left(\rho \sum_{h' \in s_i} f\left(h'|\sigma\right) \right),$$

summing over the expressions we have that:

$$\sum_{h \in \mathcal{H}_c} f(h|\sigma) > \rho \left(\sum_{h' \in \mathcal{H}} f\left(h'|\sigma\right) \right). \tag{9}$$

However, note that the sum of the frequency of all possible histories in the set $\mathcal{H}_c$ is $\frac{\rho}{1-\delta}$. Therefore, (9) implies that: $\frac{\rho}{1-\delta} > \rho \frac{1}{1-\delta}$, which is a contradiction.

The same reasoning applies to show that $\rho_i \geq \rho$. ■

Proposition 2 ***(Bound on Learning)***

For $\forall\rho,\delta,\bar{q},n$ *if* $\sigma=(\varphi,a,q)$ *is an equilibrium, then:*

$$\rho\left(\frac{\bar{q}}{1-\bar{q}}\right)^{n-1}\geq\rho_i\geq\rho\left(\frac{1-\bar{q}}{\bar{q}}\right)^{n-1},\ \textit{for}\ \forall s_i\in\mathcal{M}.$$

Proof. Consider only the case where memory is irreducible. Thus, there must exist $s_j, s_{j'}\in M\backslash\{s_n\}$ such that $\varphi_\alpha(n,j)>0$ and $\varphi_{\alpha'}(j',n)>0$, for some $\alpha,\alpha'\in\{Head,Tail\}$. This implies that

$$p_{j'}^{\alpha}\left(v_n^c-v_i^c\right)+\left(1-p_{j'}^{\alpha}\right)\left(v_n^n-v_i^n\right)\geq 0, \tag{10}$$

$\forall s_i\in\mathcal{M}$.

The posterior in a state s_i is given by:

$$p_i^H=\frac{\rho_i\bar{q}}{\pi_i},$$

and

$$p_i^T=\frac{\rho_i\left(1-\bar{q}\right)}{1-\pi_i}.$$

We also know that $p_i^H>p_i^T$, $\forall s_i$, since

$$\frac{\bar{q}}{\pi_n}>\frac{(1-\bar{q})}{1-\pi_n},$$

and the fact that in any equilibrium of this game, $\pi_i\geq 0.5$, for $\forall s_i\in\mathcal{M}$. [9]

Now suppose that $p_n^H>p_n^T>p_i^\alpha$, $\forall s_i\in\mathcal{M}$. Using (10) and lemma (2) we have that:

$$p_n^{\alpha}\left(v_n^c-v_i^c\right)+\left(1-p_n^{\alpha}\right)\left(v_n^n-v_i^n\right)>0,$$

for any $\alpha\in\{Head,Tail\}$ and $\forall s_i\in\mathcal{M}\backslash\{s_n\}$. This, implies that s_n is absorbing, which contradicts our assumption of irreducible memory state. Thus, there exists at least one memory state s_i for which $p_i^H\geq p_n^T$.

Since $p_i^H\geq p_n^T$, we have that: $\frac{\rho_i\bar{q}}{\pi_i}\geq\frac{\rho_n(1-\bar{q})}{1-\pi_n}$, thus:

$$\rho_i\geq\rho_n\frac{\pi_i\left(1-\bar{q}\right)}{\left(1-\pi_n\right)\bar{q}}\geq\rho_n\frac{\pi_i\left(1-\bar{q}\right)}{\left(1-\pi_n\right)\bar{q}}. \tag{11}$$

Given that the memory is irreducible, we must have that at least one state $s_{i'}$, where $i'\geq i$, is such that $p_i^T\leq p_j^H$: for $j<i'$. Using (11) we have that:

$$\rho_j\geq\rho_n\frac{\pi_i\pi_j\left(1-\bar{q}\right)^2}{\left(1-\pi_n\right)\left(1-\pi_i\right)\bar{q}^2}.$$

[9] This result is analogous to the full memory case. Suppose $\pi_i<0.5$ for some state $s_i\in M$. Then, the normal type benefits from playing *Head*, both in the stage game and in reputation, since $p_i^H>p_i^T$, and by sequential rationality, the uninformed player will move to a state that would give him a weakly higher payoff if his opponent is the commitment type.

We then repeat the procedure above and note that the highest upper bound for the lowest belief ρ_i is given when the transition occurs without any jump and such that $\rho_1 \leq \rho_2 \leq ... \leq \rho_n$ (thus, $\rho_n \geq \rho$). This lowest belief is such that:

$$\rho_1 \geq \rho_n \frac{\pi_{n-1}\pi_{n-2}...\pi_1 (1-\bar{q})^{n-1}}{(1-\pi_n)(1-\pi_{n-1})...(1-\pi_1)\bar{q}^{n-1}}.$$

