The Neuroeconomics of Learning and Information Processing; Applying Markov Decision Process

Sidharta Chatterjee

Andhra University

14. February 2011

The Neuroeconomics of Learning and Information Processing; Applying Markov Decision Process

Sidharta Chatterjee*
Visiting Researcher
School Of Economics, Andhra University
Email: sidharta123@yahoo.com
February 14th, 2011

Abstract

This paper deals with cognitive theories behind agent-based modeling of learning and information processing methodologies. Herein, I undertake a descriptive analysis of how human agents learn to select action and maximize their value function under reinforcement learning model. In doing so, I have considered the spatio-temporal environment under bounded rationality using Markov Decision process modeling to generalize patterns of agent behavior by analyzing the determinants of value functions, and of factors that modify policy-action-induced cognitive abilities. Since detecting patterns are central to the human cognitive skills, this paper aspires at uncovering the entanglements of complex contextual pattern identification by linking contexts with optimal decisions that agents undertake under hypercompetitive market pressure through learning which have however, implicative applications in a wide array of social and macroeconomic domains.

Keywords: Cognitive theory, Reinforcement Learning, Markov Decision Process, Glia, Action potential, policy pattern, Neuroeconomics

JEL Code: A12, C91, D83

Unpublished: Note; Usual Disclaimer Applies

---

* The author extends his gratitude and thanks to Prof. K.Rao, Faculty and Research Head of Economics Unit, School of Economics, Andhra University, India, for without his acquiescence, this research wouldn't have been possible, to all those involved in this research, and particularly, to the library staff of Dr. V.S. Krishna Central Library, Andhra University, for their assistance.

The author encourages comments from readers and collaborators to further augment the cause for quality research in the field of Cognitive Sciences, Neuroeconomics and related subjects.

Contact Information: sidharta123@yahoo.com

All errors are mine only.

I. Introduction

'The brain is the material instrument by which we retrace and combine ideas, and by which we remember, we reason, we invent'-

---- P.M. Roget (1834)

IN THIS PAPER, I attempt to undertake an interdisciplinary approach to the study of problem solving and decision making as a path dependency process which leads to realization of certain reward related value functions in a given system with predefined values. I examine at the cellular level how agents learn to process and manage new information as well, take policy actions that maximize their value functions under spatial dimensions which determines the computational and cognitive limits of agents in a complex process of decision making. Applying a simple Markovian Decision Process (MDP) based-model that generalizes the evolution of agent behavior over time, this study demonstrates the efficiency of the evolutionary process of knowledge generation based on neuroscientific theories using new information and Reinforcement learning technology as of; how agents are able to pertain these under best possible procedures, and under rational environment available to these real economic agents. This work thus, draws upon other works of the authors, in particular, Rizzello (2003), Dosi et.al. (2011), Novarese and Rizzello (2001), Bayer and Reynou (2011) and David (1997). The model that I propose in the present context helps to highlight the problem of choice in decision making when a system is given as it is; with all its initial values and several possible policy actions which when followed by these rational agents, would help to maximize their reward values. As such, it may be defined as a path dependence process, a property of complex dynamic systems where in a specified space of searching and exploration, it is impossible to know all the probable paths, and hence, to foresee all the possible outcomes. In a sense, this may lead to sub-optimal equilibriums. In strategic models of choice under risk where factors that realize the discordances of individual preferences, i.e., incomplete knowledge,

uncertainty and information asymmetry, agents are required to undertake polyhedral thinking in order to maximize choice's outcomes in terms of the functions of probability. It is through incremental innovation in perceptual learning mechanisms that agents develop certain ability as skill-sets, and discover value from certain policy actions while taking cues from past experiences from decisions that have been preceded by certain policy patterns. What I attempt in this regard, is not to develop a new model of decision making process, but apply Markov Decision Process to explain agent-based cognitive peculiarities in decision making process as well, to highlight cross-sectional consistencies in agent behavior (Bem and Allen, 1974) when they follow certain policy actions. That would, I presuppose, help generalize the complex parametrizations by integrating features of both agent-based learning methodology and reinforcement learning under MDP model framework, closely following (Oeffner, 2008) on computational agent-based dynamic equilibrium modeling. This would in such attempt, incorporate the parameters of both the models. The model simplification may be stated as;

$$\Delta C_{its} = \sum_{i}^{j} L_i \ (A_{its} + R_{its} + \varepsilon_k)$$

Where, ΔC_{its} is defined as the Cognitive capacity of agents under equilibrium, *A* and *P* as associative and reinforcement learning model, whereas, '*i*' as attributes of mental states *t* as period for session'*s*'. A_{its} and R_{its} as value taken by a dependent variable while ε_k is the constant of the dummy variable $'\varepsilon'$. This model specification is conceived to represent the cognitive equilibrium state of agent behavior under uncertainty since the finalities of the process remains unknown and where, the model includes all the variables as independent attributes required to model such an equilibrium state. Since there are multitudes of factors that determine decision-making process within economic organizations, this complex form of interdependency and interactions between the agents and their environment involving unbounded number of variables to

