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

Pathikrit Basu† Kalyan Chatterjee‡

February 25, 2018

Abstract

We study the process of decision-making and inference by a single, boundedly rational, economic agent. The agent chooses either a safe or a risky alternative in each period after receiving a signal about the state of the world in that period. The state of the world is changing according to a Markov process with some degree of persistence across time. The agent's decision rule is expressed as a finite-state automaton with a fixed number of memory states. Updating on the basis of the received signal is, for such an agent, making a transition from one state to another. The finiteness of the number of automaton states automatically suggests that beliefs are classified into categories and a signal causes a (possible) change in the category on the basis of which the next action is taken. The problem is one in partially-observable Markov decision processes (POMDP). We characterise the structure of the optimal decision rule in this setting and show how its properties pin down the categories of beliefs and explain some observed, seemingly irrational behaviour. We then address the issue of randomisation which could be interpreted as an additional measure of complexity.

*We are grateful to Yeon-Koo Che, Federico Echenique, Daniel Monte, Stephen Morris, Ariel Rubinstein, Muhamet Yildiz and conference participants at the NSF-CEME Decentralisation Conference (2016), Pennsylvania Economic Theory Conference (2016), Cornell-PSU Macro Conference (Spring 2016) and the North American Summer Meetings of the Econometric Society (2016) for comments and discussions. We also wish to thank Andrea Wilson and Yilmaz Kocer for reading an earlier draft. 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.

†California Institute of Technology. E-mail - pathkrtb@caltech.edu

‡Pennsylvania State University. E-mail - kchatterjee@psu.edu

We show that no randomisation is needed when the signal space is large, thus obtaining a purification result. Finally, we specialise to the case of a fixed state, and obtain further results on the structure of transitions in the optimal automaton.

1 Introduction

The aim of this paper is to study and characterise boundedly rational inference about an uncertain state of the world from a sequence of signals that are informative about the state. The bound on rationality in this paper is a constraint on available memory. Given this constraint, the single agent we consider is rational, in the sense of maximising lifetime discounted expected utility. The natural mathematical object to represent the bound on memory is a finite automaton; this is what we use and the nature of this object will be specified in detail later. It turns out that the use of this formalism naturally leads to a kind of qualitative reasoning through categories, where beliefs are updated by moving from one category to another. [1] The main contribution of this work is to study "changing worlds", where the state of the world itself changes by a Markov process with some degree of persistence across periods. At the end of the paper, we also extend the analysis of the structure of the solution when there is a fixed state of the world. We do this in the context of a decision problem where the individual has to use her inference to pick an action in every period. The discount factor is best interpreted as a continuation probability but our results and examples do not depend on this probability or discount factor being close to 1.

The structure of the optimal automaton given the constraint on memory is an example of a partially-observable Markov decision process (POMDP). We obtain some general results on POMDPs with a finite set of states of the world; however, we concentrate on the case where the set of states of the world is two because the categories of beliefs can be represented intuitively as intervals on the real line. In exploring the structure of optimal automata, we consider the phenomenon of inertia, about which we say more in the next paragraph, where a set of signals might lead to no updating even when a rational Bayesian without memory constraints would move his beliefs. Inertia could exist for any discount factor in our model. We also address the issue of randomisation by the individual decision-maker in choosing a belief category to update to, and discuss how we can take any automaton and reduce the randomisation (for which we construct a measure) to

[1] Reasoning through categories has been particularly popular in psychological studies : "To perceive is to categorize, to conceptualise is to categorize, to learn is to form categories, to make decisions is to categorize." Jerome S. Bruner, *Actual minds, possible worlds*, 1987 - Quoted by S. Gupta-Mukherjee. (2013[13])

a minimal level.

To consider the phenomenon of inertia briefly here, it could be that "small" amounts of bad news do not affect a person's beliefs, even if signals giving this bad news are repeated. It requires, in some cases, sufficiently strong negative or positive news to cause the decision-maker to change her beliefs (in this case belief categories). There could be several signals that are clustered together and ignored or there could be relatively few, depending on the discount factor. The financial crisis of 2007-8 is one of the most important series of economic events of the recent past and provides a powerful example of such inertia. A pertinent observation on the prelude to this crisis appeared in a Boston Federal Reserve discussion paper of 2010. Writing on the optimistic forecasts of housing prices prevalent at that time, Gerardi et al (2010[11]) observed that there were several regional indicators of a decline in housing prices (as well as a warning from Shiller, who is quoted in the paper) but somehow these were never sufficient to trigger a full-scale alarm about a crisis. Most forecasters did not revise their forecasts downwards as a result of these signals. Gerardi et al. also comment that, at the time when these forecasters were writing, there was nothing that seemed irrational about their forecasts and their processing of the available information. Two aspects of this description are important for us; first the absence of downward revision of forecasts as a result of "small" signals (something incompatible with a fully Bayesian analysis) and, second, the presumption that the state of the world itself was not fixed but could change for the worse, thereby causing worse outcomes to occur more frequently. Both of these are addressed in our model.

We now discuss the special case where there are two states of the world. We study this problem by considering a single economic agent[2], who has to choose one of two actions, labelled "Risky" and "Safe", in each period. The payoff from Risky depends on the state of the world and is positive in expectation in the good state of the world and negative in the bad state. Safe has a payoff of 0. However, unlike a two-armed bandit, the investor obtains a signal even if he or she chooses Safe in a given period. This seems reasonable when public economic information is available, say for a given stock, irrespective of whether one holds it or not.

The inference is constrained by the agent's strategy being given by a *finite automaton*. This has several consequences. Such a strategy operates by starting in a given memory state, choosing an action, receiving a signal and specifying a (possibly random) transition to another memory state. Therefore, only a finite number of probabilities (of the good state, say) can be considered by the agent, corresponding to the finite number of memory

[2]We first introduce and study a general class of decision problems in section 1.2. In section 3, we focus attention on the case of two states (good and bad) and two actions (risky and safe).

states, whilst Bayesian updating could give rise to a continuum of such values. Thus, there is an *implicit categorisation* of beliefs held by the agent. Second, if the agent is in a particular state of the automaton, he must have the same beliefs and take the same action. The finiteness of the number of memory states and the infinite horizon of the problem mean that she will eventually be uncertain about where in the dynamic process she is.

The present work is primarily concerned with the *changing worlds* setup, in which the state of the world changes over time, independently of the actions taken, though it might show a high degree of persistence. An early example of a fully Bayesian model of such a framework is in Shiryaev (1978[28]), where he looks at the problem of changing states and does a full Bayesian analysis of when a single decision-maker will "sound the alarm"after observing a series of signals. Each small signal will affect the Bayesian calculation of the posterior.[3]

We consider a somewhat more realistic case, where there is a degree of persistence in the state of the world from period to period. Note that the absence of full persistence means that arbitrarily long histories of signals will become irrelevant for deciding on an action today, since the state of the world is likely to have changed during this history.

We first establish a *revelation principle*[4] for this general environment, which states that at any memory state reached with positive probability, the beliefs in that memory state make it optimal to be there.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 of the fixed point. This gives us the possibility that even under the optimal strategy, the same (say, somewhat negative) signal may not have a significant impact and therefore will not change the optimal action by the agent.

We also demonstrate conditions under which the unconstrained optimal decision rule can be implemented by a deterministic finite state automaton. Thus sometimes the con-

[3]The most extreme example of changing worlds is the case where the state of the world is independent and identically distributed across time periods. Here a two-state automaton suffices to implement the optimal action. This is because, given a prior and given that it is optimal to play the safe action when the probability is below a threshold and the risky action otherwise, the same set of signals would be mapped into state 1, where the agent chooses, say, the safe action. If a signal outside this set arises, there will be a transition to the other state and a choice of the risky action.

[4]We shall discuss similar results obtained by others in the section on related literature.

straints don't prevent the unconstrained optimal action from being chosen.

Our next result does not depend on persistence of the state of the world. It is another result on the structure of the automaton, specifically on the use of randomised transitions. Earlier work has shown that, for the single agent problem, randomisation in actions was unnecessary but randomised transitions might need to be used. In an example, Hellman and Cover (who used a different objective function from the one we use in this work) showed in an example that such randomisation could substitute for additional memory states. This suggests that "more" randomisation is, roughly speaking, equivalent to a more complex automaton strategy, namely one with more memory states. In order to make this intuition more precise, we define a measure of randomisation and an upper bound on this measure. We then show that, given any automaton, we can construct a payoff-equivalent strategy that will minimise the randomisation required for playing that strategy (using our measure). As the number of signals goes to infinity, we show that such strategies exhibit virtually no randomisation.

Turning to the case of full persistence or fixed worlds, we introduce a weaker criterion (than optimality) for interim rationality of an automaton that we call "weak admissibility". The criterion involves a very simple necessary condition in terms of dominance, essentially that the strategy given by the automaton not be strictly inferior in payoff terms for any state of the world. 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 that 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.

