IAS School of Social Science

ECONOMICS WORKING PAPERS

On interim rationality, belief formation and learning in decision problems with bounded memory

Pathikrit Basu **Kalyan Chatterjee**

Paper Number 110

On interim rationality, belief formation and learning in decision problems with bounded memory*

Pathikrit Basu[†] Kalyan Chatterjee[‡]

September 24, 2015

Abstract

We consider decision problems where the agent has bounded memory and makes decisions using a finite state automaton. For any automaton, we show there exist equivalent automata with minimal randomization. We introduce a new criterion for interim rationality called *weak admissibility* and demonstrate that if memory states are not wasted, the nature of transitions are uniquely defined by the criterion. We show for any optimal automaton, there exists an equivalent automaton that is weakly admissible and involves no wastage of memory states. We describe more general environments (POMDP) and establish a *revelation principle* for optimality. Under a changing worlds environment, we discuss conditions under which unconstrained optimum can be attained using finite state automata. In these environments, we demonstrate characteristics of the constrained optimal automaton.

1 Introduction

In standard state space models of uncertainty, rationality assumes that decision makers apply Bayes' rule to update beliefs and undertake actions that are optimal under the beliefs held. An assumption implicit in such models is that agents are able to process information with unbounded capacity. However, in reality, available information may be too detailed to process completely and a simplification or an endogenous coarsening of it is performed to

*This version is preliminary and incomplete. We are grateful to Daniel Monte, Stephen Morris and Ariel Rubinstein for comments and discussions. Chatterjee also thanks the Richard B. Fisher Endowment at the Institute for Advanced Study, Princeton, NJ, for making his membership of the institute in 2014-15 possible.
[†]Pennsylvania State University. E-mail - pub143@psu.edu
[‡]Pennsylvania State University. E-mail - kchatterjee@psu.edu

keep the act of decision making simple. Agents have bounds to understanding and assimilating new information and *given* these bounds, contemplate optimal actions. A natural way to incorporate this feature in standard models would be to impose the bound on the agent first and then study constrained rational behaviour. The current paper is concerned precisely with this line of study. We consider standard dynamic environments where the agent's objective is to learn an unobservable payoff relevant parameter based on information (signals) that arrives at every point in time. To impose the constraint on processing ability, we assume that agents use a finite state automaton to perform inference and take actions. We study optimal behaviour under this framework and moreover, the characteristics of the optimal decision making process namely the nature of decision making and belief formation.

In dynamic environments, one important issue concerning optimal plans is that of time consistency. In the present model, the agent using an automaton finds himself in a decision problem with imperfect recall. Piccione and Rubinstein (1997)[18] point out that in such problems, optimal plans may not be time consistent but satisfy a weaker notion of time consistency that they call *modified multiself consistency*. It requires that agents not have profitable one shot deviations from the chosen strategy at any information set. If we interpret the problem as a game with information sets treated as seperate players with a common interest, modified multiself consistency is equivalent to an equilibrium of the game. In a framework with agents using a finite state automaton as in the present context, Wilson (2014)[23] establishes a natural analog of it in a model of simple hypothesis testing and Kocer (2010)[14] establishes in more general environments concerning partially observable Markov decision processes. We introduce a weaker criterion for interim rationality of an automaton that we call "weak admissibility". The criterion requires that at any memory state reached with positive probability, commiting to transition probabilities in one payoff relevant state of the world (hypothesis), the agent under the constraint created by the commitment, necessarily chooses transitions optimally for the other state of the world. One feature of this requirement is that it does not involve beliefs over the state of the world but rather a very simple necessary condition in terms of dominance. If additionally, states are not wasted (in the sense that for any two memory states reached with positive probability, the continuation values differ and neither pair of continuation values dominates the other in terms of weak dominance), then the weak admissibility condition implies that the transitions are uniquely defined and exhibit a simple "stair-case" structure. This is achieved by interpreting the weak admissibility condition as a linear program, which can be uniquely solved using a proposed algorithm which assigns transition probabilities to memory states by ranking the product of continuation values of memory states and likelihood ratios of signals. Additionally, with regard to optimality, it is shown that for any optimal automaton, there exists an equivalent automaton with the same ex-ante value that is weakly admissible and does not waste memory states.

Interpreting the discount factor as a stopping probability as in Kocer(2010), we establish a *revelation principle* for memory states reached with positive probability which states that at any memory state reached with positive probability, it is indeed optimal to stay there before any further information is received via signals under the beliefs held about the state of the world. In a generalised (compared to Wilson (2014)) investment problem with *changing worlds* (the state of the world evolves non-interactively according to a Markov chain), this result proves to be very useful in studying characteristics of the optimal automaton and in particular the nature of transitions taking place in it. In particular, it guarantess *categorisation* of beliefs in the sense that posterior beliefs lie in one of finitely many categories indexed by memory states of the automaton. Each category is exactly the set of beliefs for which its corresponding memory state is optimal. The optimal transitions can then be shown to follow simple rules depending on the signal. With each signal we associate a belief that is the fixed point of an appropriate map and transitions take place depending on whether the current belief is placed left or right to the fixed point. We also demonstrate conditions under which the unconstrained optimal decision rule can be implemented by a deterministic finite state automaton. As in Monte and Said (2014), this demonstrates the lack of need for any sophistication in the decision rule in special environments. However, unlike their model, the model studied here uses discounting, a more general signalling structure and admits an unconstrained optimal memory of more than 2 states. As an example, we also demonstrate that a simple variant (investment problem) of the framework in Shiryaev's change point detection problem (See Shiryaev (1978)[21]) allows for optimal unconstrained finite memory under simple conditions. The belief process without bounds on information processing follows an ergodic Markov process and admits a unique stationary distribution. Lastly, on a seperate but relevant note, we show that any automaton can be modified to one which gives the same performance but significantly reduces the extent of randomization involved. We derive an upper bound on randomisation effectively required and show that if the signal space is large enough, the automaton is "close" to a deterministic one. In the context of the investment problem with unchanging worlds, it is shown that requiring weak admissibility tightens the upper bound and further reduces the extent of randomisation involved.

1.1 Related Literature

The strand of literature most related to our work pertains to the issue of learning with a finite state automaton. It's starting point is marked by a seminal paper due to Hellman and Cover (1970)[9] which solves the simple hypothesis problem with finite memory. In the field of economics, as stated above, Wilson (2014), Kocer(2010), Monte and Said (2014) study related problems. Additionally, Kocer(2014) studies a bandit problem with finite memory and Monte (2007) studies a repeated sender receiver game with the receiver using a finite state automaton. Piccione and Rubinstein(1997), Aumann, Hart and Perry (1997) and Halpern(1997) discuss the notion of belief consistency in decision problems with imperfect recall. Shiryaev (1978) solves the change point detection problem but without memory constraints. Kalai and Solan (2003) discuss the need for randomisation in simple plans (automata) and establish that in non-interactive environments, randomisation in the decisions is not needed but randomisation might be required in the transitions of the optimal automaton

Other stands of the literature that analyse environments with decision making under complexity constraints are relevant but not directly related. Rubinstein(1986), Abreu and Rubinstein(1988) and Chatterjee and Sabourian(2002) consider multistage games restricting agents to use of finite state automata and study *Nash equilbria with complexity costs*. Rubinstein(1986), Abreu and Rubinstein(1988) analyse two person repeated games whereas Chatterjee and Sabourian(2002) study the dynamic n-person unanimity bargaining game.

Dow (1996)[6] studies a single person problem of a consumer searching for a low price but has limited memory to remember prices observed. Recent literature on coarse information and categorisation include Al-Najjar and Pai (2013), Peski (2011), Jackson and Fryer (2005) and Daskalova and Vriend (2015). However, these models are all based in static environments and differ in terms of contexts and approaches considered within them.

1.2 Outline

The outline of the paper is as follows : Section 2 describes an investment problem and introduces decision rules using finite state automata, optimality and modified multiself consistency. Section 3 discusses the issue of randomisation required in an automaton. Section 4 introduces the criterion of weak admissibility in relation to optimality and the extent of randomisation involved. Lastly, section 5 introduces the model with changing worlds and we study characteristics of unconstrained and constrained optimal finite state automata.

2 Investment Problem

Every period, i.i.d payoffs are realised from an unknown distribution $\theta \in \{0,1\}$. The sample space $X \subset \mathbb{R}$ is finite and its elements are interpreted as payoffs. Under $\theta = 0$, payoffs are realised according to distribution F_0. Under $\theta = 1$ they are realised according to F_1. The distributions are assumed to have density f_0, f_1. We shall assume for convenience that the pair of distributions satisfies the monotone likelihood ratio property i.e $\frac{f_0(x)}{f_1(x)}$ is strictly increasing in x[1]. It is assumed that the expectations for the two distributions satisfy $\mathbb{E}_0(x) > 0$ and $\mathbb{E}_1(x) < 0$. We shall write them as E_0 and E_1. We can interpret f_0 as a *good* distribution yielding a positive payoff on average. Similarly, we interpret f_1 as a *bad* distribution. Furthermore, we shall refer to $\theta = 0$ as the good state of the world and $\theta = 1$ as the bad state of the world. Lastly, let $\pi \in (0,1)$ be the prior probability that the true distribution is $\theta = 0$.