represent the scientific concept of abstraction cannot be fully justified. Considering the fact that human beings have the limited capacity to deal with new information where agents have cognitive limitations and where information about future events are not available or foreseeable, this needs to be modeled by means of rational approach and, by using bounded variables. Here, the choice of decision is important to these agents since; they try to maximize the value derived from their actions based on certain policies. As such, generalized abstraction, although appear to play some role while taking policy actions, however, contemplate the primer that scientific theory must be based upon abstraction as a basis for replication of the reality, may be justified only on the ground of such factual discretion where human agents endeavor to develop inimitable capacity to abstract from learning. Yet, this abstraction seems apparent only when agents are exposed to the environment with which they interact and take corrective decisional choices, and invariably, then, the problem of choice inculpates. This problem of choice seems to generate biases and heuristics in decision making. To make things simpler for themselves, agents in general, examine attributions and search for some natural '*equilibrium patterns*' that aims to offset frequency of errors in decisional choice. This is due to the fact that agents are able to identify contexts or patterns, where, patterns indicates order, and which refers to finding elements of unity among different situations or events similar to one experienced in past. This capacity to abstract in terms of perceptual recognition of contexts that help generalization of learned information about the complex world is related to the *associative* theories of storage, retrieval and learning, in addition to, the *reinforcement* model of learning- wherein, agents learn from interactions with their environment through trial-and-error. Since learning is a process that emerges from activity in a subjective and socially constructed world, the issue of embodiment in learning brings upon enduring changes at the aggregate level, which is perhaps, related to the situated nature of human cognition where cognition being a process inside the mind, is affected by mental states representing contextual aspects. Given that cognition is linked to perception and action, there exits definite

interrelationships between perception, action, behavior and goal wherein, action is required for perception in the same tune as goal is necessary for action in retrospect, whilst introspectively, action is required to achieve goals, *if, and only if*, those actions are backed by solid decisions framed on concerted policies that maximizes the utility function of actions. In effect, decision making process within economic organizations has been primarily based on the subjective expected utility theory which states that the decision maker chooses between risky or uncertain prospects by comparing their expected utility values. However, the probability of choice among decisions those agents will take on what decisional trajectory, may be determined by implicit policies, and pre-determined values of such policy outcomes- as goal. *In adeo*, the agent actions are goal oriented; which provides the theoretical framework that sociocultural aspects of cognition and learning have certain intrinsic determinants (as value function defining good in the long-run). Also, rewards tend to determine the immediate inherent desirability of environmental states using value as a secondary variable in predicting and measuring reward of a state. But, as also, due to limited inter-temporal inferential capacity of the human brain that limits the ability to explore large amount of information all at a time (due to complexity of the environment) which however, led to the development by Herbert Simon the concept bounded rationality in 1976 where agents opt to utilize fewer variables (choice of actions) and less information to generalize facts and take efficient decisions by following stabilized rules. This implies the rational selection of variables not in terms of quantity, but in quality or value, since reward must have values as also, without reward, there would be no value! Hence, action choices are based on value judgments. However, value estimation is much difficult than reward determination since rewards are given directly by the environmental states. Reinforcement learning techniques, in these scenarios, can be applied as a method for efficiently estimating values by efficient use of function optimization and search methodologies. However, one complexity still gesticulate some convexities. This is, as usual, the representation of the probabilities of rule following to take solid decisions when patterns of

probabilities are uncertain. Under classical dynamics, this stability of rules as instances of representation and perceptions determine as well guide behavioral processes where rules stay in equilibrium as long as the system is unchanged. I provide an example in the next section on the nature of representation. Thus simply put, in terms of *associative theories of learning*, perceptual recognition of familiar objects or events on account of residual activation of representation where agents understands the relation between the presence and absence of cues or patterns, whether for lexical decision tasks or choice decisions in which, they have the options that allow them to select precisely among variables that increases the quantum of predictability of a system's behavior with a higher degree of probabilistic acuity. This feature is apparently dissimilar from *reinforcement learning* which is goal directed learning from interactions eliciting a complex web of conditional behavior and interlocking goal-subgoal relationships that take advantage of experience to improve performance over time (temporality vector). This is more important in the context of expectation formation since agents often fail to derive rational outcomes under orthodox models where they face real problems while interacting with their environment to achieve goals. Invariably, it calls for adaptive expectations in the course of trial and error through search and reward on account of learning from interaction with the environment. Reinforcement learning explicitly considers the 'whole problem' of a goal directed agent with an uncertain environment that exploits what it knows as also to explore in order to make better action selections in the future (discovering new actions). Elements of reinforcement learning consist of policy, reward function, value function and model of the environment where policy and reward functions are stochastic in nature. The above mentioned two complementary learning models when combined can be expressed as; $C = A_{its} + R_{its}$. These models also allow agents in prediction and decision making.

To quote in such continuum, the word *'Learning'* can be conceptualized as the mechanism by which human beings attempt to realize the 'unknown' and discover the considerable body of *knowledge* hitherto indefinite, and which lies outside their concept *a priori* as predicate foreign to their concept. Here, knowledge may be defined as a condition of access to information defined as a *state of knowing*, and a capability of influencing action, being a path-dependent process (David, 1997, Rizzello, 2004), is a product to comprehend the *reality*, yet not illusion of reality, but explicit reality, where, the causality of conditional and relational aspects are understood in terms of cognitive consensuality (Gioia and Sims, 1986). It is through the acts of logical reasoning, deductive analysis, pure criticism, theoretical discourse and judgmental inferences that the pure essence of empirical universality is established in our faculty of representation by which we are able to differentiate the knowledge *absolutely independent* of all experience *a priori* from that of *a posteriori*- the knowledge gained through experience. However, there arises the definite need for understanding the fundamental mechanism underlying neural basis of learning and mental representation. This is so because of the need to understand how individual mental and neurobiological idiosyncrasies affect decision-making process which accounts for the inclusion of feelings, motivation, and emotions in decision-making processes. In order to provide neural explanation of agent behavior, it is essential to understand the neurobiology of mental representation and control of behavior as expressed as a series of movements and postures controlled by biological neural networks that generate differential patterns where, sensory inputs are analyzed and coordination is generated by the central neurons that precipitate the activation of a motor pattern. In the next section, I will provide a short background review within the scope of this paper, of few historical accomplishments that shaped the domain of evolutionary economics, and perhaps, provided a foreground for its newer sub-domain, cognitive economics which is now one of the most fertile interdisciplinary approach concerned with human learning and behavior.