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[16]) which solves the simple hypothesis problem with finite memory, but using a limit-of-the-means objective function rather than discounted lifetime utility.

In the field of economics, the first paper to study learning with finite automata is Wilson (2014[30]). Related problems are studied by Kocer (2010[21]), Kocer (2010[20]), Monte (2007[23]) and Monte and Said (2014[23]). Kocer (2010[21]) studies a bandit problem with finite memory and Monte (2007[22]) studies a repeated sender receiver game with the receiver using a finite state automaton. The study of decision making under finite state automata is closely related to the study of decision problems under imperfect recall as in Piccione and Rubinstein (1997[25]); Aumann, Hart and Perry (1997[3]) and Halpern (1997[14]).

We discuss certain points of comparison with the preceding papers. Interpreting the discount factor as a continuation probability as in Kocer (2010[20]), 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[30])) investment problem with *changing worlds* (the state of the world evolves non-interactively according to a Markov chain), this result is very useful in studying characteristics of the optimal automaton. We also derive conditions under which the optimal decision rule with unconstrained memory can be implemented by a deterministic finite state automaton. As in Monte and Said (2014[23]), this demonstrates the lack of need for any sophistication in the decision rule in special environments. However, they focus on symmetric environments with two signals where two-state automata suffice to implement the unconstrained optimum. Furthermore, the focus of Monte and Said (2014) is on a subclass of automata satisfying certain monotonicity, symmetry and irreducibility properties. They show that, restricting attention to this class, the optimal three-state automaton may be better than the optimal two-state automaton. The focus in our work is on a general characterisation of optimal automata and applying it to study phenomena related to inertia in beliefs and non-responsiveness to weak bad news. 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)[28]) allows for optimal unconstrained finite memory under simple conditions. In the context of constrained optimal automata, we provide an example that compares an agent's reaction to intermediate signals in a fixed world as in Wilson (2014[30]) and a highly persistent environment. Finally, classification of beliefs into categories is a well-known heuristic method of learning and conceptualisation, according to psychologists. The notion of such classification has also been used in the economics literature in varying contexts by Jackson and Fryer (2008[9]), Dow (1991[7]), Al Najjar and Pai (2014[2],

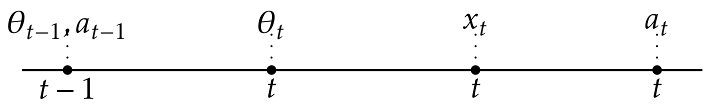


Figure 1: Timeline

Daskalova and Vriend(2014[6])).

1.2 Outline

The outline of the paper is as follows : Section 2 describes the general framework and class of decision problems we consider and further, introduces decision rules using finite state automata and notions of optimality and interim rationality. Section 3 introduces the model with changing worlds and we study characteristics of unconstrained and constrained optimal finite state automata. Section 4 discusses the issue of randomisation required in an automaton. Section 5 introduces the criterion of weak admissibility in relation to optimality and the extent of randomisation involved. The proofs for all our results can be found in the appendix.

2 Framework

In the present framework, time is discrete and the horizon is infinite i.e the environment consists of infinitely many stages $t = 0, 1, 2, 3, \ldots$. There is a finite set of states of the world Θ. The state $\theta_t \in \Theta$ evolves randomly over time and is unobservable. There is a single economic agent who takes an action every period from a finite set A and receives per period payoffs according to a state-dependent 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, though unobservable, can be inferred from signals received from a finite set $X = \{x_1, x_2, \ldots, x_n\}$. 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 $P(\theta', x|\theta, a) \equiv P(\theta, a)(\theta', x)$. Finally, there is a prior on the state of world $\pi^0 \in \Delta(\Theta)$. The diagram in figure 1 is a timeline depicting the sequence of events.

Based on information available, the agent chooses a sequence of actions in order to maximise long run discounted expected value of payoffs given by

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

where $0 < \delta < 1$ is the discount factor. The class of decision rules that we focus on are finite state automata which involve randomization. The next section introduces such decision rules and discusses the nature of decision making induced by them.

2.1 Finite state automaton and optimality

Let $M = \{1,...,m\}$ be a finite set of *memory states*. An *automaton* is defined as a tuple $\tau =< \sigma, d, g_0 >$ where $\sigma : M \times X \to \Delta(M)$ is a *transition function*, $d : M \to \Delta(A)$ is a *decision function* and $g_0 \in \Delta(M)$ is the *initial distribution* over memory states. The automaton makes decisions as follows : At a given period, when the agent is at memory state $i \in M$, he decides whether to choose the safe or risky action via the decision function $d(i)$. If he observes a signal $x \in X$, he updates the memory state randomly according to probabilities given by $\{\sigma(i,x)(j)\}_{j\in M}$ and depending on the new memory state $j \in M$ chooses the next action according to $d(j)$. Define $\mathcal{M}$ to be the set of all finite state automata defined on M. Denote the ex-ante expected long run payoff from the automaton τ to be $V(\tau)$. An optimal automaton is one which solves the maximisation problem

$$\max_{\tau\in\mathcal{M}} V(\tau).$$

Viewed as a decision rule, a finite state automaton embodies the notion of the decision maker having finite memory. One can interpret the set of memory states to be a finite set of summary statistics into which all available information (by aggregating signals) is mapped. At any point in time, the decision maker only remembers the current memory state which in turn completely describes current and future behaviour. One important observation is that this class of decision rules subsumes decision rules based on bounded recall i.e remembering only the last n signals. The memory states corresponding to the automaton will be X^n and the transition rule would be given by a deterministic function $\sigma(x_1,..,x_n)(y) = (x_2,...,x_n,y)$. Hence, for bounded recall, the transition rule is predetermined whereas it is flexible for an automaton. Another useful observation to make is that an agent using a finite state automaton finds himself in a decision problem with imperfect recall. This relates the present framework with the one considered by Piccione and Rubinstein (1997[25]), who consider general decision problems with imperfect recall with a finite extensive form. In addition to optimality, they discuss a notion of interim rationality called *modifed multiself consistency* which applies also to the present context.

Once an automaton is fixed, the long run payoffs are completely determined by the evolution of the pair (θ, i) over time since actions only depend on $i \in M$. This pair evolves according to a Markov process :

$$Q(\theta', i'|\theta, i) = \sum_{a \in A} d(i)(a) \sum_{x \in X} P(\theta', x|\theta, a)\sigma(i, x)(i'). \tag{2}$$

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 :

$$\hat{\mu}(\theta, i) = \sum_{t=0}^{\infty} (1-\delta)\delta^t \mathbb{P}^t(\theta, i). \tag{3}$$

where $\mathbb{P}^t(\theta, i)$ denotes the probability of the pair being (θ, i) at time t under the transitions induced by Q above in 3. It can be shown that $\hat{\mu}$ is the unique[5] invariant distribution of the perturbed Markov chain on the set $\Theta \times M$:

$$\hat{Q}(\theta', i'|\theta, i) = (1-\delta)g^0(i')\pi^0(\theta') + \delta Q(\theta', i'|\theta, i). \tag{4}$$

Denote the continuation values defined by each pair (θ, i) under an automaton as $V(\theta, i)$. Kocer (2010[20]) has shown modified multiself consistency of optimal automata under beliefs generated by $\hat{\mu}$. A formal definition is given below.

Definition 1 *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 :

$$a \in \arg\max_{a' \in A} \sum_{\theta \in \Theta} \hat{\mu}(\theta|i)[(1-\delta)u(a', \theta) + \delta \sum_{(\theta', x) \in \Theta \times X} \sum_{i' \in M} P(\theta', x|\theta, a')\sigma(i, x)(i')V(\theta', i')]. \tag{5}$$

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

$$j \in \arg\max_{k \in M} \sum_{\theta' \in \Theta} \hat{\pi}(\hat{\mu}(.|i), a, x)(\theta')V(\theta', k). \tag{6}$$

[5]Uniqueness follows from observing that for $(\theta', i') \in \Theta$ with $\pi^0(\theta')g^0(i') > 0$, we have $\min_{\theta, i} \hat{Q}(\theta', i'|\theta, i) > 0$. Hence, the Markov chain $\hat{Q}$ as an operator on $\Delta(\Theta \times M)$ is a contraction (See Lemma 11.3, Chapter 11, Stokey, Lucas and Prescott (1989[29])).