There is a single decision maker to whom θ is unknown. Every period, before the payoffs are realised, he chooses (based on available information) one of two actions in the set $\{I, NI\}$. If I is chosen and x is realised, the decision maker gets a payoff x in that period. If NI is chosen, then he gets 0. We assume that the signal is observed even when the decision NI is undertaken. Action I is interpreted as a decision to *invest* and NI as a decision to *not invest*. If the actual realisation is $(x_t)_t \in X^\infty$ and the decision path along $(x_t)_t$ is $(d_t)_t \in \{I, NI\}^\infty$, then the decision maker gets payoff :

$$\sum_{t=0}^{\infty}(1-\delta)\delta^t x_t \mathbb{I}_{\{d_t = I\}}$$

Where $\mathbb{I}$ is the indicator function which takes value 1 when $d_t = I$ and 0 when $d_t = NI$. The decision maker wishes to choose a decision plan to maximise discounted expected utility. Note that in doing so, it will be relevant for the decision maker to make inferences about θ using the payoff realisations x as signals. In this paper, we shall consider decision makers who use finite state automata to makes decisions and inferences.

2.1 Equivalent formulation of the problem

The problem posed above can be shown equivalent to a decision problem which involves investment decisions $a \in \{I, NI\}$ every period with state dependent payoffs being realised every period according to utility function $u(a,\theta)$ every period. The utility values are such that $u(NI,\theta) = 0$ and $u(I,0) = E_0 > 0 > E_1 = u(I,1)$. Information from signals $x \in X$

[1] This assumption is introduced for simplicity of exposition and is not crucial for obtaining results

is available to the decision maker every period. The decision problem then becomes to formulate a strategy to maximise long run discounted expected value of payoffs given by u :

$$\mathbb{E}(\sum_{t=0}^{\infty}(1-\delta)\delta^t u(a_t,\theta))$$

The above simpler formulation shall be used later as well when we focus on an environment with changing worlds where the state of the world θ is not determined once and for all but may change randomly according to a Markov process. One may consider the above problem to be an extreme case of the environment with state change : the state switches with probability zero.

2.2 Finite state automaton

An automaton is defined as a tuple $\tau =< M, X, \sigma, d, i_0 >$ where $M = \{1, ..., m\}$ is a finite set of *memory states*, $\sigma : M \times X \rightarrow \Delta(M)$ is a *transition function*, $d : M \rightarrow \{I, NI\}$ is a *decision function* and i_0 is the initial state. The interpretation of the automaton is as follows : At a given period, the agent at state i decides whether to invest or not invest determined by $d(i)$. If he observes a payoff realisation x, he updates the state randomly accordingly to $\sigma(i, x)$ and depending on the new state decides whether or not to invest according to d. Denote the ex-ante expected value from the automaton τ to be $V(\tau)$. This can be derived as follows :

- The transition function $\sigma : M \times X \rightarrow \Delta(M)$ defines two transition matrices on the state space M, $(\alpha_{ij}), (\beta_{ij})$ as

  $\alpha_{ij} = \sum_{x \in X} \Pr(\sigma(i,x) = j) f_0(x)$

  and

  $\beta_{ij} = \sum_{x \in X} \Pr(\sigma(i,x) = j) f_1(x)$

  Let their stationary distributions be denoted as μ_0, μ_1.

- Derive the values V_{i0}, V_{i1}. These are the long run expected payoffs from state i when $\theta = 0$ and $\theta = 1$. Then for $i \in M$,we have :

  $V_{i0} = E_0.\mathbb{I}_{\{d(i)=I\}} + \delta \sum_{j \in M} \alpha_{ij} V_{j0}$

  $V_{i1} = E_1.\mathbb{I}_{\{d(i)=I\}} + \delta \sum_{j \in M} \beta_{ij} V_{j1}$

- $V(\tau) = \pi V_{i_0 0} + (1-\pi) V_{i_0 1}$

An optimal automaton is one which solves the maximisation problem :

$$\max_{\tau} V(\tau)$$

Here the optimal automaton is chosen fixing the size i.e the set of states M. Notice that we focus on automata with a deterministic initial state and decision function. Even if one were to allow for randomisation, it can be shown that there exists an optimal automaton with a deterministic initial state and decision function. Hence, we can focus on this class of automata. An automaton with randomisation would be defined as $\tau =< M, X, \sigma, d, g_0 >$ where $M = \{1, ..., m\}$ is a finite set of states, $\sigma : M \times X \to \Delta(M)$ is a transition function, $d : M \to \Delta\{I, NI\}$ is a decision function and $g_0 \in \Delta(M)$ is the initial randomisation over states.

Proposition 1 *There exists an optimal automaton with a deterministic initial state and decision function.*

Proof : It can be established that there always exists an optimal automaton with randomisation. Let τ be optimal. Denote as $\Pr^{\theta,t}(i)$, the probability of the automaton τ, entering i at time t under hypothesis θ. The value generated by the automaton if the state of the world is $\theta \in \{0, 1\}$ is :

$$\begin{aligned} &\sum_{t=0}^{\infty}(1-\delta)\delta^t \sum_{i\in M} E_\theta d(i) \Pr^{\theta,t}(i) \\ &= \sum_{t=0}^{\infty}\sum_{i\in M}(1-\delta)\delta^t E_\theta d(i) \Pr^{\theta,t}(i) \\ &= \sum_{i\in M} d(i) \sum_{t=0}^{\infty}(1-\delta)\delta^t E_\theta \Pr^{\theta,t}(i) \end{aligned}$$

Hence the ex-ante value is :

$$\begin{aligned} &\pi \sum_{i\in M} d(i) \sum_{t=0}^{\infty}(1-\delta)\delta^t E_0 \Pr^{0,t}(i) + (1-\pi) \sum_{i\in M} d(i) \sum_{t=0}^{\infty}(1-\delta)\delta^t E_1 \Pr^{1,t}(i) \\ &= \sum_{i\in M} d(i)[\pi \sum_{t=0}^{\infty}(1-\delta)\delta^t E_0 \Pr^{0,t}(i) + (1-\pi) \textstyle\sum_{t=0}^{\infty}(1-\delta)\delta^t E_1 \Pr^{1,t}(i)] \end{aligned}$$

defining $e(i) := \pi \sum_{t=0}^{\infty}(1-\delta)\delta^t E_0 \Pr^{0,t}(i) + (1-\pi)\sum_{t=0}^{\infty}(1-\delta)\delta^t E_1 \Pr^{1,t}(i)$, the above expression becomes :

$$\sum_{i \in M} d(i)e(i)$$

Now since the above expression is linear in $d(i)$ and τ is optimal, it must be the case that $d(i) = 1$ if $e(i) > 0$ and $d(i) = 0$ if $e(i) < 0$. Hence, the following decision rule is optimal is optimal as well :

$$d'(i) = d(i).\mathbb{I}_{e(i)\neq 0}$$

Hence, the automata $\tau' = \tau \backslash \{d\} \cup \{d'\}$ is optimal. Now under τ', let the value from state i under hypothesis θ be denote as $V_{i\theta}$. Then the ex-ante value of the automaton is :

$$\sum_{i \in M} g_0(i)[\pi V_{i0} + (1-\pi)V_{i1}]$$

Let $i_o \in \arg\max_{i \in M} \pi V_{i0} + (1-\pi)V_{i1}$. Since τ' is optimal, the initial randomisation g_0 must be optimal and hence from the above expression it is clear that $V(\tau') = \pi V_{i_o 0} + (1-\pi)V_{i_o 1}$. Hence, then automaton $\tau'' = \tau \backslash \{g_0\} \cup \{\delta_{i_o}\}$ is optimal (here $\delta_{i_o} \in \Delta(M)$ is the delta measure on i_o) and is an automaton with deterministic initial state and decision function.

■

2.3 Modified Multiself Consistency

We discuss here the notion of modified multiself consistency. The automaton places the decision maker in a decision problem of imperfect recall and modified multiself consistency is an interim rationality requirement as introduced by Piccione and Rubinstein (1997) [18]. In order to apply it here, one would require defining beliefs at each memory state of the automaton about the state of the world. We interpret the discount factor as a continuation probability. This is as follows : At the beginning of every period, a coin of bias δ is tossed, if tail comes up (with probability $1-\delta$), then the decision problem terminates and the cumulative payoff $\sum_{\tau=0}^{t} x_t.\mathbb{I}_{\{d_t=1\}}$ is obtained. Under this interpretation, the probability of entering state i of the automaton under hypothesis θ is the following:

$$\hat{\mu}^\theta(i) = \sum_{t=0}^{\infty}(1-\delta)\delta^t \Pr^{\theta,t}(i)$$

where $\Pr^{\theta,0}(i)$ is given by the initial distribution over the states of the automaton (deterministic and independent of θ) and $\Pr^{\theta,t}(i)$ is the probability of being in state i at time t under the Markov chain induced by the automaton in hypothesis θ (these are $(\alpha_{ij}), (\beta_{ij})$ as defined before). It has been shown by Wilson(2014,[23]) that $\hat{\mu}^0$ and $\hat{\mu}^1$ are the respective stationary distributions of the following perturbed Markov chains :

$$w^0_{ij} = (1-\delta)\Pr^{0,0}(j) + \delta\alpha_{ij}$$
$$w^1_{ij} = (1-\delta)\Pr^{1,0}(j) + \delta\beta_{ij}$$

When the agent is at state i, we use the stationary distributions to define posteriors $(\hat{\pi}(i))$ about the state of the world θ by :

$$\hat{\pi}(i) = \frac{\pi\hat{\mu}^0(i)}{\pi\hat{\mu}^0(i) + (1-\pi)\hat{\mu}^1(i)} \tag{1}$$

When the agent is state i and receives signal x, his posterior, generated by the aggregate signal (i, x) is defined as :

$$\hat{\pi}(i,x) = \frac{\pi\hat{\mu}^0(i)f_0(x)}{\pi\hat{\mu}^0(i)f_0(x) + (1-\pi)\hat{\mu}^1(i)f_1(x)} \tag{2}$$

We can now define the notion of modified multiself consistency of an automaton. We do so for both notions of belief consistency defined above :

Definition An automaton is said to be *modified multiself consistent* if for all $i \in M$ and $x \in X$:

- If $\Pr(f(i,x) = j) > 0$ then

  $\hat{\pi}(i,x)V_{j0} + (1-\hat{\pi}(i,x))V_{j1} \geq \hat{\pi}(i,x)V_{k0} + (1-\hat{\pi}(i,x))V_{k1}$

  for all $k \in M$.