II. Background Review

The territory and the domains of economic science have expanded ever since the traditional intermingling of social disciplines (Psychology, Sociology, Political Science, etc.) to understand human nature in much broader perspectives. Economics being a normative science is much about social interactions and individual actions that determine material wellbeing of its subjects. This normative study of economic process is now past further than going beyond about market forces, resource allocation and equity in distribution that require both rational decision making while resources are scarce, as well, require human reasoning to foresee and solve problems of allocation and inefficiency. As such, there is genuine need for understanding economic agents' behavior related to conditions of competitive equilibrium where normative and descriptive aspects of decision theory play a greater role in understanding the power of, and the lack of equity in distribution. There arise the necessity and advocacy of positive theories of economics (Friedman, 1953) related to policy-oriented decision making to be based on sound theoretical concepts which would shed light on core fundamental, critical and essential basic public issues (Simon, 1978). In effect, economic models are developed to simulate the real world dynamics and tested as computer based models taking into consideration economic agents' preferences as experimental approach to detect the efficacy of policies that would, in otherwise, be impractical to test on real scenarios, considering the cost and temporal dimensions. In modeling economic scenarios, variables that are determined outside the model-*exogenous*, and those determined inside the model, are termed as *endogenous*. As such, models should consider taking into account optimal variables where they should be determined precisely, and in context. The contribution to the economic science of knowledge thus should be through good methods, where, it is important for the creator of these methods to *perfectly theorize* and refrain from meaningless fact gathering and piling of data end to end. There is genuine need for ordered search for empirical regularities and what should be *avoided* under these circumstances is-to *theorize*

without *knowing* it. When a model is conceived and implemented as a theory, it should be rigorously tested by additional facts to make it more systematic. This is exceedingly imperative since when economics is considered as a scientific discipline, the only true distinctive feature separating economic agents and that of natural sciences is the consistency of laws formulated to define the nature and relationship of matter and energy within some defined, fixed context as laws, which do not change. But as of in contrary to natural sciences, economic agents have differential preferences and choices which often do not follow discrete patterns giving rise to uncertainty. Hence to study the origin and nature of uncertainty and risk and to better understand agent externalities related to the existential generality of interdependencies between mind, matter, and their environment, it is prudent in reinforcing the pillars of cognitive sciences as an extended field of providing its machinery and tools to consider these problems holistically.

The Marshallian thought of parallelism detected between nature and workings of the mind and architectonic dynamics of organization in equilibrium confronted on several aspects of modeling systems that would foresee unexpected outcomes using dynamic equilibrium process, and his writings sought relevance of the mind to analyze organizations (Marshall, 1867-67, 1890). Until that time, contemporary decision making was more allied to the *expected utility theory* which states that the decision maker chooses between risky or uncertain prospects by comparing their expected utility values (Mongin, 1997). The question arises, whether decision makers always rely on probabilities? This pertains to the EUT process which has since been generalized using non-probabilistic decision theories since Allais (1953) invention of a thought-provoking problem widely termed as Allais paradox. Another inference is how to compute a system's expected utility values or payoffs? Is it simply by adding the utility values of outcomes multiplied by their respective probabilities? Or, is it through the cognitive skills of human mind developed from learning processes of individuals that determines individual

agent's payoff? While accounting for the problems of choice in decision making and given the computational and time limits of the internal environments, agents often incur systematic mistakes under a situation of strategic uncertainty. This was defined in Allai's Paradox which violates the theory of expected utility. Then, in case of uncertainty, the question of probability does arise (as choices' outcomes in terms of function of probability) whereas in case of risk, probabilities are not explicitly a part of agent's decision problem (Mongin, 1997). Thus, two standard distinctions of the theory appeared with one-Subjective Expected Utility Theory (SEUT) related to uncertainty, and the other, related to risk theory as the von Neumann-Morgenstern Theory (VNMT). However, these two theories raised questions on the limits of human rationality (uncertainty and risk) which led to the development of Simon's interdisciplinary approach in understanding decision making wherein, Simon's work gathered momentum on economic agents' rationality in decision making process. Although in terms of cooperative games as in 'Nash Equilibrium', repeated game learning matters where learning processes of individuals lead to Pareto efficiency. A Pareto outcome allows no wasted welfare; i.e., the only way one person's welfare can be improved is to lower another person's welfare. This may discretely lead to possibility of predetermining the outcome of repeated games with non-completely foreseeable trajectories (since determining all the possible paths and their value functions would be utterly complex, yet not unfeasible under procedural rationality). As such, what I have attempted here is, to identify some definite patterns of trajectories and hence compute and optimize the total value function (value normalization) of a system given some possible trajectories by considering some amount of probability distributions for uncertain value functions, with some approximations, and simplifications. However, drawing definite trajectories may be easier, but assigning values to them is easier said than done, since, it's these value functions which would likely determine the nature and characteristics of reward, added further, when choices are to be made under uncertainty and risky environment. This is generally described as value function normalization under the Prospect Theory-the theory

proposed by Kahneman and Tversky (1979) to model decision making under risk. In the Prospect Theory, choices among risky prospects are determined by replacing probabilities with decision weights. Applications of the *Prospect Theory,* where, biological and emotional dimensions into decision making are considered that make possible to determine functions of value, which is, by far, one of the mainstays of this theory. Thus, it may seem from the theory that to some extent, uncertainty and risk is itself a pattern and, by recounting the cost of uncertainty and risk is what that predominates by overweighting of low probabilities in both insurance and gambling.