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 :

$$\begin{aligned}
\hat{\pi}(\hat{\mu}(.|i), a, x)(\theta') &= \frac{\sum_{\theta \in \Theta} \hat{\mu}(\theta|i) d(i)(a) P(\theta', x|\theta, a)}{\sum_{\theta'' \in \Theta} \sum_{\theta \in \Theta} \hat{\mu}(\theta|i) d(i)(a) P(\theta'', x|\theta, a)}, \\
&= \frac{\sum_{\theta \in \Theta} \hat{\mu}(\theta|i) P(\theta', x|\theta, a)}{\sum_{\theta'' \in \Theta} \sum_{\theta \in \Theta} \hat{\mu}(\theta|i) P(\theta'', x|\theta, a)}. \qquad (7)
\end{aligned}$$

The multiselves approach studied by Piccione and Rubinstein (1997[25]) is motivated as a requirement of time consistency. The agent in our framework moves either by taking an action at a given memory state $i \in M$ via the decision function d or by making transitions at $(i, x) \in M \times X$ via transition function σ. Hence, choices (potentially involving randomisation) are made by the agent at $|M| + |M||X|$ information sets. Modified multiself consistency requires that such choices be optimal in the interim : fixing future behavious as was initially planned, the agent does not have an incentive to deviate from (d, σ). We now state and prove a *revelation principle* for optimal automata. This establishes that if the agent reaches a particular memory state, his optimal choice is indeed to stay there and not proceed with a different initial state.

Proposition 1 *(Revelation Principle) Consider any optimal automaton τ. Let $i \in M$ such that $\hat{\mu}(i) > 0$. We have,*

$$i \in \arg\max_{k \in M} \sum_{\theta \in \Theta} \hat{\mu}(\theta|i) V(\theta, k). \qquad (8)$$

The revelation principle, along with modified multiself consistency, allows us to characterise transitions in an optimal automaton in the following manner : i) The probability simplex $\Delta(\Theta)$, which gives the set of beliefs about the state of the world, is divided into finitely many *categories*, each correspond to a memory state of the automaton reached with positive probability ii) Each category contains a *representative belief*, which is used to update beliefs about the state of the world given the action undertaken and the signal received iii) The automaton transitions with probability one to the set of categories containing the updated belief.

To represent the above formally, we first establish some notation. Define the set $\bar{\Pi}(i) =$

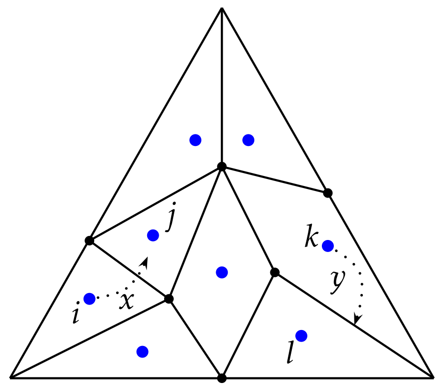


Figure 2: Nature of Characterisation

$\{\pi \in \Delta(\Theta) : \sum_{\theta\in\Theta} \pi(\theta)V(\theta,i) \geq \sum_{\theta\in\Theta} \pi(\theta)V(\theta,k)$ for all $k \in M\}$. By the revelation principle, the set $\bar{\Pi}(i)$ is non-empty for each $i \in M$ such that $\hat{\mu}(i) > 0$. The collection $\{\bar{\Pi}(i)\}_{i\in M}$ is the set of categories and $\{\hat{\mu}(.|i)\}_{i\in M}$ are the representative beliefs corresponding to them. The following result establishes the desired characterisation.

Proposition 2 *(Characterisation) The following statements hold true :*

1. *(Categorisation)* $\bigcup_{i\in M} \bar{\Pi}(i) = \Delta(\Theta)$ *and the collection of sets* $\{\bar{\Pi}(i)\}_{i\in M}$ *along with the intersections of all its sub-collections and the empty set, constitutes a polytopal complex*[6].

2. *(Representative Beliefs) For all* $i \in M$, *such that* $\hat{\mu}(i) > 0$, $\bar{\Pi}(i)$ *is non-empty and is a polytope. Furthermore, we have* $\hat{\mu}(.|i) \in \bar{\Pi}(i)$ *i.e the representative belief* $\hat{\mu}(.|i)$ *belongs to its corresponding category* $\bar{\Pi}(i)$.

3. *(Transitions) For all* $(i,x) \in M \times X$, *we have* $\sigma(i,x)(j) > 0$ *only if* $\hat{\pi}(\hat{\mu}(.|i),a,x)$ *belongs to* $\bar{\Pi}(j)$ *for all* $a \in A$ *with* $d(i)(a) > 0$.

Illustration : Let us now describe the nature of the characterisation established in Proposition 2 by means of a diagram. Suppose $|\Theta| = 3$ and $|M| = 8$. Then, we obtain the picture as in Figure 2

The different parts of the 2-dimensional simplex above respresent the categories corresponding to the memory states M of the automaton. The representative beliefs are

[6]The following definition is borrowed from Ziegler (1995[31]). A *polytope* is the convex hull co(E) of a finite set of points E in $\mathbb{R}^{|\Theta|}$. A *polytopal complex* $\mathcal{C}$ is a collection of polytopes in $\mathbb{R}^{|\Theta|}$ such that : the empty polytope is in $\mathcal{C}$; if $P \in \mathcal{C}$, then all the faces of P are also in $\mathcal{C}$; the intersection $P \cap Q$ of two polytopes $P, Q \in \mathcal{C}$ is a face of both P and Q.

depicted by the blue dots and the transitions are given by arrows. Hence, if the automaton transitions from i to j upon receiving signal x with probability one, then the updated belief $\hat{\pi}(\hat{\mu}(.|i), a, x)$ must necessarily belong to the category corresponding to memory state j. The diagram also reveals what would happen if randomisation takes place in the optimal automaton. Suppose $\sigma(k,y)(k), \sigma(k,y)(l) > 0$, then it must be the case that $\hat{\pi}(\hat{\mu}(.|k), a, y)$ belongs to the intersection of the categories corresponding to k and l.

2.2 Non-interactive Decision Problems

In this section we study the situation where the underlying state of the word $\theta \in \Theta$ and signals $x \in X$ evolve independent of actions via $P(\theta', x|\theta)$. Hence, the environment is *non-interactive* i.e the chance moves undertaken by nature do not depend on the actions taken by the agent. In such non-interactive environments, it follows that no randomisation is needed in the decisions and the initial state. Hence, we can focus on automata with a deterministic initial state and decision function i.e $\tau = <\sigma, d, \delta_{i_0}>$ where $\sigma : M \times X \to \Delta(M)$ is a transition function as before but now $d : M \to A$ and $i_0 \in M$.

Proposition 3 *Suppose the environment is non-interactive. For any automaton τ, there exists an automaton τ' with a deterministic initial state and decision function such that $V(\tau) \leq V(\tau')$.*

The above proposition allows us to focus attention, without loss of generality, on the sub-class of automata which involve no randomization in the decisions and initial memory state. However, for the transition function, randomization is needed. This is a very crucial feature of the environment we consider and Section 4 is devoted to analysing the question of how much randomization is needed when the signal space X is sufficiently large. In what follows, we shall refer to any automaton with a deterministic transition function as a *deterministic automaton*.

Due to randomisation in the transition function, any sequence of signals is mapped stochastically to a memory state. The following result states that any automaton with memory states M essentially *garbles* the true sequences of signals[7]. For any history of signals $h = (x_1, x_2 ..., x_t) \in X^t$ reached with positive probability, let $\hat{\pi}(.|h) \in \Delta(\Theta)$ denote the belief the agent holds about state of the world in time period t after h (in the unconstrained memory case when Bayes' rule can be applied). Further, let $X^* = \bigcup_{t \geq 0} X^t$ denote the set of all signal histories.

[7] The terminology is borrowed from Blackwell (1951[4],1953[5])

Proposition 4 *(Garbling) : Let $< g^0, \sigma, d >$ be an automaton with memory states M and let $\hat{\mu} \in \Delta(\Theta \times M)$ denote the stationary distribution of the perturbed markov chain as defined in 4. There exists a joint distribution $\nu \in \Delta(\Theta \times X^* \times M)$ such that :*

1. *$\nu(\theta|h) = \hat{\pi}(\theta|h)$ for all $h \in X^*$ reached with positive probability*
2. *$\nu(\theta|i) = \hat{\mu}(\theta|i)$ for all $i \in M$ reached with positive probability*
3. *$\nu(i|\theta, h) = \nu(i|h)$ i.e. $\nu(i|\theta, h)$ is independent of the value of θ. Hence, there exists a stochastic map $\hat{\sigma} : X^* \to \Delta(M)$ such that $\hat{\sigma}(h)$ equals the probability vector $\{\nu(i|\theta, h)\}_{i \in M}$*

For any transition function $\sigma : M \times X \to \Delta(M)$, define the stochastic map $\hat{\sigma} : X^* \to \Delta(M)$ by $\sigma(x_1, ..., x_t)(i) := \sum_{i_0 \in M} g^0(i_0) \sum_{i^t \in M^t : i_t = i} \prod_{\tau=1}^{t} \sigma(i_{\tau-1}, x_\tau)(i_\tau)$. Hence, if the true history of signals is $h = (x_1, ..., x_t) \in X^*$, then $\hat{\sigma}(h)(i)$ is the probability that the automaton will be in memory state i at time period t. The following is a useful corollary of the above proposition :