- $d(i) = 1$ if

  $\hat{\pi}(i)E_0 + (1-\hat{\pi}(i))E_1 > 0$

  and $d(i) = 0$ if

  $\hat{\pi}(i)E_0 + (1-\hat{\pi}(i))E_1 < 0$

It has been shown in Wilson(2014) that optimal automata are modifed multiself consistent when considering beliefs defined by (1) and (2). Kocer(2010) establishes modified multiself consistency for environments where agents face partially observable Markov decision processes. One may note that the above investment problem is an example of a POMDP. As can be seen there modified multiself consistency allows one to see how the transitions take place. In section 4 below, we shall consider a weaker requirement which uniquely describes the transitions.

It is important to note that the right notion of modifed multiself consistency ought to depend on the interpretation of the discount factor in the model. If the discount factor δ is assumed as a factor that is used to measure next period payoffs from the point of the view of the current period, it seems implausible to consider it as a continuation probability that enters the extensive form of the environment. In this case, the stationary distributions of $\{\alpha_{ij}\}$ and $\{\beta_{ij}\}$ should be used to define beliefs. However, in the definition of the environment, if the discount factor is indeed a continuation probability, beliefs should correspond to ones defined by (1) and (2). From a purely technical standpoint, both notions are well defined and may be used in defining interim rationality. The use of the right notion depends on the specification of the environment.

3 Limits to randomization

The issue of randomisation in an automaton has been studied by Kalai and Solan (2003) and Hellman and Cover (1971). The former study the need for it in the optimum and the latter establish that in certain learning problems, a two-state automaton with randomisation yields a payoff higher than any deterministic automaton of a fixed size m. In this section we define a measure of randomisation and show that any automaton can be modified to an equivalent one where the extent of randomisation is reduced and does not exceed a universal upper bound. This implies that effectively, minimal randomisation suffices to target a desired payoff. As the number of states and number of signals become large virtually no randomisation is required and the automaton is "close" to a deterministic one. We require the following definition of the measure below :

Definition : Two finite state automata τ and τ' are said to be equivalent if they generate the same transition probabilities for all $\theta \in \{0,1\}$, have the same decision function and initial state.

Before proceed towards the result, we would need to define a measure of the amount of randomization taking place in the automaton τ. For each state $i \in M$, consider the number $n_i^\tau = |\{(x,j) : \sigma(i,x)(j) > 0\}|$. The measure of randomization is :

$$n(\tau) = \frac{\sum_i n_i^\tau}{|M||M||X|}$$

Notice that the measure of randomisation $n(\tau)$ is minimized if annd only if τ is a deterministic automaton. Additionally, the measure also strictly increases if there is marginally "more" randomisation. This means that if a state and signal combination in the transition functions adds more memory states in the support of the transition, then the measure strictly increases. The result deriving the upper bound is presented as follows :

Proposition 2 : Let $|M| \geq 3, |X| \geq 3$, then for any automaton τ, there exists an equivalent automaton τ' such that :

$$\frac{1}{|M|} \leq n(\tau') \leq \frac{1}{|M|} + \frac{2}{|X|}$$

Proof : Let τ be an automaton yielding transition probabilities α_{ij}, β_{ij}. For any equivalent automaton and $i \in M$, the transition function $\hat{\sigma}_i(x, j) = \sigma(i, x)(j)$ must satisfy the following conditions :

$$\begin{aligned} &\sum_x \hat{\sigma}_i(x,j) f_1(x) = \alpha_j \ \forall j \\ &\sum_x \hat{\sigma}_i(x,j) f_1(x) = \beta_j \ \forall j \\ &\sum_j \hat{\sigma}_i(x,j) = 1 \ \forall x \\ &\hat{\sigma}_i(x,j) \geq 0 \ \forall x, j. \end{aligned} \tag{3}$$

Viewing $\hat{\sigma}$ as an element in $\mathbb{R}^{X \times M}$, we observe that $\{\hat{\sigma} : \hat{\sigma}$ satisfies (7) $\}$ is a polytope of the form $\{x \in \mathbb{R}^{X \times M} : Ax \leq b\}$. It is true (see Schrijver(2013) [20]) that if z is an extreme point of the submatrix A_z defined by the set of inequalities which bind has $Rank(A_{\hat{\sigma}}) = |X| \times |M|$. In the above problem, there are $2|M| + |X| + |M| \times |X|$ inequalities. Consider an extreme point σ' of the polytope. For this point $Rank(A_{\hat{\sigma}}) = |X| \times |M|$. Since the first $2|M| + |X|$ constraints are all equalities, to satisfy the rank condition atleast $|M| \times |X| - 2|M| - |X| \geq 0$ of the last constraints should hold at equality. Let τ' be the automaton that replaces the transition function in τ with σ'. Then we have $n_i^{\tau'} \leq 2|M| + |X|$, hence :

$$n(\tau) = \frac{\sum_i n_i^{\tau}}{|M||M||X|} \geq \frac{|M|(2|M| + |X|)}{|M||M||X|} = \frac{1}{|M|} + \frac{2}{|X|}$$

The lower bound $\frac{1}{|M|}$ comes from the fact that the measure of randomization for deterministic automata is the lowest and equals $\frac{1}{|M|}$.

□

4 Weak admissibility and structure of optimal M-state automata

In this section, we define a new notion of interim rationality for an automaton to be obeyed in the context of the investment problem called *weak admissibilty.* The criterion has a flavor of an "admissibility" requirement as in statistical decision theory and classical hypothesis testing and is weaker than modified multiself consistency. When the automaton additionally does not "waste" memory states, the requirement uniquely identifies the nature of transitions.

Suppose the set of states of the automaton is $M = \{1, ..., m\}$. Automaton τ generates the following value function equations :

$$V_i^0 = d(i)E_0 + \delta \sum_{j \in M} \alpha_{ij} V_j^0$$
$$V_i^1 = d(i)E_1 + \delta \sum_{j \in M} \beta_{ij} V_j^1$$

Consider the transition from a state $i \in M$ that is reached with positive probability. Suppose under the automaton transitions dictate probabilities β_i in the state of the world $\theta = 1$. Fixing β_i, it would be natural to require that the transitions in the good state of the world are optimal i.e :

$$\begin{aligned} \max_{\alpha} \quad & \sum_j \alpha_j V_j^0 \\ \text{s.t.} \quad & \alpha \in K(\beta_i) \end{aligned} \tag{4}$$

Where for a probability distribution $\beta \in \Delta(M)$, $K(\beta)$ is defined as :

$$K(\beta) = \{\alpha \in \Delta(M) : \exists \sigma : X \to \Delta(M) \text{ s.t } \sum_x f_0(x)\sigma(x) = \alpha \text{ and } \sum_x f_1(x)\sigma(x) = \beta\}$$

The above condition has the flavor of optimal tests in classical hyopthesis testing. The transition proability vector β_i may be interpreted as the "size of the test" and we choose optimally from the constrained choice of α_i. Weak admissibility is hence defined as follows :

Definition : An automaton τ is said to be *weakly admissible* if at every state i reached with positve probabilitiy, the transition probabilities (α_i, β_i) satisfy :

$$\begin{aligned} \alpha_i \in \arg\min_{\alpha} \quad & \sum_j \alpha_j V_j^0 \\ \text{s.t.} \quad & \alpha \in K(\beta_i) \end{aligned}$$

We refine the class of automata of interest by requiring that no wastage of states occurs in the performance of the automaton. By this, it is meant that no two states reached with positive probabilitiy generate the same continuation values in both states of the world $\theta \in \{0,1\}$ and also that one does not "weakly dominate" in terms of continuation values. We present a formal definition below :

Definition : An automaton τ satisfies *no wastage* if for any two states i and j reached with positive probability :

1. $(V_i^0, V_i^1) \neq (V_j^0, V_j^1)$
2. $(V_i^0, V_i^1) \ngeq (V_j^0, V_j^1)$ and $(V_j^0, V_j^1) \ngeq (V_i^0, V_i^1)$