In the realms of evolutionary economics, thus slowly yet steadily marched a few gathering of *avant-garde* economists, most notably, Alchian (1950), Hayek (1952), Carl Menger, Boulding (1956), Allais (1953), Kahneman and Tversky, and Social Psychologists, foremost among them, Herbert Simon, who confounded on the archetypical theories related to systems in equilibrium that aroused much debate on the clinical aspects of economic theory. They diagnosed lack of equilibrium in the equilibrium theory itself, that is, when a system is not in equilibrium, what rationality played on the part of the agents in decision making process? In understanding the structural characteristics and dynamics of organizations with an eye on decision making process invaluable to organization science, these critics raised questions on the limits of human rationality.

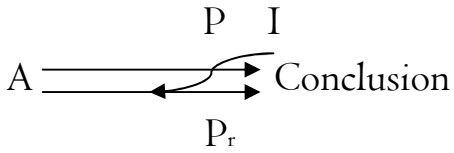
It should be mentioned here that the science of information processing is much based on the art of learning. Specific forms of adaptive learning to augment problem solving efficiency of different forms of hierarchical governance led to building formal models of organizations as information processing and problem solving entities. However, organizations faced some critical challenges while dealing with imperfect information as well; uncertainty related to payoff function of these new decision making models (Marengo et.al. 2000). For example, the market economics of resource allocation related to the

problem-solving activities of a firm confronted on decision choice with respect to its product pricing, cost minimization and profit maximization where firms often face tough decisional choices when they find their profits shrinking, although with rising sales. Here, the crisis originates on account of defining organization behavior in a dynamic decision problem with diversity in decision rules while solving complex decision problems under uncertainty using models for drawing inferences about the nature, and the existence of a number of decision rules present in a system. More complexity props up while defining a system's behavior considering the non-homogeneity in agent learning processes and heterogeneity in cognitive efforts among agents. Several studies noted (Bayer and Renou, 2011, Bracha and Brown, 2010, Houser et. al. 2004), this instability of behavior in agents confronted with complex decision problems and the diversity in the types of decision rules agents use attempting to solve complex problems, often resulting in confusion and bias where, agents end up adopting suboptimal decision rules. These behavioral dispositions led to two arguments; possibly, accounting for the factor of risk and uncertainty related to choices among prospects weighted on probabilities (Kahneman and Tversky, 1979), and the other being, the limits of human rationality, that led to the development of the concept '*Bounded Rationality*'- as the destruens dimension of Herbert Simon's contribution to neoclassical theory. Bounded rationality reduces the amount of variables to manage by using frugal heuristics, i.e., rules which uses just fewer variables and thus increases the capacity to generalize to deal with a complex world. This was not a straightforward as it became practically impossible to model such a system without considering the holistic dimensions of a system's behavior, including that of its agents' preferential diversities under representational spatiality underneath temporal domains. This problem did not go unheeded and well in the mid of twentieth century, Alchian (1950) led the foundation of evolutionary economics that marked the beginning of a new era which today presents a very large gamut of application. However, before this period, there arose the question of existential generality of interdependencies between mind and matter, and that of mentalism and

materialism that brought in the subjective treatments of feelings, emotion and motivations in decision making process, alongside, treating perception, memory and thinking in terms of cognitive psychology. Hayek's (1952) model of mind certainly led to this consideration of transversal subjects in understanding more about the nature and evolution of organizations and institutions in a subjectivist framework of perception, knowledge and cultural evolution. The Neoclassical theory thus proved to be too counterintuitive is answering these notions[1] (Rizzello 2003) where contemporary economists avidly observed the persistence of inconsistencies related to absolute rationality in agent behavior. Also, there arose the question of whether equilibrium theory can still be considered a unifying theory. It is on this otter that I apply Markov decision process to understand the economics of marginal utility under differentiated choices by deriving a mathematical formulation for value function determination.

III. On the Nature of Representation

On the philosophy of contextual representation, agents may be faced with a problem representing its diverse contexts involving decision choices. Yet, all abstraction of these contexts through the act of analytical reasoning leads to two generalized events in natural sciences-the *cause* and its *effect*. Analysis of human analogical reasoning leads to inferences that distinguish arguments from a simple collection of propositions. In the eyes of philosophical thought, an argument inferentially derived from its premises as the truth of its conclusion can be represented as;



Propositions expressed by declarative sentences to describe human reasoning through mental acts of affirmation by judgments connecting truth of one proposition with the truth of another. Here, identifying contexts can make clear

[1]Readers may refer to the series of CESMEP working papers by Rizzello, Novarese and Edigi (2003, 2004, and 2006) for some lucid accounts on the history of economic thought related to the domain of cognitive economics.

the direction of movement by directness of a path trajectory. In understanding the infinitesimal factuality of representation, it is thus essential to understand the structural features of logical arguments as well as methods of representing logical arguments. The general cause and effect principality can however, be represented by following an example that may represent the scenario of multiple contextuality, that is, a problem can solved by a single decision rule encompassing a diverse sub-methodological syntax but the result or goal would have to be the same. Reciprocally, a problem can be represented in diverse contexts to reach the same goal where the units of causes can be represented in various relations with respect to its effect. A simple mathematical example can be postulated; 7+5=12 or, 7+x=12 can also be written as $a + b = c$. Where, the sum of a+b=c and where, a=7, b=5, and c=12. Here, *a* and *b* are both the causes of the effect *c*.[2] However, the real effect 'c' may have a multiple or varied causes, but I content this study with this particular equation to analyze the causes with the goal of attaining the similar effect as for '*c*'. Here, for the given causes *a* and *b* are set as *a priori*, or given, and 'c', the *posteriori*. The system so far is in equilibrium since all the causal variables are known. Each cause, '*a*' and '*b*', may be represented in different contexts as they are made from i.e., each unit of one counts of seven gives 7 for *a*, as well for each unit of one makes 5 for *b*. This embodiment of the importance of *contexts* in real world in the faculty of our representation, where everything is described in nomological context where methods are engineered and designed as closed systems that are bounded by the