Corollary 5 *Consider any automaton τ and a convex subset $\Pi \subseteq \Delta(\Theta)$. Let $i \in M$ such that $\hat{\mu(i)} > 0$ and suppose $\{\pi(.|h)|\ h \in X^*$ and $\hat{\sigma}(h)(i) > 0\} \subseteq \Pi$. Then, it is the case that $\hat{\mu}(.|i) \in \Pi$*

3 Changing Worlds and Optimal Finite Memory

In the rest of the paper, we consider a specific class of decision problems involving two states of the world that evolve according to a markov process. Hence, $\Theta = \{0, 1\}$ and θ_t evolves over time according to Markov process $P(\theta'|\theta)$. The transition probabilities are as follows :

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

The decision problem faced by the agent involves, every period, to choose either a *risky* action r or a *safe* action s. Therefore, the action set is $A = \{r, s\}$. The utilities from actions are defined as follows : $u(r, \theta) = E_\theta$ and $u(s, \theta) = 0$ 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 structure $< f_0, f_1 >$. By informative, we mean $f_0 \geq f_1$. Further, we assume $f_\theta \in \text{int}(\Delta(X))$ for each $\theta \in \Theta$. This implies $f_\theta(x) > 0$ for each $(\theta, x) \in \Theta \times X$. Hence, there exists no signal in X that would fully reveal the underlying state. When Δ, ϵ are small, the decision problem can be imagined as one close to the case

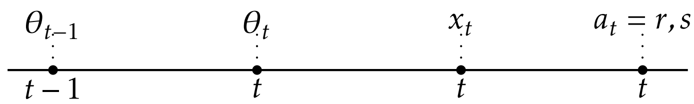


Figure 3: Timeline

where the state of the world is determined at the beginning of time and stays unchanged henceforth. Figure 3 is a timeline for the sequence of events :

Such models are also called hidden Markov models and fall under the category of non-interactive partially observable Markov decision processes as introduced in Section 2. The optimal decision rule in the case of unconstrained memory is simple : At time t, if the belief about the current state being $\theta = 0$ is π_t, then choose r if and only if $\pi_t \geq \pi^*$, where π^* is such that $\pi^* E_0 + (1-\pi^*)E_1 = 0$. In the subsequent sections 3.1 and 3.2, we study i) when in such environments the unconstrained optimal decision rule is implementable by a finite state automaton ii) characteristics of the optimal automaton under memory constraints. Furthermore, we assume for the rest of this section, that $\Delta, \epsilon \in (0,1)$. Hence, the environment involved *changing worlds* i.e the state of the world can, with positive probability, change or stay the same.

3.1 Unconstrained optimal automata

In this case, 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 itself. At a given time if the belief about $\theta = 0$ is π , then the belief about the state next period, after receving the signal x will be :

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

If $\Delta = \epsilon = 0$, we would enter an environment with a *fully persistent* state of the world. In such environments, it is impossible to implement the unconstrained decision rule by any finite state automaton[8]. The changing worlds environment is special in the sense that

[8]This follows from the well-known Myhill-Nerode Theorem from the theory of computation (See Hopcroft, Motwani and Ullman (2006[18])). If the unconstrained optimal decision rule were indeed implementable, then the set $\mathcal{S} = \{h \in \bigcup_{t=0}^{\infty} X^t : \hat{\pi}(\pi_0, h) < \pi^*\}$ must be computable by some finite state automaton. From the Myhill-Nerode theorem, the following equivalence relation on X^* would then have finitely many equivalence classes : $h \sim h'$ if there exists $h'' \in \bigcup_{t=0}^{\infty} X^t$ such that either $\{h \circ h'', h' \circ h''\} \subseteq \mathcal{S}$ or $\{h \circ h'', h' \circ h''\} \subseteq \mathcal{S}^c$. However, under full persistence, this is not possible.

in some instances, even the unconstrained optimal decision rule can be implemented by a finite state automaton. The following result and its corollary demonstrate that even with unconstrained memory, only a finite memory capacity is needed to attain the optimal payoff under some conditions :

Proposition 6 *For any informative signalling structure $< f_0, f_1 >$ with full support, let $\bar{x} \in \arg\max_{x \in X} \dfrac{f_0(x)}{f_1(x)}$ be the signal with the highest likelihood ratio. Let $\bar{\pi}_{\bar{x}}$ be such that $\hat{\pi}(\bar{\pi}_{\bar{x}}, \bar{x}) = \bar{\pi}_{\bar{x}}$. If $\bar{\pi}_{\bar{x}} < \pi^*$, and $1 - \epsilon > \Delta$, then the optimal decision rule can be implemented by a deterministic finite state automaton.*

Monte and Said (2014[23]) also study a changing worlds environment under bounded memory and provide sufficient conditions under the optimal decision rules is implementable by a two-state automaton which simply responds the current period signal. The condition we provide here differs from theirs and further, in the present case, the implementation is via an automaton which keeps track of signals till a time period T after which the agent permanently switches to the safe action. This would require $|\cup_{t \geq 0} X^t|$ many memory states in the automaton.

The dynamics of the belief process are governed by the position of the fixed points $\{\bar{\pi}_i\}_i$ corresponding to the signals $X = \{x_i\}_i$ (ordered by i according to likelihood ratio) defined by the condition $\hat{\pi}(\bar{\pi}_i, x_i) = \bar{\pi}_i$. For each signal $x \in X$, the function $\hat{\pi}(., x)$ is strictly increasing and strictly concave (convex) if and only if the likelihood ratio of x is strictly greater (less) than one. Furthermore, the fixed points are globally stable in a specific sense. If the belief is to the left (right) of a fixed point i.e $\pi < \bar{\pi}_i$ ($\pi > \bar{\pi}_i$), then perpetually receiving signal x_i keeps beliefs to the left (right) and ultimately converges to $\bar{\pi}_i$.

3.2 Constrained optimal automata

In the previous, 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 an agent not bounded by memory constraints may be replicated by one who faces bounds. Here, we consider the constrained problem, where the agent chooses the optimal automaton with a fixed number of states. The definition of an automaton $\tau =< \sigma, d, i_0 >$ and optimality are as in section 2. 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 :

$$Q(\theta', i'|\theta, i) = P(\theta'|\theta) \sum_{x \in X} f_{\theta'}(x)\sigma(i,x)(i') \tag{10}$$

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

$$Q(\theta', i'|\theta, i) = P(\theta'|\theta)\alpha_{ii'}^{\theta'} \tag{11}$$

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[20]) 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 :

$$\begin{aligned} \hat{Q}(\theta', i'|\theta, i) &= (1-\delta)\pi_0(\theta')\delta_{i_0}(i') + \delta P(\theta'|\theta)\alpha_{ii'}^{\theta'} \\ &= (1-\delta)\pi_0(\theta')\delta_{i_0}(i') + \delta Q(\theta', i'|\theta, i) \end{aligned} \tag{12}$$

In this setting of two states of the world, an automaton τ is said to be *modified multiself consistent* if for the invariant distribution $\hat{\mu}$ (corresponding to Markov chain $\hat{Q}$), the following are satisfied :

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

$$d(i) \in \arg\max_{a \in \{r,s\}} [\hat{\mu}(0|i)E_0 + \hat{\mu}(1|i)E_1]\mathbb{I}_{\{a=r\}} \tag{13}$$

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

$$j \in \arg\max_{k \in M} [\hat{\pi}(\hat{\mu}(0|i), x)V(0,k) + (1 - \hat{\pi}(\hat{\mu}(0|i), x))V(1,k)] \tag{14}$$

It is possible to modify the optimal automaton in a way such that the values generated by the memory states with positive probability in the invariant distribution are all different. If two memory states i and j provide the same the continuation values i.e $(V(0,i), V(1,i)) = (V(0,j), V(1,j))$, then modify the automaton shifting all transitions to i in the original automaton to memory state j instead. Now, if we interpret each j as a decision taken under uncertainty yielding payoff $V(\theta, j)$ in state $\theta \in \Theta$, modified multiself consistency tells us that states j with $\hat{\mu}(j) > 0$ are admissible[9] in the set :

$$\mathcal{V} = \{(V(0,k), V(1,k)) \in \mathbb{R}^2 : k \in M\}$$

The revelation principle from Section 2 here implies that at the belief $\hat{\mu}(0|i) > 0$, staying at i and receiving continuation values $(V(0,i), V(1,i))$ is optimal. In the current framework with two states and the state of the evolving according to a Markov chain, we obtain the following characterisation :

Proposition 7 *(Categorisation) : For the optimal automaton τ^*, 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 $\hat{\pi}(\hat{\mu}(0|i), x) \in [\underline{\pi}_j, \underline{\pi}_{j+1}]$*

The following example illustrates the phenomenon of inertia. Unfavourable signals, which would influence a unconstrained Bayesian's to move downwards, would not change beliefs.