4.1 Structure of weakly admissible automata with no wastage

The optimization problem posed by the weak admissiblity constraint is a linear program and for every $\beta \in \Delta(M)$, $K(\beta)$ is a polytope in $\mathbb{R}^M$. It can be written more explicitly as follows :

$$\begin{aligned} \max_{\sigma} \quad & \sum_j \sum_x \sigma(x,j) f_0(x) V_j^0 \\ \text{s.t.} \quad & \sum_x \sigma(x,j) f_1(x) = \beta_j \ \forall j \\ & \sum_j \sigma(x,j) = 1 \ \forall x \\ & \sigma(x,j) \geq 0 \ \forall x, j. \end{aligned} \tag{5}$$

Here, the $\sigma(x,j)$ denotes the probability of transition to state j upon receiving signal x. By thinking of σ as a matrix of dimension $|X| \times |M|$ with non-negative entries, the first constraint says that the columns up to a vector of ones and the rows sum to the vector β. It can be shown the above problem is equivalent to the following problem :

$$\begin{aligned} \max_{u} \quad & \sum_j \sum_x u(x,j) \frac{f_0(x)}{f_1(x)} V_j^0 \\ \text{s.t.} \quad & \sum_x u(x,j) = \beta_j \ \forall j \\ & \sum_j u(x,j) = f_1(x) \ \forall x \\ & u(x,j) \geq 0 \ \forall x, j. \end{aligned} \tag{6}$$

The equivalence is guaranteed by noting the relation $u(x, j) = \sigma(x, j) f_1(x)$. Note that the feasible set is the set of all non-negative bi-stochastic matrices whose columns add to f_1 and rows add to β. Note also that since the above problem is considered from a state $i \in M$ reached with positive probability, then clearly any $j \in M$ with $\beta_j > 0$ implies that j is reached with positive probability as well. Hence, from no wastage, all the vectors in $\{(V_j^0, V_j^0) : j \text{ such that } \beta_j > 0\}$. A solution exists due to compactness of $K(\beta)$ and continuity of the objective and the following algorithm guarantees a unique optimal solution u^*.

Algorithm :

1. Define $u_1 = \mathbf{0}$
2. Suppose u_n is achieved at the nth step of algorithm. Search for (x, j) with the highest value of $\frac{f_0(x)}{f_1(x)} V_j^0$ such that $\min\{f_1(x), \beta_j\} - u_n(x, j) > 0$. If none exists, return u_n otherwise return the matrix u_{n+1} defined by :

$$u_{n+1}(x', j') = \begin{cases} \min\{f_1(x), \beta_j\} & \text{if } (x', j') = (x, j) \\ u_n(x', j') & \text{o.w} \\ \end{cases}$$

The above algorithm terminates and returns the unique optimal solution with the following "staircase" form where the order of row components i.e x's is done via likelihood ratios $\frac{f_0(x)}{f_1(x)}$ and the ordering over columun components i.e $j's$ is done via the continuation values V_j^0 (from no wastage this ordering is unique).

$$\sigma = \begin{pmatrix} * & * & 0 & 0 & \dots & 0 \\ 0 & * & 0 & 0 & \dots & 0 \\ 0 & * & * & * & \dots & 0 \\ 0 & 0 & 0 & * & \dots & 0 \\ \vdots & & & & \vdots & \ddots \\ 0 & 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & \dots & * \\ \end{pmatrix}. \tag{7}$$

This leads us to the following proposition :

Theorem 1 : Every automaton that is weakly admissible and satisfies no wastage necessarily admits a transition function of a staircase form as in (7). Moreover, is the unique solution to the LP problem imposed by the weak admissibility constraints.

Proof : The algorithm above terminates uniquely to say σ^* since from no wastage the values $\frac{f_0(x)}{f_1(x)}V_j^0$ are distinct over all (x, j)'s. Consider any optimal solution $\hat{\sigma}$. It is sufficient to show that either the first row or the first column has the intersecting term ((x, j) with highest value of $\frac{f_0(x)}{f_1(x)}V_j^0$)) as the only non zero elements. If not, then there exists a (y, k) such that $\frac{f_0(y)}{f_1(y)} < \frac{f_0(x)}{f_1(x)}$ and $V_k^0 < V_j^0$ such that $\hat{\sigma}(y, j) > 0$ and $\hat{\sigma}(x, k) > 0$. Now clearly, $\epsilon > 0$ amount of weight can be taken from $\hat{\sigma}(y, j) > 0$ and $\hat{\sigma}(x, k) > 0$ and shifted to $\hat{\sigma}(x, j) > 0$ and $\hat{\sigma}(y, k) > 0$ to strictly improve the solution. Hence, $\hat{\sigma}$ agrees with the first step of the algorithm. By induction, it can be shown that it agrees with all steps. Hence $\hat{\sigma} = \sigma^*$ □

Using the above result, we can now show that weakly admissible automata with no wastage improve the upper bound on the extent of randomization involved when using the measure of randomisation as introduced in section 3. We present it as the following result.

Proposition 4 : Let τ be a weakly admissible automaton with no-wastage, then :

$$\frac{1}{|M|} \leq n(\tau') \leq \frac{1}{|M|} + \frac{1}{|X|}$$

Proof : The algorithm terminates to an extreme point of the feasible set in the linear program. The argument is similar to the proof of proposition 2 but with tighter constraints. □

Now, for optimal automata, the following can be said :

Proposition 5 : For every optimal automaton τ, there exists an equivalent automaton with the same ex-ante value that is weakly admissible and satisfies no wastage.

Proof : Call two memory states reached with positive probability i and j *equivalent* if they generate the same continuation values $(V_i^0, V_i^1) = (V_j^0, V_j^1)$. Notice that due to modified multiself consistency, it is not possible that one of them weakly dominates the other. For any two equivalent memory states i and j one can shift all transitions entering i to memory state j with the same probabilities. This eliminates memory states, inducing the same continuation values and we are left with an automaton with no wastage. We now have an optimal automaton with no wastage. Since optimal automata are modified multiself consistent, at memory state i, we know that in the interim, the decision rule (transitions)

$\sigma_i : X \to \Delta(M)$ is ex-post optimal for the decision problem with prior $\bar{\pi}(i)$ and state dependent utilities $U(k,\theta) := V_k^0$. Hence, it must be ex-ante optimal as well. Since, σ_i generates β_i, it must hence yield higher payoff than all decision rules which generate $K(\beta_i)$. Hence, σ must solve the problem imposed by the weak admissibility constraint.
□

5 Changing worlds and optimal finite memory

In this section we study a more general decision problem where the underlying state of the word θ can potentially change over time evolving according to a Markov process P. Consider $\Theta = \{0, 1\}$ and suppose that the underlying state evolves according to :

$$P(\theta = 0|\theta = 0) = 1 - \epsilon$$
$$P(\theta = 1|\theta = 1) = 1 - \Delta$$

where $0 < \Delta, \epsilon < 1$ and the stationary distribution is $(\frac{\Delta}{\Delta+\epsilon}, \frac{\epsilon}{\Delta+\epsilon})$

5.1 Unconstrained optimal automata

The decision problem every period is to invest or not every period depending on the current state and maximise long run discounted expected expected utility. The utilities for actions are : $u(I, \theta) = E_\theta$ and $u(NI, \theta) = 0$ and $E_0 > 0 > E_1$. However, the states are not observable and the agent can only observe an informative signal from a finite set X about the current state via the signalling structures $< f_0(x), f_1(x) >$, where $f_0(x), f_1(x) > 0$. We assume that payoffs are unobservable every period hence ommiting the possibility of inference of the true state from the period payoff. When Δ, ϵ are small, the decision problem can be imagined as one close to the investment problem in section 1, where the probability of staying at a state is close to one.

Such models are also called hidden Markov models and fall under the category of non-interactive partially observable Markov decision processes. The memory unconstrained ex-ante optimal decision rule is simple : At time t if belief about the current being $\theta = 0$ is π_t, then invest if and only if $\pi_t \geq \pi^*$, where π^* is such that $\pi^* E_0 + (1 - \pi^*)E_1 = 0$.

Clearly the decision path is completely described by the belief process. Hence, to understand the optimal decision rule, it is important to understand the dynamics of the belief process. At a given time if the belief about $\theta = 0$ is π , then the belief about the next state after receving the signal x will be :

$$\hat{\pi}(\pi, x) = \frac{\pi(1-\epsilon)f_0(x) + (1-\pi)\Delta f_(x)}{\pi(1-\epsilon)f_0(x) + (1-\pi)\Delta f_0(x) + \pi\epsilon f_1(x) + (1-\pi)(1-\Delta)f_1(x)} \tag{8}$$

Before proceeding to the result the following definition would be important :

Definition : A signalling structure $< f_0(x), f_1(x) >$, has *full support* if for all $\theta \in \{0, 1\}$, for all $x \in X$, $f_\theta(x) > 0$.

Definition : A signalling structure $< f_0(x), f_1(x) >$ with full support is said to be *informative* if $f_0 \neq f_1$

We now present the result :

Proposition 6 : If $\frac{\Delta}{\Delta+\epsilon} \neq \pi^*$, and $1 - \Delta > \epsilon$, then there exists an informative signalling structure $< f_0(x), f_1(x) >$, with full support such that the optimal decision rule can be implemented by a deterministic finite state automaton.