[2] To be noted, *a* is *a* and nothing else as well as for *b* and the *c* gives nothing but as stated product of the effects of *a*+*b*. From this representation, it seems that the law of this equation is bounded which is rational for which, there may be a fixed number of contexts to define either '*a*' or '*b*' while, the '*c*' remains unchanged. Herein, if we assign the concept of cause *b* as unknown=X, a requirement $7 + x = 12$ to discover outside the concept of '*a*' a predicate '*b*' foreign to this concept (related to the concept as cause of '*a*') is sought for, and solving for x gives $x = 5$. Now, a different representation of the context as given by 7+ $(x^2 + 1) = 12$ or $7 + (x^2 - 4) = 12$ must be as a subproduct of the representation b=5 and nothing else so as $7 + 5 = 12$. However, similarly, the cause may by represented as well by $(d + 1) + 5 = 12$ *or* $(d - 1) + 5 = 12$ wherein, the product of $(d - 1)or(d + 1)$ must be, 7. While, a causality of $7 + (x + y) = 12$ gives $x = -y + 5$. Now by substituting $7 + (-y + 5) = 12$ yields the value for $-y = 0$ and the rationality of $7 + x = 12$ is established. Similarly, $7 + \left(x + \frac{y}{z}\right) = 12$ would require more methodic iterations to establish the empirical universality of 'b=5'. Thus, the rules remain the same but the iterations and subrmethods may alter to attain the same goal. This is a simple example of multiple representational states of a single causality.

nature of laws. It is by controlling these closed systems that we see laws without interferences through associations between measurable quantities or values. Hence, explaining contexts is a phenomenon of representational generalization where we may define a context (Ballinger, 2008) as *'a contingent concept of conjecture under spatiotemporal circumstances with the presence of these conjectures in multiple factors as multiple chains of causations being path dependent'.* This is an important deduction since; we can homogeneously reduce this definition as a foundational analog of '*patterns*', which are, in essence, path dependent.

IV. The Model

A. Pattern recognition and Neurophysiology of Associative and Reinforcement Learning

One of the primary foundational questions on interdisciplinary area involving agent behavior and cognitive science encompasses diverse areas of other scientific domains, i.e., behavioral psychology, neurophysiology, neuroscience, artificial intelligence, ethology, behavioral economics, and social anthropology. Thus, the domain of cognitive economics is becoming an assorted sphere that incorporates diverse new subjective domains into the economics jargon-evolving as *'Neuroeconomics'*. Neuroeconomics is hence, the study of the "*mechanisms of how the embodied brain interacts with its physical external and biological internal environment to produce economic behavior".* This multidimensional approach enables investigators to undertake inquisitive research by providing some wide array of innovative tools to explore the fundamental questions involving cognitive sciences. Below is drawn a schematic representation of the science of Neuroeconomics and its related origins, which is just for the reference of readers, not an elaborate origin tree however.