Example (*Two-state Automaton*) : Consider the case of an optimal automaton with two memory states $M = \{L, H\}$ and $X = \{l, m, h\}$. In this case, both memory states will be used since one state automata always take the same action. From Proposition 7, we can argue that there exists a point of division $\underline{\pi} \in (0,1)$ such that transitions to L are optimal for beliefs in $[0, \underline{\pi}]$ and transitions to H are optimal on $[\underline{\pi}, 1]$. That the point $\underline{\pi}$ is in the interior follows from the fact that $0 < \hat{\mu}(0|L) \leq \underline{\pi} \leq \hat{\mu}(0|H) < 1$ since memory states can reached with positive probability in both the good and bad state. Now, suppose it is the case that 1) $\frac{f_0(l)}{f_1(l)} < \frac{f_0(m)}{f_1(m)} < 1 < \frac{f_0(h)}{f_1(h)}$ 2) $\bar{\pi}_m < \pi^* < \underline{\pi} < \hat{\pi}(\hat{\mu}(0|H), m) < \hat{\mu}(0|H)$ and 3) $d(L) = s$ and $d(H) = r$. Hence the automaton stays in memory state H and chooses the risky action after receiving an unfavourable signal m and would continue to do so if m is continually received. In constrast, an unconstrained Bayesian starting at belief $\hat{\mu}(0|H)$ would eventually switch his action from risky to safe. The appendix provides sufficient conditions under which any optimal two-state automaton would exhibit such features. Moreover, this is possible for any fixed discount factor $0 < \delta < 1$ and likelihood ratio for the intermediate signal m. Figure 4 illustrates the transitions. The green, orange and red squares depict the fixed points corresponding to the signals h, m, l respectively.

A full analysis of the example is contained in the appendix and parametric conditions under which inertia occurs are provided. This extends even to the case of any signal space

[9]For a subset $\mathcal{V} \subseteq \mathbb{R}^n$, a point $x^* \in V$ is *admissible* if there does not exist another $x \in V$ such that x weakly dominates x^* coordinate-wise.

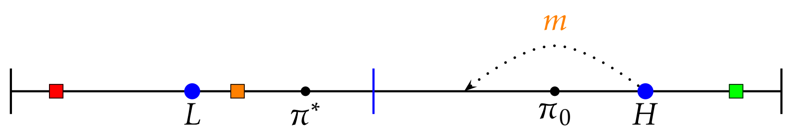


Figure 4: Inertia in optimal two-state automaton

where there exist clusters of signals i.e $X = X_l \cup X_m \cup X_h$ such that $X_i \neq \emptyset$ and $X_i \cap X_j = \emptyset$ for $i \neq j$ where $i, j \in \{l, m, h\}$. Here, X_l and X_m (X_h) are to be interpreted as a class of low and intermediate (high) signals which are indicative of the bad state of the world i.e all of them have likelihood ratio less than 1. However, signals in X_l are stronger (lower likelihood ratio) compared to X_m. Under a counterpart of parametric conditions, the optimal automaton involves all signals in X_m being ignored and also admits that any signal in X_l leads to playing the safe action. Hence, intermediate signals are not taken into account whereas the agent is reactive to a class of worse signals. Moreover, this is again consistent with any discount factor and high levels of persistence in the state of the world. This is in contrast with the results in Hellman and Cover (1970[16]) and Wilson (2014[30]) who establish that with high discount factors all but the only two extreme signals would be ignored.

4 Limits to randomization

The class of decision problems we consider in this section are general non-interactive problems introduced in section 2.1. In such problems, Proposition 3 allows us to focus on automaton which only require randomisation in the transition function. We analyse here, the extent to which randomisation is actually needed. This issue has been studied by Kalai and Solan (2003[19]) and Hellman and Cover (1971[17]). 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. The lesson from both of these papers is that randomisation, although not required in the decisions and initial state, is crucial for the transitions[10]. In

[10] See also Hellman (1972[15]) who considers the flip side of the issue. The paper argues that under full persistence (with the limit of means criterion), for every m there exists K such that the optimal m-state stochastic automaton gives a lower payoff than the optimal $m+K$-state deterministic automaton

a related context, Piccione and Rubinstein (1997[25]) establish that there is strict benefit from randomisation in decision problems of imperfect recall. 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 introduce our randomisation measure below :

Definition 2 *For each state $i \in M$, consider the number $n_i^\tau = |\{(x,j) : \sigma(i,x)(j) > 0\}|$. The measure of randomization is :*

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

For any deterministic automaton τ_d, we have $n(\tau_d) = \dfrac{1}{|M|}$ and observe that deterministic automata provide the lowest level of randomisation according to the measure. 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. Before proceeding towards our result, we first need to define the notion of equivalence between two automata. This notion is stronger than payoff equivalence and requires that conditional on θ, the total probability of transitioning between two memory states be equal for the two automata.

Definition 3 *Two finite state automata τ and τ' are said to be equivalent if they generate the same transition probabilities i.e. $\alpha_{ij}^\theta = \alpha_{ij}'^\theta$ for all $\theta \in \Theta$, have the same decision function and initial state.*

The main result of this section is the following.

Proposition 8 *Suppose $|M||X| - |\Theta||M| - |X| \geq 0$, then for any automaton τ, there exists an equivalent automaton τ' such that :*

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

Note that as $|X|$ approaches infinity, the upper bound converges to the measure of randomisation in any deterministic automaton. Moreover, the result is true for any arbitrary automaton and not necessarily one which is optimal. One may view this as a "purification" result for this class of decision rules.

Remark : Suppose we focus attention on the changing worlds framework with two states of the world as introduced in section 3. From Proposition 9, it follows that in the optimal $|M|$-state automaton, due to the partitioning of the unit interval into further subintervals, the transition function would randomise between atmost two memory states. These would correspond to two adjacent categories and randomisation between them would take place if the updated belief $\hat{\pi}(\hat{\mu}(0|i), x)$ is exactly their common end point. This observation does not generalise to the case of more than two states of the world. From Figure 2, we observe that with three states of the world there could be transitions which randomise at more than three memory states. In general, we cannot obtain an upper bound on randomisation simply by virtue of optimality. However, appealing to the above result, we can argue that if there were a large number of signals available, any automaton (not necessarily optimal) would involve low amount of randomisation.

5 Fixed worlds, weak admissibility and structure of optimal automata

In this section, we consider the case the state of the world stays fixed i.e $P(\theta|\theta) = 1$. We define a new notion of interim rationality for an automaton to be obeyed in the context of the decision problem considered in section 3 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 $M = \{1, ..., m\}$. An automaton τ generates the following equations for deriving the continuation values :

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

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 :

$$\begin{aligned} \max_{\alpha} \quad & \sum_j \alpha_j V_j^0 \\ \text{s.t.} \quad & \alpha \in K(\beta_i) \end{aligned} \tag{15}$$

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 4 *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 :*

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

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 the other in terms of continuation values. We present a formal definition below.

Definition 5 *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* $(V_j^0, V_j^1) \ngeq (V_i^0, V_i^1)$

5.1 Structure of weakly admissible automata with no wastage

The optimization problem posed by the weak admissibility 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 :

$$\begin{aligned} \max_{\sigma} \quad & \sum_j \sum_x \sigma(x,j) f_0(x) V_j^0 \\ \text{s.t.} \quad & \sum_x \sigma(x,j) f_1(x) = \beta_j \ \forall j \\ & \sum_j \sigma(x,j) = 1 \ \forall x \\ & \sigma(x,j) \geq 0 \ \forall x, j. \end{aligned} \tag{16}$$

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 sum up to a vector of ones and the weighted sum of rows (according to weights $f_1 \in \Delta(X)$) equals the vector β. It can be shown the above problem is equivalent to the following problem :

$$\begin{aligned} \max_{u} \quad & \sum_j \sum_x u(x,j) \frac{f_0(x)}{f_1(x)} V_j^0 \\ \text{s.t.} \quad & \sum_x u(x,j) = \beta_j \ \forall j \\ & \sum_j u(x,j) = f_1(x) \ \forall x \\ & u(x,j) \geq 0 \ \forall x, j. \end{aligned} \tag{17}$$

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 $\{(V_j^0, V_j^0) : j \text{ such that } \beta_j > 0\}$ are distinct and for any two vectors, neither weakly dominates the other coordinate-wise. 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 :

$$\sigma_i = \begin{pmatrix} * & * & 0 & 0 & \dots & 0 \\ 0 & * & 0 & 0 & \dots & 0 \\ 0 & * & * & * & \dots & 0 \\ 0 & 0 & 0 & * & \dots & 0 \\ \vdots & & & & \vdots & \ddots \\ 0 & 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & \dots & * \\ \end{pmatrix}. \quad (18)$$

Figure 5: Staircase structure

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 column components i.e $j's$ is done via the continuation values V_j^0 (from no wastage this ordering is unique).