Proof : We shall prove this for the case when $\frac{\Delta}{\Delta+\epsilon} < \pi^*$. The proof for $\frac{\Delta}{\Delta+\epsilon} > \pi^*$ will be analogous. It will be convenient to define represent updated beliefs in the following way :

$$\hat{\pi}(\pi, x) = \frac{\pi(1-\epsilon)f_0(x) + (1-\pi)\Delta f_(x)}{\pi(1-\epsilon)f_0(x) + (1-\pi)\Delta f_0(x) + \pi\epsilon f_1(x) + (1-\pi)(1-\Delta)f_1(x)} \tag{9}$$

$$= \frac{1}{1 + \frac{\pi\epsilon f_1(x) + (1-\pi)(1-\Delta)f_1(x)}{\pi(1-\epsilon)f_0(x) + (1-\pi)\Delta f_0(x)}} \tag{10}$$

$$= \frac{1}{1 + \frac{f_1(x)}{f_0(x)}\frac{\pi\epsilon + (1-\pi)(1-\Delta)}{\pi(1-\epsilon) + (1-\pi)\Delta}} \tag{11}$$

$$= \frac{1}{1 + \frac{f_1(x)}{f_0(x)}\frac{\pi\epsilon + (1-\pi)(1-\Delta)}{\pi(1-\epsilon) + (1-\pi)\Delta}(\frac{\pi}{1-\pi})(\frac{1-\pi}{\pi})} \tag{12}$$

The condition $1-\Delta > \epsilon$ guarantees that the function $\hat{\pi}(\pi, x)$ is strictly increasing in π. For $l > 0$ define the function $g^l(\pi)$ defined as :

$$g^l(\pi) = \frac{1}{1 + \frac{1}{l}\frac{\pi\epsilon + (1-\pi)(1-\Delta)}{\pi(1-\epsilon)+(1-\pi)\Delta}}$$

Notice that that g is continuous, strictly increasing and maps $[0,1]$ to $[0,1]$ and has a unique fixed point. The point $\bar{\pi}$ is a fixed point if and only if :

$$\frac{1}{l}\frac{\pi\epsilon + (1-\pi)(1-\Delta)}{\pi(1-\epsilon)+(1-\pi)\Delta} - \frac{\pi}{1-\pi} = 0$$

In addition, the fixed points are strictly increasing and continuous in l. Let $\bar{\pi}_l$ be the unique fixed point associated with $l > 0$. Notice that $\bar{\pi}_1 = \frac{\Delta}{\Delta+\epsilon}$. By continuity of π_l in l, there exists $l^* > 1$ such that

$$\tfrac{\Delta}{\Delta+\epsilon} < \bar{\pi}_{l^*} < \pi^*$$

Now, there exists an informative signalling structure $< f_0(x), f_1(x) >$ with full support such that for $\max\{\frac{f_0(x)}{f_1(x)} : x \in X\} < l^*$: We show that this signalling structure has the desired property. Order the x's according to likelihood ratios as $\{x_1, .., x_n\}$. We can assume for convenience that likelihood ratios are distinct hence strictly increasing in i. For each $i \in \{1, ..., n\}$, define $\bar{\pi}_i$ as the unique fixed point associated with $\frac{f_0(x_i)}{f_1(x_i)}$. Hence, we have :

$$0 < \bar{\pi}_1 < \bar{\pi}_2 ... \leq \tfrac{\Delta}{\Delta+\epsilon} \leq \bar{\pi}_{n-1} < \bar{\pi}_n < \pi^* < 1$$

Now, consider the partition $[0, \bar{\pi}_1) \cup [\bar{\pi}_1, \bar{\pi}_n] \cup (\bar{\pi}_n, 1]$ of $[0,1]$. We can show that if the belief enters the interval $[\bar{\pi}_1, \bar{\pi}_n]$, there it will forever stay there with probabilitiy one. If the prior π belongs to the interval $(\bar{\pi}_n, 1]$, then if the signal with the highest likelihood ratio is received repeatedly, the belief drops below π^* for sure. Starting in $[0, \bar{\pi}_1)$, we forever stay below π^* irrespective of signals. Notice in all three cases, there is a time $\bar{t}$ after which the decision maker should not invest for sure. Finite automata can replicate all such decision rules

□

Corollary : For any informative signalling structure $< f_0(x), f_1(x) >$ with full support, let $\bar{l} = \max_x \frac{f_0(x)}{f_1(x)}$. If $\bar{\pi}_{\bar{l}} < \pi^*$, and $1-\epsilon > \Delta$, then the optimal decision rule can be implemented by a deterministic finite state automaton.

We now show the following :

Proposition 7 : If signals have full support and are informative and $1 - \epsilon > \delta$, then the belief process is Markov and has a unique invariant distribution.

Proof : The state space for the process is $[0,1]$ and transitions are defined by $\hat{\pi}(\pi, x)$ making the belief process Markov. By definition, it is clear that monotonicity and the feller property are satisfied. The third condition to check is the mixing condition (See SLP [22]). Since signals are informative, the fixed points associated with the likelihood ratios can be ordered according to the likelihood ratios as in the proof of the above proposition. Let Q be the transition function associated with the Markov process then for $c = \frac{\Delta}{\Delta+\epsilon}$, $Q^n(0, [c, 1]) > 0$ is positive by repeatedly receiving the highest signal and $Q^n(1, [0, c]) > 0$ by repeatedly receiving the lowest signal. Hence, the mixing condition is satisfied for c as defined and $\delta = \min\{Q^n(0, [c, 1]), Q^n(1, [0, c])\}$. Hence there exists a unique invariant distribution. $\square$

Example : *Variant of Shiryaev's Disruption Problem* In the Shiryaev decision problem the transitions take place according to

	0	1
0	$1-\epsilon$	ϵ
1	0	1

The results presented above depends on the threshold belief π^*. In the Shiryaev problem, one can obtain a stronger result which does not depend on the threshold. The statement is as follows :

Proposition 8 : Let $\bar{l} = \max \frac{f_0(x)}{f_1(x)}$. If $\bar{l}(1-\epsilon) < 1$, then belief process converges to 0 almost surely. Hence, deterministic finite state automata can execute the optimal decision rule.

Proof : For the transition matrix above :

$$g^l(\pi) = \frac{\pi l(1-\epsilon)}{\pi(l-1)(1-\epsilon)+1}$$

Notice that g^l is strictly increasing, continuous and differentiable over $[0,1]$ with $g^l(0) = 0$ for all $l > 0$. The derivative is :

$$g^{l\prime}(\pi) = \frac{l(1-\epsilon)}{(\pi(l-1)(1-\epsilon)+1)^2}$$

Since $\pi(l-1)(1-\epsilon)+1 = l\pi(1-\epsilon)+(1-\pi(1-\epsilon)) > 0$ we have that $g^{l\prime}$ is strictly decreasing in π for $l > 1$ and strictly increasing in π for $l < 1$. Hence, g^l is concave for $l > 1$ and convex for $l < 1$.

$$g^{l\prime}(0) = l(1-p)$$

Now since, $\bar{l}(1-p) < 1$, for each $x \in X$, $g^{\frac{f_0(x)}{f_1(x)}}(\pi) < \pi$ for all $\pi > 0$. Hence, beliefs always strictly decrease even under the effect of the highest signal and as a result converge to 0. $\square$

The propositions 6 and 8 mainly serve to demonstrate that under certain parametric conditions, optimality is achieved with finite memory. Hence, in these environments, agents with a large memory capacity might choose to underutilise their capacity without any loss in payoff and in turn achieve the unconstrained optimum. These results are in line with Monte and Said (2014) who exhibit examples of "coarseness" in the decision making process. We extend their results in allowing for more general signalling structures and additionally shed light on the belief formation process which has a Markovian and "categorical" nature. The result in Proposition 7 takes us a step further and shows using the Markovian nature of belief that asymptotically, they converge to a stationary distribution. This stationary distribution may be used to obtain the limit of means payoff since the myopic belief based threshold rule would apply there as well.