We again use the fact that in any equilibrium of this game, $\pi_i \geq 0.5$, for $\forall s_i \in \mathcal{M}$ to obtain the following inequality:

$$\rho_1 \geq \rho \left(\frac{1-\bar{q}}{\bar{q}}\right)^{n-1}. \tag{12}$$

To obtain an upper bound on the highest belief we use the same reasoning as above, but in the opposite direction. I.e., for an irreducible memory we must have that $p_1^H \geq p_i^T$ for some state $s_i \in \mathcal{M} \backslash \{s_1\}$. Similarly, $p_i^H \geq p_{i+1}^T$. This gives us the following inequality:

$$\rho_n \leq \rho_1 \left(\frac{\bar{q}}{1-\bar{q}}\right)^{n-1} \frac{(1-\pi_n)(1-\pi_{n-1})...(1-\pi_2)}{\pi_{n-1}\pi_{n-2}...\pi_1},$$

which implies that:

$$\rho_n \leq \rho \left(\frac{\bar{q}}{1-\bar{q}}\right)^{n-1}. \tag{13}$$

■

We know from (12) and (13) that $\rho \left(\frac{\bar{q}}{1-\bar{q}}\right)^{n-1} \geq \rho_i \geq \rho \left(\frac{1-\bar{q}}{\bar{q}}\right)^{n-1}$, for $\forall s_i \in \mathcal{M}$. Thus:

$$\lim_{\bar{q} \to \frac{1}{2}} \rho_i = \rho.$$

Corollary 1 ***(No Learning when Game is Noisy)***
For any memory size n, and any $\varepsilon > 0$, there exists a $\bar{q} > 0.5$ such that any equilibrium in the game is such that:

$$\rho + \varepsilon \geq \rho_i \geq \rho - \varepsilon, \ \textit{for } \forall s_i \in \mathcal{M}.$$

Remark 1 *Note that under full memory, the result on learning by the uninformed player holds even for $\bar{q}$ arbitrarily close to 0.5 (as long as $\bar{q} > 0.5$).*

Corollary 2 ***(No Learning when Memory is Small)***
For any $\bar{q} > 0.5$, and any $\varepsilon > 0$, there exists a memory size $\bar{n}$ such that any equilibrium in a game in which the uninformed player is restricted to a memory of size $n < \bar{n}$ is such that:

$$\rho + \varepsilon \geq \rho_i \geq \rho - \varepsilon, \ \textit{for } \forall s_i \in \mathcal{M}.$$

5.2 Reducible case

We now turn to the reducible case. Our result is that the uninformed player cannot learn more when using a reducible memory than when using an irreducible one. The intuition for this is that there are less states to "dilute" the posteriors. We prove the result in this section.

Consider a memory with k recurrent classes $\mathcal{R}_1, \mathcal{R}_2, ..., \mathcal{R}_k$ and a set of transient states $\mathcal{T}$. We want to show that in this case there is also a bound on learning and, in fact, it is smaller than in the irreducible case.

Proposition 3 ***(Bound on Learning: the Reducible Case)***
For $\forall \rho, \delta, n$ if $\sigma = (\varphi, a, q)$ is an equilibrium in the $n-$memory state game with k recurrent classes and one transient state, then:

$$\rho \left(\frac{\bar{q}}{1-\bar{q}} \right)^{n-1} \geq \rho_i \geq \rho \left(\frac{1-\bar{q}}{\bar{q}} \right)^{n-1}, \text{ for } \forall s_i \in \mathcal{M}.$$

Proof. For every recurrent class $\mathcal{R}_i$ let s_i be the memory state for which p_i^T is lower. Also denote s_h as the memory state in the transient class that has the highest posterior, denoted $p_{\mathcal{T}}^H$, among all states that reach some state $s_{i'} \in \mathcal{R}_i$ with positive probability:

$$s \in \{ s_h \in \mathcal{M} | \varphi_\alpha (s_h, s_{i'}) > 0, \forall \alpha \in \{Head, Tail\}, s_{i'} \in \mathcal{R}_i \}.$$

From an argument identical to the one in lemma (3), which we omit here, it must be the case that $p_{\mathcal{T}}^H \geq p_i^T$. Informally, we say that this is true since all the mass of normal types that enter a recurrent class will stay there forever. Similarly, if $p_{\mathcal{T}}^T$ is the lowest posterior in $\mathcal{T}$, and p_j^H is the highest posterior in some recurrent class, then: $p_j^H \geq p_{\mathcal{T}}^T$.

The lowest bound is achieved (or highest possibility of learning) when every state in the transient class connects to each other and every state in the recurrent classes also connect to each other. Using the same reasoning as in proposition 2 we have that the lower bound on the uninformed player's belief is given by:

$$\rho_i \geq \rho \left(\frac{1-\bar{q}}{\bar{q}} \right)^{\sum_{i=1}^k n_i + n_{\mathcal{T}} - 1},$$

where n_i is the number of memory states in $\mathcal{R}_i$ that are reached through each other and $n_{\mathcal{T}}$ is the number of memory states in $\mathcal{T}$ that are reached through each other. This number is smallest when $\sum_{i=1}^k n_i + n_{\mathcal{T}} = n$: the bound is smallest when all the states communicate with each other.