This leads us to the following proposition :

Proposition 9 *Every automaton that is weakly admissible and satisfies no wastage necessarily admits a transition function of a staircase form as in 5. Moreover, it is the unique solution to the LP problem imposed by the weak admissibility constraints.*

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

Proposition 10 *For every optimal automaton τ, there exists an equivalent automaton with the same ex-ante value that is weakly admissible and satisfies no wastage.*

In Section 4, we introduced a measure of randomisation for finite state automata and established a universal upper bound on the extent of randomisation involved. In the present context, substituting the $|\Theta| = 2$, the upper bound becomes $\frac{1}{|M|} + \frac{2}{|X|}$. We can now show that weakly admissible automata with no wastage improve the upper bound leading to the following result :

Proposition 11 *Let τ be a weakly admissible automaton with no-wastage, then :*

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

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 Gennaioli, Shleifer and Vishny (2013[10])) 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. Suppose, for example, that there are three signals and fifty states with the first two signals being "bad news" 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 finanancial 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 implied by weak admissibility. 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 Hellman and Cover (1971[17]) and Kalai and Solan (2003[19]).

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] Blackwell, David. "Comparison of experiments." Proceedings of the second Berkeley symposium on mathematical statistics and probability. Vol. 1. 1951.

[5] Blackwell, David. "Equivalent comparisons of experiments." The annals of mathematical statistics 24.2 (1953): 265-272.

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

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

[8] Ferguson, Thomas S. "Mathematical statistics." (1967).

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

[10] Gennaioli, Nicola, Andrei Shleifer, and Robert W. Vishny. "A model of shadow banking." The Journal of Finance 68.4 (2013): 1331-1363.

[11] Gerardi, Kristopher, Christopher L. Foote, and Paul Willen. ”Reasonable people did disagree: Optimism and pessimism about the US housing market before the crash.” FRB of Boston Public Policy Discussion Paper 10-5 (2010).

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

[13] Gupta-Mukherjee, Swasti, et al. ”Categorical Thinking in Portfolio Choice.” (2013).

[14] Joseph Y Halpern, On Ambiguities in the Interpretation of Game Trees, Games and Economic Behavior, Volume 20, Issue 1, 1997, Pages 66-96

[15] Hellman, Martin. ”The effects of randomization on finite-memory decision schemes.” IEEE Transactions on Information Theory 18.4 (1972): 499-502.

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

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

[18] Hopcroft, John E., Rajeev Motwani, and Jeffrey D. Ullman. ”Automata theory, languages, and computation.” International Edition 24 (2006).

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

[20] Kocer, Yilmaz, ”Optimal Plans with Bounded Memory are Modified Multiself Consistent”, working paper 2010

[21] Kocer, Yilmaz. ”Endogenous learning with bounded memory.” Economic Theory Center Working Paper 001-2011 (2010).

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

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

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

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

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

[27] Schrijver, A., ”A Course in Combinatorial Optimization” 2013

[28] Shiryaev, Albert N., ”Optimal stopping rules” *Springer* 1978

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

[30] Wilson, Andrea. ”Bounded memory and biases in information processing.” Econometrica 82.6 (2014): 2257-2294.

[31] Ziegler, G. M. ”Lectures on polytopes.” Vol. 152. Springer Science & Business Media, 2012.

7 Appendix

7.1 Proofs from Section 2

Proof of Proposition 1 : Suppose not. Then there exists a $j \in M \backslash \{i\}$ such that :

$$\sum_{\theta \in \Theta} \hat{\mu}(\theta|i) V(\theta, j) > \sum_{\theta \in \Theta} \hat{\mu}(\theta|i) V(\theta, i) \tag{19}$$

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 :

$$\nu(\theta, l, s, \theta'', j) := \hat{\mu}(\theta, l)(1-\delta)\pi(\theta''^0(j) \text{ for all } (\theta, k, s, \theta'', j) \in \Theta \times M \times \{s\} \times \Theta \times M$$
$$\nu(\theta, l, n, \theta'', y, j) := \hat{\mu}(\theta, l)\delta d(l)(a) P(\theta'', x|\theta, a)\sigma(l, y)(j) \text{ for all}$$
$$(\theta, l, n, a, \theta'', y, j) \in \Theta \times M \times \{n\} \times A \times \Theta \times X \times M$$

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 :

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

2. 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 $\Pi(j,i)^c$.

3. We now prove the result. There are two cases to consider :

   (a) *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).

   (b) *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).

Proof of Proposition 2 :

1. Let $\mathcal{V} = \{V(.,j) \in \mathbb{R}^{\Theta} : j \in M\}$. Now, let $P = \text{conv}(\mathcal{V})$ be the polytope defined by the convex hull of $\mathcal{V}$. For each face F of the polytope, define the cone

   $N_F = \{c \in \mathbb{R}^{\theta} : c.w \geq c.w'$ for all $w \in F$ and $w' \in P\}$.

   Let $\mathcal{F}(P) = \{F : F$ is a face $P\}$. The collection of cones $\{N_F\}_{F \in \mathcal{F}(P)}$ is known as the normal fan of the polytope P and is a polyhedral complex (See Chapter 7, Ziegler (2012[31])). It then follows that the collection $\{N_F \cap \Delta(\Theta) : F \in \mathcal{F}(P)\}$ is a polytopal complex. We shall now show that collection $\{N_F \cap \Delta(\Theta) : F \in \mathcal{F}(P)\} \cup \{\emptyset\}$ equals $\{\cap_{k \in M'} \bar{\Pi}(k) : \emptyset \neq M' \subseteq M\} \cup \{\emptyset\}$. Consider any set $N_F \cap \Delta(\Theta)$. Now, since F is a face of P, it follows that there is subset $M' \subseteq M$ such that

   $N_F = \{c \in \mathbb{R}^{\theta} : c.V(.,k) \geq c.w'$ for all $k \in M'$ and $w' \in P\}$.

   where $\{V(.,k)\}_{k \in M'}$ is the set of extreme points (vertices) of F. Clearly, $N_F \cap \Delta(\Theta) = \cap_{k \in M'} \bar{\Pi}(k)$. Now, consider any set of the form $\cap_{k \in M'} \bar{\Pi}(k)$. If the set is empty, then we are done. Hence, assume that the set is non-empty. Let $\pi \in \cap_{k\in M'}\bar{\Pi}(k)$. Now consider the face $F' = \{w \in P : w.\pi \geq w'.\pi \text{ for all } w' \in P\}$ of the polytope P. There exists a subset $\mathcal{V}' \subseteq \mathcal{V}$ such that $F' = \mathrm{conv}(\mathcal{V}')$. It follows that $N_{F'} \cap \Delta(\Theta) = \cap_{k\in M'}\bar{\Pi}(k)$

2. Now consider $i \in M^+$. From the revelation principle in Proposition 1, it is true that $\hat{\mu}(.|i) \in \bar{\Pi}(i)$. This also establishes that the set $\bar{\Pi}(i)$ is non-empty.

3. Follows directly from modified multiself consistency.

Proof of Proposition 4 : Let $\theta \in \Theta$, $i \in M$ and $h = (x_1, x_2..., x_t) \in X^*$ reached with positive probability. We define the joint distribution as :

$$v(\theta,h,i) =$$
$$(1-\delta)\delta^t\Big[\sum_{\theta_0\in\Theta}\pi^0(\theta_0)\sum_{\theta^t\in\Theta^t:\theta_t=\theta}\prod_{s=1}^{t}P(\theta_s,x_s|\theta_{s-1})\Big]\Big[\sum_{i_0\in M}g^0(i_0)\sum_{i^t\in M^t:i_t=i}\prod_{\tau=1}^{t}\sigma(i_{\tau-1},x_\tau)(i_\tau)\Big]$$

Here, $\pi^0 \in \Delta(\Theta)$ denotes the prior on the states of the world. In order to interpret the above expression, consider the following interpretation of the decision problem : 1) Nature chooses the horizon length t with probability $(1-\delta)\delta^t$ 2) Once the horizon length is determined, the states of world and signals about the states of the world are determined according to $P(\theta', x|\theta)$. Hence, the probability that the state of world will be θ at time t and signals $h = (x_1, x_2..., x_t)$ will be generated at time period $1,..t$ is given by $C(\theta,h) = \Big[\sum_{\theta_0\in\Theta}\pi^0(\theta_0)\sum_{\theta^t\in\Theta^t:\theta_t=\theta}\prod_{s=1}^{t}P(\theta_s,x_s|\theta_{s-1})\Big]$ 3) Once the signals are determined, the memory states of the automaton evolve according to the transition function σ. The probability that the automaton will be in memory state i at time t upon receiving signals $h = (x_1, x_2..., x_t)$ is $D(h,i) = \Big[\sum_{i_0\in M}g^0(i_0)\sum_{i^t\in M^t:i_t=i}\prod_{\tau=1}^{t}\sigma(i_{\tau-1},x_\tau)(i_\tau)\Big]$. Hence, the joint distribution ν can be written as :

$$\nu(\theta,h,i) = (1-\delta)\delta^{|h|}C(\theta,h)D(h,i)$$

Consider part 1 of the proposition. Note that $\nu(\theta,h) = (1-\delta)\delta^t C(\theta,h)$ and hence :

$$\begin{aligned}
\nu(\theta|h) &= \frac{(1-\delta)\delta^{|h|}C(\theta,h)}{\sum_{\theta'}(1-\delta)\delta^{|h|}C(\theta',h)} \\
&= \frac{\sum_{\theta_0\in\Theta}\pi^0(\theta_0)\sum_{\theta^t\in\Theta^t:\theta_t=\theta}\prod_{s=1}^{t}P(\theta_s,x_s|\theta_{s-1})}{\sum_{\theta'\in\Theta}\sum_{\theta_0\in\Theta}\pi^0(\theta_0)\sum_{\theta^t\in\Theta^t:\theta_t=\theta'}\prod_{s=1}^{t}P(\theta_s,x_s|\theta_{s-1})} \\
&= \hat{\pi}(\theta|h)
\end{aligned}$$