5.2 Constrained optimal automata

In the section above, we focussed on conditions under which the optimal value under full memory could be achieved by a finite state automaton i.e actions taken by decision maker not bounded by memory constraints may be replicated by one who is not. Here, we consider the constrained problem, where the agent chooses the optimal automata with a fixed number of states. The definition of an automaton $\tau =< i_0, M, \sigma, d >$ and optimality are as in section 1. Note here that the payoff sequence is completely determined by the evolution of the pair (θ, i) which gives payoff $E_\theta d(i)$. These transitions follows a Markov chain Q on $\Theta \times M$ and take place in the following manner :

$$Q(\theta', i'|\theta, i) = P(\theta'|\theta) \sum_{x \in X} f_{\theta'}(x)\sigma(i,x)(i') \tag{13}$$

defining the transition probabilities $\alpha^\theta_{ij} := \sum_{x \in X} f_\theta(x)\sigma(i,x)(j)$, the above simplifies to :

$$Q(\theta', i'|\theta, i) = P(\theta'^{\theta'}_{ii'} \tag{14}$$

Let $V(\theta, i)$ denote the value generated by Q starting from (θ, i). Interpreting the discount factor as a stopping probabilitiy, from Kocer(2010) [14] it is known that optimal automata in POMDP environments are modified multiself consistent. In the current environment, it is satisfied for beliefs generated by the unique invariant distribution $\hat{\mu}$ on $\Theta \times M$ of the perturbed Markov chain :

$$\hat{Q}(\theta', i'|\theta, i) = (1-\delta)P(\theta'|\theta)\pi(\theta')\delta_{i_0}(i') + \delta P(\theta'^{\theta'}_{ii'} \tag{15}$$

Modified multiself consistency means the following :

Definition : An automaton τ is said to be *modified multiself consistent* if for the invariant distribution $\hat{\mu}$, the following is satisfied :

1. *Decisions* : For all $i \in M$,

$$d(i) \in \arg\max_{a \in \{I, NI\}} [\hat{\mu}(0|i)E_0 + \hat{\mu}(1|i)E_1]\mathbb{I}_{\{a=I\}} \tag{16}$$

2. *Transitions* : For $(i, x) \in M \times X$, if $\sigma(i, x)(j) > 0$, then :

$$j \in \arg\max_{k \in M}[\hat{\pi}(\hat{\mu}(0|i), x)V(0, k) + (1 - \hat{\pi}(\hat{\mu}(0|i), x))V(0, k)] \tag{17}$$

We may modify the optimal automaton in a way such that the values generated by the states with positive probability in the invariant distribution are all different. If we interpret each j as a decision taken under uncertainty yielding payoff $V(\theta, j)$ in state $\theta \in \Theta$, MMC tells us that states j with $\hat{\mu}(\theta, j) > 0$ are admissible in the set :

$$V = \{(V(0, k), V(1, k)) : k \in M\}$$

Moreover it can be shown (see appendix) that a *revelation principle* (As in Wilson(2014)) holds here as well i.e at a belief $\hat{\mu}(0|i) > 0$, the decision $(V(0, i), V(1, i))$ is optimal.

Proposition 9 (*Revelation Principle*) : Let $i \in M$ such that $\hat{\mu}(0|i) > 0$, then :

$$i \in \arg\max_{k \in M} \hat{\mu}(0|i)V(0, k) + \hat{\mu}(1|i)V(1, k) \tag{18}$$

Proof : Follows from a more general result in the appendix
□

Hence, if an individual reaches a state that recurs, his optimal choice is indeed to stay there. Using arguments from Wilson(2014) the following useful result can be shown :

Proposition 10 (*Categorisation*) : There exist thesholds $0 = \underline{\pi}_1 < .. \leq \underline{\pi}_i \leq \underline{\pi}_{i+1} \leq ... < \underline{\pi}_{m+1} = 1$ such that $\sigma(i,x)(j) > 0$ only if $\hat{\pi}(\hat{\mu}(0|i), x) \in [\underline{\pi}_j, \underline{\pi}_{j+1}]$

Proof : Since the values are $(V(0,k), V(1,k))$ are all distinct, we can order the states in M according to the value $V(0,k)$. Notice that the revelation principle implies that :

$$V(0,k) > V(0,l) \implies \hat{\mu}(0|k) \geq \hat{\mu}(0|l)$$

because if $\hat{\mu}(0|k) < \hat{\mu}(0|l)$, then the agent strictly prefers l over k at state k. Hence, the ordering is consistent with the beliefs at the states of the automaton. Now, define the thresholds as :

$$\underline{\pi}_i := \min\{\pi' : i \in \arg\max_k \pi' V(0,k) + (1-\pi')V(1,k)\}$$
$$\underline{\pi}_{m+1} = 1$$

We shall show that the values defined above are indeed the desired thresholds. First we make the following observations :

1. For state $1 \in M$, $\underline{\pi}_1 = 0$. This is true since $V(1,1) > V(1,k)$ for all $k \in M \backslash \{1\}$.
2. For each $i \in M$, $\underline{\pi}_{i+1} = \max\{\pi' : i \in \arg\max_k \pi' V(0,k) + (1-\pi')V(1,k)\} =: c$. Suppose not. By definition of $\underline{\pi}_{i+1}, c$, $\underline{\pi}_{i+1}V(0,i+1) + (1-\underline{\pi}_{i+1})V(1,i+1) = \underline{\pi}_{i+1}V(0,i) + (1-\underline{\pi}_{i+1})V(1,i)$ and $cV(0,i+1) + (1-c)V(1,i+1) = cV(0,i) + (1-c)V(1,i)$. Hence $c = \underline{\pi}_{i+1}$
3. Notice that $\underline{\pi}_i \leq \underline{\pi}_{i+1}$. This is true since $V(0,i+1) > V(0,i)$.

Now we show the result. Suppose $\sigma(i,x)(j) > 0$. Then by modified multiself consistency, it must be the case that $\hat{\pi}(\hat{\mu}(0|i), x) \in \{\pi' : j \in \arg\max_k \pi' V(0,k) + (1-\pi')V(1,k)\}$. From 2 above, it follows that $\hat{\pi}(\hat{\mu}(0|i), x) \in [\underline{\pi}_j, \underline{\pi}_{j+1}]$
□

The above result shows that beliefs before and after receiving signals always lie in some category of beliefs. This observation is very useful to understand the nature of the transitions taking place in an optimal automaton (as we shall see below). First, we define some notation. As in the previous section the fixed points induced by the signals $X = \{x_1, ..., x_n\}$

will be denote $\bar{\pi}_j$ or alternatively as $\bar{\pi}_x$ (the fixed point associated with the signal $x \in X$). For each $x \in X$ define the set $M_x^- = \{k \in M : \underline{\pi}_k < \bar{\pi}_x\}$, and the sets $M_x^+ = \{k \in M : \underline{\pi}_k > \bar{\pi}_x\}$.

Proposition 11 : (*Optimal Automata*) For the optimal automata τ^* the following are true :

1. (*Bounded Jumps*) For each signal $x \in X$:
   (a) If $i \in M_x^-$, then
   If $\sigma(i,x)(j) > 0$, then $j \notin \{k \in M : k < i\} \cup M_x^+$
   (b) If $i \in M_x^+$, then
   If $\sigma(i,x)(j) > 0$, then $j \notin \{k \in M : k > i\} \cup M_x^-$
2. (*Bias*) For $x \in X$, if $\bar{\pi}_x < \pi$, then for all $i \geq i_0$
   If $\sigma(i,x)(j) > 0$ then $j \leq i$
   Hence, it is possible to go to a lower state even when the likelihood ratio of the signal is strictly greater than 1.

Proof :

1. We only prove (a). The proof for (b) is analogous. Let $i \in M_x^-$, then $\hat{\pi}(\hat{\mu}(0|i), x) = g^{l(x)}(\hat{\mu}(0|i))$. We show the claim using properties of the function $g^l(.)$ discussed in the previous section. There are two cases to consider :
   (a) *CASE 1* : Suppose that $\hat{\mu}(0|i) \leq \bar{\pi}_x$. If equality is satisfied, $\hat{\mu}(0|i) = \hat{\pi}(\hat{\mu}(0|i), x) = \bar{\pi}_x$ and we are done. Suppose $\hat{\mu}(0|i) < \bar{\pi}_x$ From the properties of the function $g^{l(x)}(.)$, we know that $\bar{\pi}_x$ is the unique fixed point of $g^{l(x)}$ and that $\hat{\mu}(0|i) < \hat{\pi}(\hat{\mu}(0|i), x) = g^{l(x)}(\hat{\mu}(0|i)) < \bar{\pi}_x$. This implies from Proposition 10 that $\underline{\pi}_i < \hat{\pi}(\hat{\mu}(0|i), x)$. Hence, there cannot be transitions to $\{k \in M : k < i\}$. Since $\hat{\pi}(\hat{\mu}(0|i), x) < \bar{\pi}_x$, we have $\hat{\pi}(\hat{\mu}(0|i), x) < \underline{\pi}_k$ for all $k \in M_x^+$. Hence, transitions to the set M_x^+ are not possible.
   (b) *CASE 2* : Now suppose that $\bar{\pi}_x < \hat{\mu}(0|i)$. Again, from properties of the function $g^{l(x)}$, it is the case that $\bar{\pi}_x < \hat{\pi}(\hat{\mu}(0|i), x) < \hat{\mu}(0|i) \leq \underline{\pi}_k$ for all $k \in M_x^+$. Hence,
2. Note that it is sufficient to show the claim for the initial state. We consider two cases :

(a) *CASE 1* : Suppose $\hat{\mu}(0|i_0) \leq \bar{\pi}_x$. By optimality, we know that $\pi \in [\underline{\pi}_{i_0}, \underline{\pi}_{i_0+1}]$. Hence, for all $j > i_0$ it is the case that $\hat{\pi}(\hat{\mu}(0|i_0), x)\bar{\pi}_x < \pi \leq \underline{\pi}_j$. Hence transitions to $j > i_0$ cannot happen.

(b) *CASE 2* Suppose $\hat{\mu}(0|i_0) > \bar{\pi}_x$, then the result follows from arguments in 1.

□

A consequence of the changing worlds environment is that a decision maker may receive unfavourable signals for a long time and still not go below a threshold of beliefs. Since the state changes are also taking places randomly every period, the effect of the Markov chain enters the beliefs as well. Hence, if signals are not strong enough, they may impact beliefs for a long time. Only extreme signals (if there exist any) can change beliefs significantly.