Similarly, the upper bound on the belief of the uninformed player is given by:

$$\rho_i \leq \rho \left(\frac{\bar{q}}{1-\bar{q}} \right)^{\sum_{i=1}^k n_i + n_{\mathcal{T}} - 1},$$

which achieves its maximum when $\sum_{i=1}^k n_i + n_{\mathcal{T}} - 1 = n$. ■

6 Conclusion

A celebrated recent result in the literature on reputation and repeated games with incomplete information is that the play of the game converges asymptotically to the play of a complete information game. This means that players can profit from a "false" reputation only in the short-run. Constant opportunistic behavior will lead to statistical revelation of the actual type, which means no long-run reputation.

Thus, Cripps, Mailath and Samuelson (2004, p409) conclude that: "We view our results as suggesting that a model of long-run reputations should incorporate some mechanism by which the uncertainty about types is continually replenished." This leaves an open question in the study of reputation games: how to explain long-term relationships when preferences are not changing over time?

We show that under bounded memory we may not have learning (or type separation) even in the long-run. Therefore, bounded memory on the uninformed player can explain how long-term reputations can be sustained, even in the extreme case where agents have opposite preferences. In fact, players can learn close to nothing if their memory is small enough compared to the noise (commitment type's mixed strategy) in the game.

The recent results on reputations and long-term relationships are shown to be robust to different underlying games and different monitoring technologies. From what we show in this paper, though, it may not be robust to cognitive constraints on the individuals.

References

[1] D. Abreu, A. Rubinstein, "The Structure of Nash Equilibrium in Repeated Games with Finite Automata," *Econometrica*, 56, 1259-1281, (1988).

[2] R. Aumann, S. Hart, and M. Perry, "The Absent-Minded Driver," *Games and Economic Behavior*, 20, 102-116, (1997).

[3] R. J. Aumann, M. B. Maschler, and R. E. Stearns. "*Repeated games with incomplete information,*" MIT Press, Cambridge, (1995).

[4] H. Bar-Isaac, "Something to prove: Reputation in teams" *RAND Journal of Economics*, Volume 38(2), 495-511, (2007).

[5] R. Benabou, and G. Laroque, "Using Privileged Information to Manipulate Markets: Insiders, Gurus, and Credibility," *Quarterly Journal of Economics*, 921-58, (1992).

[6] M. Cripps, G. Mailath and L. Samuelson, "Imperfect Monitoring and Impermanent Reputation," *Econometrica*, March, 72.2, 407-432, (2004).

[7] M. Cripps, G. Mailath and L. Samuelson, "Disappearing Private Reputations," *Journal of Economic Theory*, (2007).

[8] M. Ekmecki, "Sustainable Reputations with Rating Systems," job market paper, Princeton University, (2005).

[9] D. Fudenberg and D. Levine, "Reputation and Equilibrium Selection in Games with a Patient Player," *Econometrica*, July, 57.4, 759-778, (1989).

[10] D. Fudenberg and D. Levine, "Maintaining a Reputation When Strategies Are Imperfectly Observed," *Review of Economic Studies*, July, 59.3, 561-579, (1992).

[11] B. Holmstrom, "Managerial Incentive Problems - A Dynamic Perspective," *Review of Economic Studies*, January, 66.1, 169-182, (1999).

[12] M. Jackson and E. Kalai, "Reputation versus Social Learning," *Journal of Economic Theory*, 88(1), 40-59, (1999).

[13] E. Kalai and E. Lehrer, "Rational Learning Leads to Nash Equilibrium," *Econometrica*, vol. 61, No. 5, 1019-1045, (1993).

[14] E. Kalai, W. Stanford, "Finite Rationality and Interpersonal Complexity in Repeated Games," *Econometrica*, 56, No. 2, 387-410, (1988).

[15] G. Mailath, and L. Samuelson, "Who Wants a Good Reputation?," *Review of Economic Studies*, 68, 415-441, (2001).

[16] D. Monte, "Reputation and Bounded Memory: A Theory of Inertia and Skepticism," Simon Fraser University, (2007).

[17] A. Neyman, "Bounded Complexity Justifies Cooperation in the Finitely Repeated Prisoner's Dilemma," *Economic Letters*, 19, 227-229, (1985).

[18] M. Piccione, and A. Rubinstein, "On the Interpretation of Decision Problems with Imperfect Recall," *Games and Economic Behavior* 20, 3-24, (1997).

[19] A. Rubinstein, "Finite Automata Play the Repeated Prisoners' Dilemma," *Journal of Economic Theory*, 39, 83-96, (1986).

[20] J. Sobel, "A Theory of Credibility," *Review of Economic Studies*, LII, 557-573, (1985).

[21] S. Sorin, "Merging, Reputation, and Repeated Games with Incomplete Information," *Games and Economic Behavior*, 29, 274-308, (1999).

[22] A. Wilson, "Bounded Memory and Biases in Information Processing", job market paper, Princeton University, (2003).