Now consider part 2. Note that :

$$\begin{aligned}
\nu(\theta,i) &= \sum_{h\in X^*}(1-\delta)\delta^{|h|}C(\theta,h)D(h,i) \\
&= \sum_{t=0}^{\infty}\sum_{x^t\in X^t}(1-\delta)\delta^t C(\theta,x^t)D(x^t,i) \\
&= \sum_{t=0}^{\infty}(1-\delta)\delta^t Q^t(\theta,i) \\
&= \hat{\mu}(\theta,i)
\end{aligned}$$

where Q is the markov chain on pairs in $\Theta\times M$ as defined in section 3.2. Clearly, from the above equality, we obtain $\nu(\theta|i) = \hat{\mu}(\theta|i)$ for all $i\in M$ reached with positive probability.

Finally, consider part 3. Note that

$$\begin{aligned}
\nu(i|\theta,h) &= \frac{\nu(\theta,h,i)}{\nu(\theta,h)} \\
&= \frac{(1-\delta)\delta^{|h|}C(\theta,h)D(h,i)}{\sum_{j}(1-\delta)\delta^{|h|}C(\theta,h)D(h,j)} \\
&= D(h,i)
\end{aligned}$$

which is independent of θ. The desired stochastic map is given by $\hat{\sigma}(h)(i) := D(h,i)$

Proof of Corollary 5 : The result follows from Blackwell's theorem on comparison of experiments (See Blackwell (1951[4],1953[5]). Using the three properties of the joint distribution derived in the above proposition, we obtain that the memory states garble the sequences of signals. Since we have $\{\pi(.|h)|\ h \in X^* \text{ and } \hat{\sigma}(h)(i) > 0\} \subseteq \Pi$, from Blackwell's theorem, it follows that $\hat{\mu}(.|i)$ is in the convex hull of the set $\{\pi(.|h)|\ h \in X^* \text{ and } \hat{\sigma}(h)(i) > 0\}$. Since Π is convex, it follows that $\hat{\mu}(.|i) \in \Pi$.

7.2 Proofs from Section 3 and Analysis of Example

Proof of Proposition 6 : 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 represent updated beliefs in the following way :

$$\hat{\pi}(\pi, x) = \frac{\pi(1-\epsilon)f_0(x) + (1-\pi)\Delta f_0(x)}{\pi(1-\epsilon)f_0(x) + (1-\pi)\Delta f_0(x) + \pi\epsilon f_1(x) + (1-\pi)(1-\Delta)f_1(x)} \tag{20}$$

$$= \frac{1}{1 + \dfrac{\pi\epsilon f_1(x) + (1-\pi)(1-\Delta)f_1(x)}{\pi(1-\epsilon)f_0(x) + (1-\pi)\Delta f_0(x)}} \tag{21}$$

$$= \frac{1}{1 + \dfrac{f_1(x)}{f_0(x)} \dfrac{\pi\epsilon + (1-\pi)(1-\Delta)}{\pi(1-\epsilon) + (1-\pi)\Delta}} \tag{22}$$

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 :

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

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

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

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 $\bar{\pi}_1 = \frac{\Delta}{\Delta+\epsilon}$. For likelihood of $\bar{x}$, say $l^* > 1$, we have

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

Now, for the given signalling structure $< f_0, f_1 >$ we have that $\max\{\frac{f_0(x)}{f_1(x)} : x \in X\} < l^*$.Order the x's according to likelihood ratios as $\{x_1, .., x_n\}$. The likelihood ratios are distinct and 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 :

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

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 T after which the decision maker should choose the risky action for sure. Now, a deterministic automaton which keeps track of signals upto time T can implement the optimal unconstrained decision rule.

Proof of Proposition 7 : 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. We argue as follows. Firstly, notice that $V(0,k) > V(0,l)$ implies $V(1,l) > V(1,k)$ since otherwise $V(1,k) \geq V(1,l)$ would imply that $\hat{\mu}(0|l)V(0,k) + (1-\hat{\mu}(0|l))V(1,k) > \hat{\mu}(0|l)V(0,l) + (1-\hat{\mu}(0|l))V(1,l)$ violating the revelation principle at l. Secondly, state l is preferred over state k at belief π if and only if $\pi \leq \pi_{l,k} := \dfrac{V(1,l)-V(1,k)}{[V(1,l)-V(1,k)]+[V(0,k)-V(0,l)]}$. By the revelation principle at l this implies $\hat{\mu}(0|l) \leq \pi_{l,k}$. Since $\hat{\mu}(0|k) < \hat{\mu}(0|l)$, the conclusion follows.

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 the above, it follows that $\hat{\pi}(\hat{\mu}(0|i),x) \in [\underline{\pi}_j, \underline{\pi}_{j+1}]$

Analysis of Example : The analysis below pertains to the example in section 3.2. It demonstrates inertia with respect to intermediate signals in the optimal two-state automaton. Further, when multiple clusters of signals exist, a set of parametric conditions similar to those in the hypothesis of Proposition 14 below guarantees strong reaction to a class of low non-extreme signals. The following lemma shall be useful.

Lemma 12 *Fix $0 < \delta < 1$. Consider the set of points $\mathcal{W} = \{((w_H^0, w_H^1),(w_L^0, w_L^1)) \in [E_1, E_0]^2 \times [E_1, E_0]^2 | w_H^0 \geq w_L^0 \, , \, w_L^1 \geq w_H^1\}$, then :*

1. *$\mathcal{W}$ is compact*

2. *$\pi_{\max} := \max_{(w_H, w_L) \in \mathcal{W}} \dfrac{\delta(w_L^1 - w_H^1) + (1-\delta)(-E_1)}{\delta[(w_L^1 - w_H^1) + (w_H^0 - w_L^0)] + (1-\delta)(E_0 - E_1)} < 1$*

3. *$\pi^* < \pi_{\max}$*

Proof : Part 1 of the claim follows from the fact that $\mathcal{W}$ is closed and bounded. We now establish part 2. Since $\mathcal{W}$ is compact, it suffices to show that for all $(w_H, w_L) \in \mathcal{W}$, the condition in part 2 is satisfied. Hence, let $(w_H, w_L) \in \mathcal{W}$. Consider the following 2 by 2 state dependent utility function $u(a,\theta)$:

	0	1
a_0	$(1-\delta)E_0 + \delta w_H^0$	$(1-\delta)E_1 + \delta w_H^1$
a_1	δw_L^0	δw_L^1

By definition of $\mathcal{W}$, we have : $u(a_0, 0) > u(a_1, 0)$ and $u(a_1, 1) > u(a_0, 1)$. Hence, there exists a unique $\mu^* \in (0,1)$ such that $\mu^* u(a_0,0) + (1-\mu^*)u(a_0,1) = \mu^* u(a_1,0) + (1-\mu^*)u(a_1,1)$. Now, note that

$$\mu^* = \frac{\delta(w_L^1 - w_H^1) + (1-\delta)(-E_1)}{\delta[(w_L^1 - w_H^1) + (w_H^0 - w_L^0)] + (1-\delta)(E_0 - E_1)} < 1$$

Now consider part 3. Consider the values $(w_H^0, w_H^1) = (E_0 - c, E_1)$ and $(w_L^0, w_L^1) = (0,0)$ where $c > 0$ is small enough so that $E_0 - c > 0$. For the defined (w^H, w^L) if we construct a state dependent utility function $u(a,\theta)$ as above, then the threshold μ^* corresponding to $u(a,\theta)$ is strictly greater than π^*. From part 2, it now follows that $\pi^* < \pi_{\max}$. ■

The following result derives sufficient conditions under which the optimal two-state automaton exhibits the features highlighted in example. Further, it demonstrates that they indeed will arise for an open set of parameter values.

Lemma 13 *Fix the discount factor $0 < \delta < 1$, the payoffs $E_1 < 0 < E_0$ and the likelihood ratio $l_m < 1$ of the intermediate signal. There exists an open set of paramater values for $(\pi_0, \Delta, \epsilon, f_0, f_1)$ such that :*

1. $\pi_{\max} < \hat{\pi}(\pi_0, m) < \pi_0$
2. $\max\{\bar{\pi}_m, \hat{\pi}(1, \{l, m\})\} < \pi^*$

Here, $\hat{\pi}(1, \{l, m\})$ denotes the belief update of $\pi = 1$ upon knowing that the signal is in the set $\{l, m\}$.