6 Conclusions

In this paper, we have investigated some properties of inference with bounded memory. We have assumed that the economic agent uses a finite automaton to process external signals about the true state of the world and to make decisions (with payoffs dependent on the true state). There has been recent interest in this problem, with the earlier papers cited previously. We consider a different decision problem than these others, with our aim being to explain the phenomenon that agents will often effectively ignore "small" pieces of bad news and continue to take the same actions until some extreme bad news leads to a change. This is arguably (see [8]) one reason why investors kept on repeating their choices even when evidence was accumulating that something was wrong. This phenomenon is difficult to explain under fully Bayesian decision-makers; this is true even with bounded decision-making, if the automaton used a large number of states and the *state of the world* does not change. Suppose, for example, that there are three signals and fifty states with the first two signals being "bad news" in terms of the likelihood ratios being less than one. If the agent starts off in a state near the middle, a large number of instances of the middle signal will drive the process into the next lower states of the automaton. With changing worlds, however, which might reflect the reality of the fundamentals at the time in question, the optimal automaton is characterised by a sequence of fixed points, one for each signal. Even if the middle signal is repeated, the belief might not go into the next lower category because there is a countervailing belief that the state of the world might have improved. Our characterisation of the optimal automaton, therefore demonstrates that the behaviour of agents during the financial crisis was not necessarily sub-optimal; it was optimal to stay in the same qualitative category until a piece of really bad news forced the decision-maker into a lower category of beliefs (about the good state of the world).

Our paper also characterises the structure of automata (with a fixed state of the world) that satisfy a notion of optimality familiar in statistical decision theory, namely *weak admissibility.* The structure derived is intuitively appealing; a lower signal leads to a lower state, thus exhibiting some kind of monotonicity in the transitions. The result is derived using dominance arguments and does not involve long-run beliefs over the state the automaton is in.

A third main result we have shown involves a measure and bound on randomisation arising from the structure mentioned in the previous paragraph. Since the transition matrix among states in the automaton has the specified structure the extent of randomisation is limited. Also, consistent with intuition, the larger the number of states or the number of signals the lower the measure of randomisation. This result completes the analysis of randomisation in optimal automata begun in [10] and [13].

References

[1] Abreu, Dilip, and Ariel Rubinstein. "The structure of Nash equilibrium in repeated games with finite automata." Econometrica: Journal of the Econometric Society (1988): 1259-1281.

[2] Al-Najjar, Nabil I., and Mallesh M. Pai. "Coarse decision making and overfitting." Journal of Economic Theory 150 (2014): 467-486.

[3] Aumann, Robert J., Hart, Sergiu and Perry, Motty, 1997. "The Absent-Minded Driver," *Games and Economic Behavior*, Elsevier, vol. 20(1), pages 102-116, July.

[4] Chatterjee, Kalyan, and Hamid Sabourian. "Multiperson bargaining and strategic complexity." Econometrica (2000): 1491-1509.

[5] Daskalova, Vessela, and Nicolaas J. Vriend. "Categorization and Coordination", Cambridge INET, No. 719. 2014.

[6] Dow, J. "Search Decisions with Limited Memory" *Review of Economic Studies* (1991) 58 (1): 1-14.

[7] Ferguson, Thomas S. "Mathematical Statistics." (1967).

[8] Gennaioli, Nicola, Andrei Shleifer, and Robert W Vishny. 2013. “A Model of Shadow Banking.” Journal of Finance 68 (4): 1331-1363.

[9] M. E. Hellman and T. M. Cover, "Learning with Finite Memory,", Vol. 41, June 1970, pp. 765-782. *Ann. of Math. Stat.* 1991.

[10] Hellman, Martin E., and Thomas M. Cover. "On memory saved by randomization." The Annals of Mathematical Statistics (1971): 1075-1078.

[11] Fryer, Roland, and Matthew O. Jackson. "A categorical model of cognition and biased decision making." The BE Journal of Theoretical Economics 8.1 (2008).

[12] Grove, Adam J., and Joseph Y. Halpern. "On the expected value of games with absent-mindedness." Games and Economic Behavior 20.1 (1997): 51-65.

[13] Kalai, Ehud, and Eilon Solan. "Randomization and simplification in dynamic decision-making." Journal of Economic Theory 111.2 (2003): 251-264.

[14] Kocer, Yilmaz, "Optimal Plans with Bounded Memory are Modified Multiself Consistent", working paper 2014

[15] Monte, Daniel "Learning with bounded memory in games." Ph.D thesis, Yale University (2007).

[16] Monte, Daniel, and Maher Said. "The value of (bounded) memory in a changing world." Economic Theory 56.1 (2014): 59-82.

[17] Peski, Marcin. "Prior symmetry, similarity-based reasoning, and endogenous categorization." Journal of Economic Theory 146.1 (2011): 111-140.

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

[19] Rubinstein, Ariel. "Finite automata play the repeated prisoner's dilemma." Journal of economic theory 39.1 (1986): 83-96.

[20] Schrijver, A., "A Course in Combinatorial Optimization" 2013

[21] Shiryaev, Albert N., "Optimal stopping rules" *Springer* 1978

[22] Stokey, Nancy L. Recursive methods in economic dynamics. Harvard University Press, 1989.

[23] Wilson, Andrea, "Bounded Memory and Biases in Information Processing", Econometrica, Vol 82 No. 6 , pp 2257-2294, 2014

7 Appendix

7.1 A revelation principle for POMDP under bounded memory

This section establishes a revelation principle in the sense of Wilson(2014) for optimal plans uncer bounded memory in environments where the decision maker faces a partially observable Markov decision process (POMDP). POMDP environments are general decision problems with state unobservability and encompass the investment problems with unchanging and changing worlds discussed in the paper. We develop the framework as in Kocer (2010) and then establish a revelation principle using modified mutliself consistency.

There is a finite set of states of the world Θ. The agent takes an actions every period from a finite set A and receives per period payoffs according to utility function $u : A \times \Theta \to \mathbb{R}$. Hence, if in a given period action $a \in A$ is chosen and the current state of the world is $\theta \in \Theta$, the agent receives payoff $u(a, \theta)$ in that period. The state of the world is however unobservable to the agent, evolves randomly and can inferred from a signals received from finite set X. The state evolution and signal generation process is determined jointly according to a transition function $P : \Theta \times A \to \Delta(\Theta \times X)$. The interpretation is as follows : If the current state is $\theta \in \Theta$ and current action taken is $a \in A$, then the next state $\theta' \in \Theta$ and signal $x \in X$ are both jointly determined according to probability $P(\theta', x|\theta, a) \equiv P(\theta, a)(\theta', x)$. The objective of the agent is to maximise long run discounted expected utility according to discount factor $\delta \in (0, 1)$ i.e maximise :

$$\mathbb{E}[\sum_{t=0}^{\infty}(1-\delta)\delta^t u(a_t, \theta_t)] \tag{19}$$

The agent uses an automaton $\tau =< g_0, M, \sigma, d >$ to maximise (24). Here, $g_0 \in \Delta(M)$, $\sigma : M \times X \to \Delta(M)$ and $d : M \to \Delta(A)$. Note that once an automaton is fixed the long run payoffs are completely determined by the evolution of the pair (θ, i) over time. This happens according to a Markov process :

$$Q(\theta', i'|\theta, i) = \sum_{a \in A} d(i)(a) \sum_{x \in X} P(\theta', x|\theta, a)\sigma(i, x)(i') \tag{20}$$

Interpreting the discount factor as a continuation probability that explicitly enters the extensive form of the decision making environment, we obtain a long run distribution $\hat{\mu}$ over pairs in $\Theta \times M$ determined by :

$$\hat{\mu}(\theta, i) = \sum_{t=0}^{\infty}(1-\delta)\delta^t \mathbb{P}^t(\theta, i) \tag{21}$$

where $\mathbb{P}^t(\theta, i)$ denotes the probability of the pair (θ, i) occuring at time t under the transitions induced by Q above in (25). It can be shown that $\hat{\mu}$ is the unique invariant distribution on $\Theta \times M$ of the perturbed Markov chain :

$$\hat{Q}(\theta', i'^0(\theta'^0(i') + \delta Q(\theta', i'|\theta, i) \tag{22}$$

Denote the continuation values defined by each pair (θ, i) under an automaton be denoted as $V(\theta, i)$. Kocer(2010) has shown modified multiself consistency of optimal automata under beliefs generated by $\hat{\mu}$. A formal definition is given below :

Definition : An automaton τ is said to be *modified multiself consistent* if for the invariant distribution $\hat{\mu}$, the following conditions are satisfied :

1. *Decisions* : For all $i \in M$ with $\hat{\mu}(i) > 0$, $d(i)(a) > 0$ implies :

$$a \in \arg\max_{a' \in A} \sum_{\theta \in \Theta} \hat{\mu}(\theta|i)[u(a', \theta) + \delta \sum_{(\theta', x) \in \Theta \times X} \sum_{i' \in M} P(\theta', x|\theta, i)V(\theta', i')] \tag{23}$$