<Diagram here>

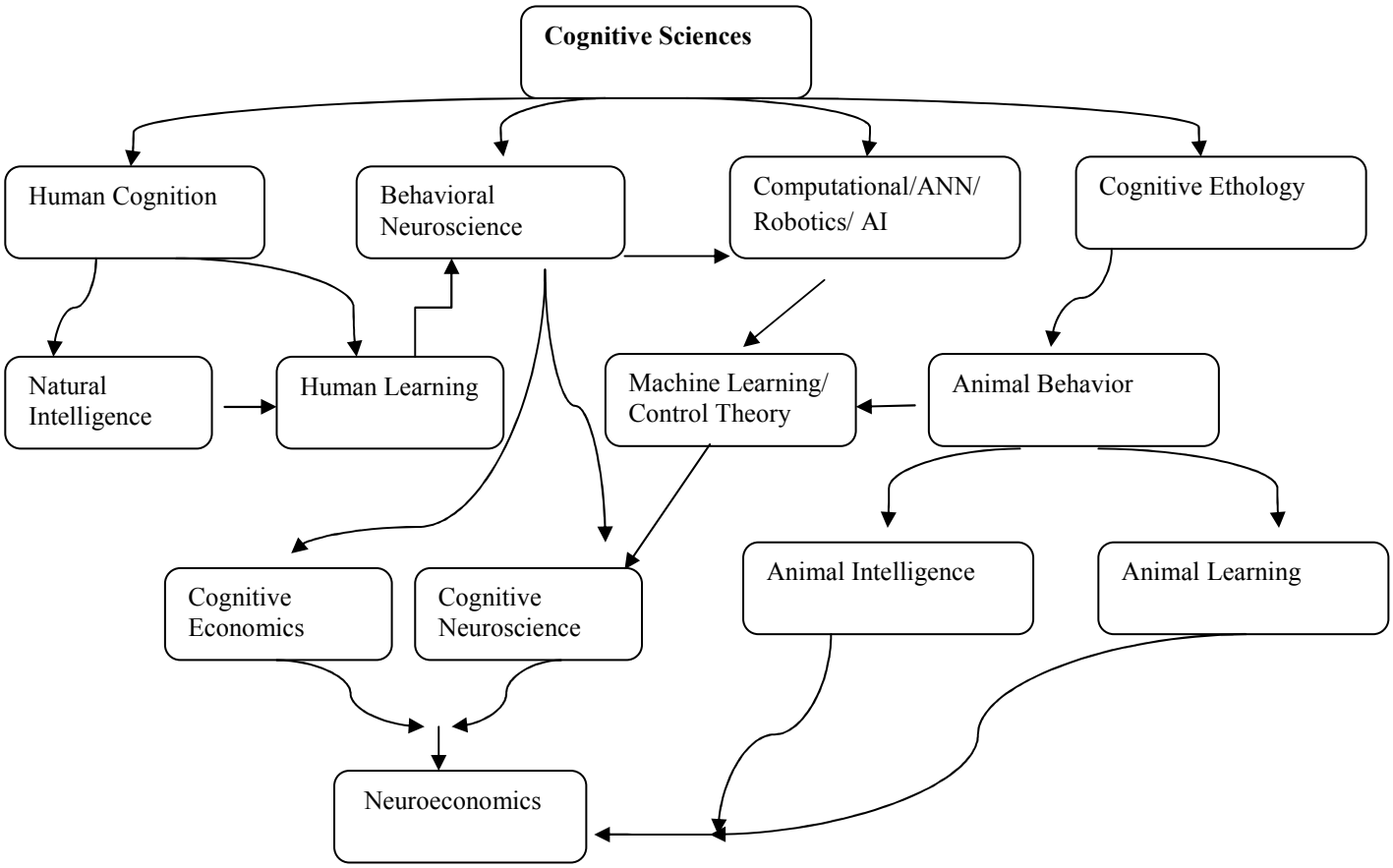


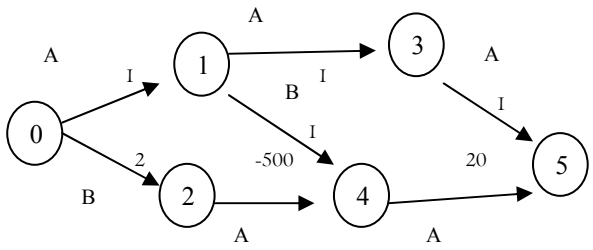
Chart1. Schematic representation of the evolution of Neuroeconomics

The micro-architecture of neural network in the human brain undergoes evolutionary adaptation in response to higher brain functions related to reasoning and pattern recognition and complex cognitive computations (of thinking and conation). We shall see in the next section how '*Glial Cell Theory*' contributes to this fascinating phenomenon of cognitive evolution in hominids. Since the overall number of neuron populations tends to remain same or diminutive with no new neuronal growth, the newly established role of glial cells seem to compensate for the general loss of synaptic connections. In the course of re-establishing new connections among the existing neurons by forming new synapses and interneuron networks, it facilitates nerve impulse conduction and impulse routing among newly available synapses that aid in complex functional efficiency of the brain.

A. Associative Learning and Reinforcement Learning; Agent-based Model

The fundamental theory of *Reinforcement Learning* method is based on the principles of learning to maximize efficiency through trial and error which is selectional in character and where, agents' interactions with the environment improve their performance over time. It is based on learning from interaction. Biological evolution produces organisms with skilled behavior. However, evolutionary methods like supervised learning methods are different from reinforcement methods of learning in a sense that, reinforcement models incorporate planning into the learning system, while supervised systems involved control theory of associations as tasks under conditions of complete knowledge, as opposed to R.L. In evolutionary models, there is not association, but selection methods. It allies more on function optimization and search methods like genetic algorithm and simulated annealing. By sensing the environment, agents choose actions to influence their environment. In supervised method, learning is associative and not by selection. However, John Holland's 'General Theory of Adaptive Systems' is based on selectional principles where, Holland's (1986) classifier systems consists of a true reinforcement learning systems including association and value function. A typical explanation given by Thorndike (1911) based on trial-and-error learning methods in animals proposed the Law of Effect which consisted of selectional and associative aspects of learning. Ron Holland (1960) instituted the policy iteration method for Markovian Decision Process (MDP), a model of Bellman where optimal return function can be computed from a dynamic programming. The components of R.L. method incorporate policy, reward function, value function and the model of the environment. Agents generally observe states and decide on an action. Following actions, they observe the new state and recognize reward thus learning from experience where the process is repeated altogether. A general example involving MDP model of R.L. can be given involving decision choices and discrete actions that is followed by a reward function. This is formally

modeled after Puterman's (1995) MDP model[3]; a simple straightforward model may be stated as;



For MDP, a set of States, actions and reward function is defined by;

1. A set of states '*S*' *as* S=$\{s_1,\ s_2 \ldots . s_n\}$
2. A set of actions '*A*' as A=$\{a_1,\ a_2 \ldots . a_m\}$
3. Reward functions R: S×A×S→$\Re$

There are 3 policies for this MDP;

1. 0 →I →3 →5
2. 0 →I →4 →5
3. 0 →2 →4 →5

Now, based on the policies, reward function can be computed from the above diagram as;

1. 0 →I →3 →5 = I + I + I = 3
2. 0 →I →4 →5 = I + I + 20 = 22
3. 0 →2 →4 →5 = 2 –500 + 20 = -478

I define a more complex model of the above MDP method with multiple decision states and with complex reward functions; (See appendix for function tables). The schema below is a topology diagram of a decision-reward which is however, not all inclusive of back-propagating state-function policies. For the simplicity of the context, I have kept the set of actions and reward function limited which however, may be expanded to a maximum of 37 forward propagating policies.