Proof Choose π_0 high enough such that $\pi_{\max} < \dfrac{\pi_0}{\pi_0 + \left(\frac{1}{l_m}\right)\pi_0} < \pi_0$. Now, π_0 is fixed. Next, choose $\Delta, \epsilon > 0$ low enough so that $\bar{\pi}_m < \pi^*$ and $\pi_0 = \dfrac{\Delta}{\Delta + \epsilon}$. The former is possible since it can be guaranteed for low values of $\Delta, \epsilon > 0$ and the latter can be ensured via an appropriate choice of $\Delta, \epsilon > 0$ to make the stationary distribution of P equal to the prior. This implies $P(\pi_0) = \pi_0$ and $\hat{\pi}(\pi_0, m) = \dfrac{\pi_0}{\pi_0 + \left(\frac{1}{l_m}\right)\pi_0}$. Now, $\Delta, \epsilon > 0$ are fixed.

We now wish to choose $< f_0, f_1 >$ such that $\dfrac{f_0(m)}{f_1(m)} = l_m$ and $\hat{\pi}(1, \{l, m\}) < \pi^*$. We proceed as follows. The condition $\hat{\pi}(1, \{l, m\}) < \pi^*$ is equivalent to

$$\frac{1}{1 + \left[\frac{f_1(l) + f_1(m)}{f_0(l) + f_0(m)}\right]\frac{\epsilon}{1 - \epsilon}} < \pi^*$$

which in turn is equivalent to :

$$\frac{f_0(l) + f_0(m)}{f_1(l) + f_1(m)} < l^* \tag{23}$$

for some $l^* > 0$. We construct $< f_0, f_1 >$ as follows. First pick $f_0(l), f_1(l) \in (0,1)$ so that $\dfrac{f_0(l)}{f_1(l)} < \min\{l^*, l_m\}$. Now choose $f_0(m), f_1(m) \in (0,1)$ low enough so that $\dfrac{f_0(l) + f_0(m)}{f_1(l) + f_1(m)} <$

$\min\{l^*, l_m\}$ and $\dfrac{f_0(m)}{f_1(m)} = l_m$. We now have all the desired properties for $< f_0, f_1 >$ namely the $\dfrac{f_0(l)}{f_1(l)} < \dfrac{f_0(m)}{f_1(m)} < 1 < \dfrac{f_0(h)}{f_1(h)}$ and that $\hat{\pi}(1, \{l, m\}) < \pi^*$.

Further note that conditions 1 and 2 in the statement of the lemma all involve strict inequalities. Since $\hat{\pi}$ is continuous, the conditions are satisfied for an open set of parameter values. ■

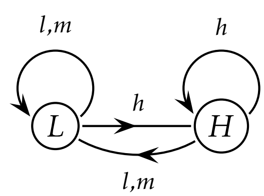
We now establish the desired properties of the optimal two-state automaton in the example.

Proposition 14 *Let $\pi^* < \pi_0 = \frac{\Delta}{\Delta+\epsilon}$ and suppose conditions 1 and 2 stated in Lemma 13 are satisfied. Then, the optimal two-state automaton with memory states $\{H, L\}$ satisfies the following properties :*

1. *$\hat{\mu}(0|L) < \pi^* < \hat{\mu}(0|H)$ i.e the automaton choose the risky action at H and the safe action at L*

2. *$\underline{\pi} < \hat{\pi}(\hat{\mu}(0|H), m)$ where $\underline{\pi}$ denotes the point of division for the two categories of beliefs. This implies that the automaton stays in H upon receiving the intermediate signal m*

Proof We first show part 1. We wish to show that the optimal automaton does not ignore information meaning that the beliefs $\hat{\mu}(0|H), \hat{\mu}(0|L)$ lie on either side of the threshold for action π^*. It suffices to find some automaton which satisfies this property. Consider the following two state automaton :



the intial state is H and the automaton takes the risky (safe) action at H (L). Note that this automaton simply follows the signal i.e whenever it receives a signal $\{l, m\}$, it takes the safe action and risky otherwise. Let $\hat{\mu}^f \in \Delta(\Theta \times M)$ denote the long run distribution over states of the world and memory states for the above automaton. It suffices to show that $\hat{\mu}^f(0|L) < \pi^*$. Now at L, we know that the true sequence of signals ends in $\{l, m\}$. From Corollary 5 , we know that $\hat{\mu}^f(0|L)$ is the expectation over the beliefs induced by all such signal sequences. From Condition 2, it then follows that $\hat{\mu}^f(0|L) < \pi^*$.

Now consider part 2. From part 1, it follows that the optimal automaton chooses the risky action at H and the safe action at L. Now, the point of division $\underline{\pi}$ is such that $\underline{\pi}V(0,H)+(1-\underline{\pi})V(1,H)=\underline{\pi}V(0,L)+(1-\underline{\pi})V(1,L)$. The continuation values $V(\theta,i)$ are as follows :

$$\begin{aligned} V(0,H) &= (1-\delta)E_0+\delta W_H^0 \\ V(1,H) &= (1-\delta)E_1+\delta W_H^1 \\ V(0,L) &= \delta W_L^0 \\ V(1,L) &= \delta W_L^1 \end{aligned}$$

where $W_i^\theta := \sum_{j\in M}[\Pr(0|\theta)\alpha_{ij}^0 V(0,j)+P(1|\theta)\alpha_{ij}^1 V(1,j)]$ is the continuation value obtained at (θ,i) after taking the decision and employing transitions according to $\sigma(i,.)$. Note from modified multiself consistency, at H, given belief $\hat{\mu}(0|H)$, it is optimal to use the transitions $\sigma(H,.)$ rather than $\sigma(L,.)$ (similar argument for L and belief $\hat{\mu}(0|L)$). Since $\hat{\mu}(0|H)>\hat{\mu}(0|L)$, it follows that $W_H^0\geq W_L^0$ and $W_L^1\geq W_H^1$. Hence, we have $(W_H,W_L)\in\mathcal{W}$, where $\mathcal{W}$ is the set defined in Lemma 12. This implies that $\underline{\pi}<\pi_{\max}$. Finally, from condition 1, we obtain $\underline{\pi}<\pi_{\max}<\hat{\pi}(\hat{\mu}(0|H),m)$ since $\pi_0<\hat{\mu}(0|H)$. ■

7.3 Proofs from Section 4

Proof of Proposition 8 : Let τ be an automaton yielding transition probabilities α_{ij}^θ. For any equivalent automaton and memory state $i\in M$, the transition $\hat{\sigma}_i(x,j):=\sigma(i,x)(j)$ must satisfy the following conditions :

$$\begin{aligned} &\sum_x \hat{\sigma}_i(x,j)f_\theta(x)=\alpha_{ij}^\theta \ \forall j, \\ &\sum_j \hat{\sigma}_i(x,j)=1 \ \forall x, \\ &\hat{\sigma}_i(x,j)\geq 0 \ \forall x,j. \end{aligned} \tag{24}$$

Viewing $\hat{\sigma}$ as an element in $\mathbb{R}^{X\times M}$, we observe that $\{\hat{\sigma}:\hat{\sigma}$ satisfies (6) $\}$ is a polytope of the form $\{x\in\mathbb{R}^{X\times M}:Ax\leq b\}$. It is true (see Theorem 2.2, Chapter 2, Schrijver(2013)[27]) that if z is an extreme point of the polytope, then 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 $|\Theta||M|+|X|+|M|\times|X|$ inequalities. Consider an extreme point σ' of the polytope. For this point $Rank(A_{\hat{\sigma}})=|X|\times|M|$. Since the first $|\Theta||M|+|X|$ constraints are all equalities, to satisfy the rank condition atleast $|M| \times |X| - |\Theta||M| - |X| \geq 0$ 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 $n_i^{\tau'} \leq |\Theta||M| + |X|$, hence :

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

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

7.4 Proofs from Section 5

Proof of Proposition 9 : The algorithm above terminates uniquely to say σ^* since at every step of the algorithm, the choice of (x, j) that maximises $\frac{f_0(x)}{f_1(x)} V_j^0$ over $\min\{f_1(x), \beta_j\} - u_n(x, j) > 0$ is unique (note from no wastage, the choice is unique at the first step of the algorithm). At every step, when the optimal choice is (x, j), either the row corresponding to x or the column corresponding to j stays fixed for the rest of the iterations of the algorithm. Now, 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 element. 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 $\hat{\sigma} = \sigma^*$

Proof of Proposition 10 : 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.

Proof of Proposition 11 : We use the fact that the algorithm described in Section 5 terminates to a unique solution that is an extreme point of the feasible set. The argument is similar to the proof of Proposition 8 but with weaker constraints.