2. *Transitions* : For $(i, x) \in M \times X$, if $d(i)(a) > 0$ and $\sigma(i, x)(j) > 0$, then :

$$j \in \arg\max_{k \in M} \sum_{\theta' \in \Theta} \hat{\pi}(\hat{\mu}(.|i), a, x)(\theta')V(\theta', k) \tag{24}$$

where $\hat{\pi}(\hat{\mu}(.|i), a, x) \in \Delta(\Theta)$ is the belief about the next state when action $a \in A$ is taken and signal $x \in X$ is received :

$$\begin{aligned}
\hat{\pi}(\hat{\mu}(.|i), a, x)(\theta') &= \frac{\sum_{\theta \in \Theta} \hat{\mu}(\theta|i)d(i)(a)P(\theta', x|\theta, a)}{\sum_{\theta'' \in \Theta}\sum_{\theta \in \Theta} \hat{\mu}(\theta|i)d(i)(a)P(\theta'', x|\theta, a)} \\
&= \frac{\sum_{\theta \in \Theta} \hat{\mu}(\theta|i)P(\theta', x|\theta, a)}{\sum_{\theta'' \in \Theta}\sum_{\theta \in \Theta} \hat{\mu}(\theta|i)P(\theta'', x|\theta, a)}
\end{aligned} \tag{25}$$

We now state and prove the following revelation principle :

Proposition 12 : For each $i \in M$ such that $\hat{\mu}(i) > 0$:

$$i \in \arg\max_{k \in M} \sum_{\theta \in \Theta} \hat{\mu}(\theta|i) V(\theta, k) \tag{26}$$

Proof : Suppose not. Then there exists a $j \in M \backslash \{i\}$ such that :

$$\sum_{\theta \in \Theta} \hat{\mu}(\theta|i) V(\theta, j) > \sum_{\theta \in \Theta} \hat{\mu}(\theta|i) V(\theta, i) \tag{27}$$

Defining the set $\Pi(j,i) = \{\pi \in \Delta(\Theta) : \sum_{\theta \in \Theta} \pi(\theta) V(\theta, j) > \sum_{\theta \in \Theta} \pi(\theta) V(\theta, i)\}$ we observe that $\hat{\mu}(.|i) \in \Pi(j,i)$. Notice that $\Pi(j,i), \Pi(j,i)^c$ are both convex.

Now define the set $Z = (\Theta \times M \times \{s\} \times \Theta \times M) \cup (\Theta \times M \times \{n\} \times A \times \Theta \times X \times M)$. Notice that Z can be interpreted as a tree which branches out into distinct paths at history (θ, k) either through s or n. We define a probability measure ν on Z :

$\nu(\theta, l, s, \theta'', j) := \hat{\mu}(\theta, l)(1-\delta)\pi(\theta''^0(j)$ for all $(\theta, k, s, \theta'', j) \in \Theta \times M \times \{s\} \times \Theta \times M$

$\nu(\theta, l, n, \theta'', y, j) := \hat{\mu}(\theta, l)\delta d(l)(a) P(\theta'', x|\theta, a)\sigma(l, y)(j)$ for all $(\theta, l, n, a, \theta'', y, j) \in \Theta \times M \times \{n\} \times A \times \Theta \times X \times M$

For $z \in Z$, denote as $z(\theta), z(l)$,$z(c)$ (where $c \in \{s, n\}$) $z(\theta''), z(y), z(j)$ as the value of the appropriate nodes on the path z. We now establish the result via the following steps :

(a) Define the sets $Z_i = \{z \in Z : z(j) = i\}$,$Z_{\theta'} = \{z : z(\theta'') = \theta'\}$ then :

$$\begin{aligned}
\nu(Z_{\theta'}|Z_i) &= \frac{\nu(Z_{\theta'} \cap Z_i)}{\nu(Z_i)} \\
&= \frac{\sum_{\theta,l} \hat{\mu}(\theta,l)[(1-\delta)\pi(\theta'^0(i) + \delta \sum_{a \in A} d(i)(a) \sum_{x \in X} P(\theta',x|\theta,a)\sigma(i,x)(i')]}{\sum_{\theta,l} \hat{\mu}(\theta,l)[(1-\delta)\sum_{\theta''}\pi(\theta''^0(i) + \delta \sum_{\theta''}\sum_{a \in A} d(i)(a) \sum_{x \in X} P(\theta',x|\theta,a)\sigma(i,x)(i')]} \\
&= \frac{\sum_{\theta,l} \hat{\mu}(\theta,l)\hat{Q}(\theta',i|\theta,l)}{\sum_{\theta''}\sum_{\theta,l} \hat{\mu}(\theta,l)\hat{Q}(\theta'',i|\theta,l)} \\
&= \frac{\hat{\mu}(\theta',i)}{\sum_{\theta''}\hat{\mu}(\theta'',i)} \\
&= \hat{\mu}(\theta'|i)
\end{aligned}$$

Hence we have $\{\nu(Z_{\theta'}|Z_i)\}_{\theta' \in \Theta} \in \Pi(j,i)$.

(b) Define the sets $Z_s = \{z \in Z : z(c) = s\}$, $Z_n = \{z \in Z : z(c) = n\}$, $Z_n = \{z \in Z : z(a') = a\}$,$Z_k = \{z \in Z : z(l) = k\}$,$Z_x = \{z \in Z : z(y) = x\}$, . Clearly $Z_s \cap Z_n = \emptyset$. Now consider the set $L = \{(k,a,x) : \nu(Z_i \cap Z_n \cap Z_k \cap Z_a \cap Z_x) > 0\}$. Note that if $(k,a,x) \in L$ then $d(k)(a) > 0$ and $\sigma(k,x)(i) > 0$. For all $(k,a,x) \in L$ we have :

$$\begin{aligned}
\nu(Z_{\theta'}|Z_i \cap Z_n \cap Z_k \cap Z_a \cap Z_x) &= \frac{\nu(Z_{\theta'} \cap Z_i \cap Z_n \cap Z_k \cap Z_a \cap Z_x)}{Z_i \cap Z_n \cap Z_k \cap Z_a \cap Z_x} \\
&= \frac{\sum_{\theta} \hat{\mu}(\theta,k)\delta d(k)(a)P(\theta',x|\theta,a)\sigma(k,x)(i)}{\sum_{\theta''}\sum_{\theta} \hat{\mu}(\theta,k)\delta d(k)(a)P(\theta'',x|\theta,a)\sigma(k,x)(i)} \\
&= \frac{\sum_{\theta} \hat{\mu}(\theta,k)P(\theta',x|\theta,a)}{\sum_{\theta''}\sum_{\theta} \hat{\mu}(\theta,k)P(\theta'',x|\theta,a)} \\
&= \hat{\pi}(\hat{\mu}(.|k),a,x)(\theta')
\end{aligned}$$

By modified multiself consistency, we get that transitioning to i must be optimal at any $(k,a,x) \in L$ and hence must be better than j. This implies $\{\nu(Z_{\theta'}|Z_i \cap$

$Z_n \cap Z_k \cap Z_a \cap Z_x)\}_{\theta' \in \Theta} = \hat{\pi}(\hat{\mu}(.|k), a, x) \in \Pi(j,i)^c$. From the convexity of $\Pi(j,i)^c$ it follows that :

$< \nu(Z_{\theta'}|Z_i \cap Z_n) >_{\theta'} = < \sum_{(k,a,x) \in L} \nu(Z_{\theta'}|Z_i \cap Z_n \cap Z_k \cap Z_a \cap Z_x) \nu(Z_k \cap Z_a \cap Z_x | Z_i \cap Z_n) >_{\theta' \in \Theta}$

belongs to the set $\Pi(j,i)^c$.

(c) We now prove the result. There are two cases to consider :

i. *Case 1* : $g^0(i) = 0$. In this case, $\nu(Z_s|Z_i) = 0$. So, $\nu(Z_n|Z_i) = 1$ hence : $< \hat{\mu}(\theta'|i) >_{\theta' \in \Theta} = < \nu(Z_{\theta'}|Z_i) >_{\theta' \in \Theta} = < \nu(Z_{\theta'}|Z_i \cap Z_n) >_{\theta' \in \Theta} \in \Pi(j,i)^c$ which contradicts (10).

ii. *Case 2* : $g^0(i) > 0$. By optimality it must be the case that $\pi^0 \in \Pi(j,i)^c$. Moreover, $\pi^0 = < \nu(Z_{\theta'}|Z_i \cap Z_s) >_{\theta' \in \Theta}$. Also in this case $\nu(Z_s|Z_i) > 0$. Now from the convexity of $\Pi(j,i)^c$, we have :

$$\begin{aligned}
< \hat{\mu}(\theta'|i) >_{\theta' \in \Theta} &= < \nu(Z_{\theta'}|Z_i) >_{\theta' \in \Theta} \\
&= < \nu(Z_{\theta'}|Z_i \cap Z_s) \nu(Z_s|Z_i) + \nu(Z_{\theta'}|Z_i \cap Z_n) \nu(Z_n|Z_i) >_{\theta' \in \Theta} \\
&= < \pi(\theta') \nu(Z_s|Z_i) + \nu(Z_{\theta'}|Z_i \cap Z_n) \nu(Z_n|Z_i) >_{\theta' \in \Theta} \\
&\in \Pi(j,i)^c
\end{aligned}$$

which again contradicts (10).

□