[3]Readers can refer to a presentation on Reinforcement Learning guide by Bill Smart, 2005 at this address: http://www.cse.wustl.edu/~wds/

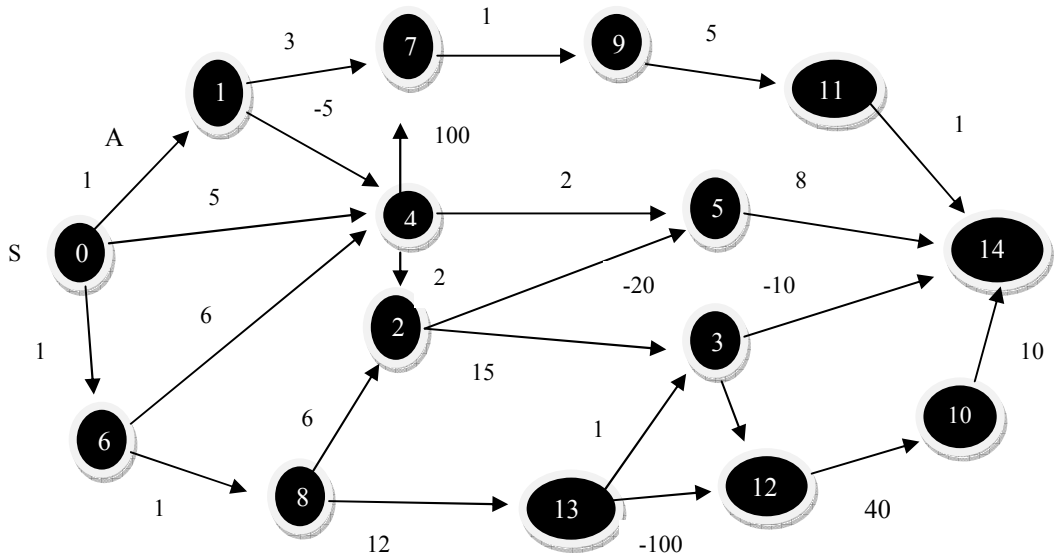


Fig.1 General Topology of a Reinforcement Learning Model with State, Action, Decision and Reward Function

In the above depicted topology model of a decision-reward system, the overall observable states and their value function are givens. Agents are required to decide on trajectory for action and observe reward in a continuous trial-and-error basis until they find the best actions as decision rules that generate maximal reward from the system. As an example of R.L., all actions starts from state '0' defined henceforth as set of states S={S1 + S2…Sn}.This model may in the simplest of form represent a neural network however, without feedback actions (back-propagation). Although the model could be more complex by placing arrows between states which have been omitted, but given the complexity of task, this seems optimal in this context. The function tables (see appendix) provides policy table and their value functions for each action to be followed by agents. The states and values are distributed more or less randomly between the policy trajectories to seek out patterns among outcomes. There are 14 states and 22 possible policies in the model (many more can be conceived). Each of the agents may follow only one policy at a time, and subsequently, by following a particular class of policies through trial-and-error method, agents

learn about the efficiencies of both an individual policy and a class of policies that tends to maximize their reward values, and thus enhance the reward functions. Applying the above reward function variables, I perform statistical computations to find correlations among individual policies (not reported) and their classes as well, obtain the summary statistics on the value and reward functions. The mean and standard deviation of the total reward functions in the system based on 22 policies defined are given in the appendix; I find certain interesting patterns among the states when they are assigned as variables i.e., a, b, d, e, f and arbitrarily, *c* and *g*. The total value of the system when summing up from given values assigned to different states is $V = \sum_{i=1}^{j} V_i(X_{i,1} + X_{2\ldots} X_{i,n})$ is 85, where, X_n is the additional policy values. Following certain policies, agents can maximize their rewards and on three occasions, the policies yield 103, 112 and 114 correspondingly. I define policy pattern classification (PPC) by identifying the symmetrical nature of policy directions that would lead to reward functions. These policies may be grouped and characterized as:

1. (a_1+a_2)=12+74=86
2. $(b_1+b_2+b_3)$=-36+5+67=36
3. $(d_1+d_2+d_3)$=75-4+13=84
4. $(e_1+e_2+e_3)$=-12+3+65=56
5. $(f_1+f_2+f_3)$=-3+13+76=86, and
6. $(g_1+g_2+g_3)$=112+103+114=329

Total Reward Values unlocked by agents (patterned): 892 (or 96.85%)
Let us define the functions in terms of mathematical expressions as;

$$\mathrm{P}(x)=(\alpha_1 + \alpha_2 \ldots x_n) \qquad (1)$$

Then, I may define in terms of definite integral the sum of policy functions as,

$$\int_1^0 x^2(dx) + (\alpha_1 + \alpha_2 \ldots x_n)$$

(2)

Now, all the product of the sum of policy variables is to be defined in terms of positive integer variables by the following formula

$$[(\cos + \sin(x)] = \gamma \tag{3}$$

Where, $$x = \mathrm{P}(x) = (\alpha_1 + \alpha_2 \dots x_n)$$

I derive the values of policy functions from (3) as; 6.0765, 6.0082, 6.4784, 7.7332, 6.0765 and 7.76 respectively. Now let us define the equation for rationalizing the value functions in terms of products. The equation is defined as,

$$\sin \sum_{i=1}^{j} \sum_{1=2}^{j} V_i \tag{4}$$

Where, $V_i = \mathrm{g1} + \mathrm{g2} + \mathrm{g3}$, highest value from all the variables.

By combining (2) and (4), and where, $(\alpha_1 + \alpha_2 \dots + x_n)$ are additional policies, I derive

$$\int_1^0 x^2(dx) + (\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 + \alpha_6) + \sin \sum_{i=1}^{j} \sum_{1=2}^{j} V_i \tag{5}$$

The agents generally have expectations value where, their expectation is a function of probability under linearity of time domain. Under linear time-invariant system, bounded inputs produces a bounded output, as such, I consider that the agents' expectations are rational. I derive the functional part of the equation from Laplace transformation where, the probability density function f is given by,

$$(\mathcal{L}f)(s) = E[e^{-tx}] \tag{5.1}$$

Laplace transformation is used for solving boundary value problems. In this context, it is necessary to derive a function which can determine the conditional probabilities of additional policies in terms of payoff value. In stochastic process such as Markov chains, '*s*' is applied as Laplace transform to move in-between time domains and where, the Laplace variable 's' is the operator variable in the *L* domain to determine the linearity of time domain. This is based on a model proposed by McFarland (1970) to derive mathematical analogy of motivational systems using topological methods of systems analysis. Using inverse Laplace transform, functions of time are transformed into functions of '*S', (f).*A linear dynamical system is given by;

$$f(t) = \mathcal{L}^{-1}\{F(s)\}^{(s^{-1})}$$
$$g(t) = \mathcal{L}^{-1}\{G(s)\}$$

The formula for solving dynamic equations on time scales is given by,

$$Z'\{x|z|\} = \frac{Z\{x[z+1]}{z+1}$$

In equation (14), I define the function in terms of

The optimal value function equation is given by modifying the f as $\int_{e+1}^{e^{-1}} xdx.$ To define the time domain of the probability density function, the agents expectation value should be rationalized, and this may be modified in order to include time-bound variances in the first passage times of Markovian stochastic chain where it is defined by absolute convergence of the Laplace transform in a linear dynamical system, and I modify $\int_{e+1}^{e^{-1}} xdx$ as, $\int_{e-\frac{-\frac{1}{0}}{1}+e}^{e^{-\frac{-\frac{1}{0}}{1}+e}} xdx$. This would

further rationalize probability distribution among the variables. I define the optimization equation as,

$$V = a + \sin(b) + \frac{\sin(b)}{sin(b+b)} \tag{6}$$

Where, $a = \int_1^0 x^2(dx) + (\alpha_1 + \alpha_2 \dots x_n)$, and, $b = \sum_{i=1}^{j} \sum_{1=2}^{j} V_i$ respectively. By substituting the variables in (6), the final optimal value function equation thus stands as,

$$\int_1^0 x^2(dx) + (\alpha_1 + \alpha_2 \dots x_n) + \sin \sum_{i=1}^{j} \sum_{1=2}^{j} V_i + \frac{\sin \sum_{i=1}^{j} \sum_{1=2}^{j} V_i}{\left(\sin \sum_{i=1}^{j} \sum_{1=2}^{j} V_i + \sin \sum_{i=1}^{j} \sum_{1=2}^{j} V_i\right)} \tag{7}$$

Where, the optimality modulator is defined as x^2 in $\int_1^0 x^2(dx) + (\alpha_1 + \alpha_2 \dots x_n)$ of equation (2). Now, again by solving (6) as polynomials, I obtain

$$V = \frac{2\sin b + 2a + 1}{2} \tag{8}$$

Substituting the variables in (8), I derive,

$$\frac{2 * \sin\left(\sum_{i=1}^{j} \sum_{1=2}^{j} V_i\right) + 2\left|\int_1^0 x^2(dx) + (\alpha_1 + \alpha_2 \dots x_n)\right|}{2} + 1 \tag{9.1}$$

$$2 * \sin\left(\sum_{i=1}^{j}\sum_{1=2}^{j} V_i\right) + 2\left|\int_{1}^{0} x^2(dx) + (\alpha_1 + \alpha_2 \dots x_n)\right| + 1 \tag{9.2}$$

From equation (9), we derive the optimal efficiency of agents as reward value gained by each of the agents as 44.66 for the full system, close to our real value of 41. Given by the equation (9.2), in order to compute the percentage probability of reward values to be obtained by the agents by following patterned policy choices after all forward-propagating state policies that can be derived from the system have been computed, we derive a value of 88.33% or conversely, 12% for non-patterned policy functions. To be noted, in our previous computation with real values (See Table 1), the total reward values (as %) gained by the agents while following patterned policy actions was 96.85%. Whilst, to determine the total unrealized value of the system had all the policies been implemented, it is possible to derive approximation by modifying the equation (6) slightly as;

$$V = \sin b + 2a + 1 \tag{10}$$

Where,

$$sin\left(\sum_{i=1}^{j}\sum_{1=2}^{j} V_i\right) + 2\left|\int_{1}^{0} x^2(dx) + (\alpha_1 + \alpha_2 \dots x_n)\right| + 1 \tag{11}$$

By equation (11), we get an efficiency of 78.28% or 721.23, while the primary derivation from equation (7) is 676.23, the lowest possible value that can be derived from the system, and whereby, substituting (10), we get;

$$V = sin\left(\sum_{i=1}^{j}\sum_{1=2}^{j} V_i\right) + \left|\int_{1}^{0} x^2(dx) + (\alpha_1 + \alpha_2 \dots x_n)\right| + \left|\int_{1}^{0} x^2(dx) + (\alpha_1 + \alpha_2 \dots x_n)\right| + 1 \quad (12)$$

the maximum total reward value that the agents can realize in close approximation given the similar state and value function from all possible policy actions(without backward propagation) is 1353.39 from (12), whilst from(11), we derive 721.33 as reward values. Now, the initial reward value computed from all the 22 policy functions is 921. Thus, the maximum possible reward value which agents will be able to derive when every possible state policy is covered, V=721+921=1642, which is, and in approximation, the real value maximization of the system using all network trajectories. This approach is much based on the method of frugal heuristics to determine tautomeric equilibrium of boundary value problem using variations of Laplace functions, where, it may be possible to determine such unknown total value to be realized from a system. A possible application could be in the field of oil and energy sectors, wherein, it might aid in determining the approximate total global or regional reserves of crude oil or gas given some of the known reserve values as policy functions, although, similar methods are usually applied statistically to solve such problems. This method may also aid to develop parallel homeostatic systems related to quantitative predictions about behavior of a system. Thus, I am able to derive the iterated product from (9.2) using a similar, yet modified expression using, the parameter($\prod_{i=1}^{j} V_i$) for optimization vector,

$$\varphi=\frac{\left[1+\left[\prod_{i=1}^{j}V_i+\frac{\prod_{i=1}^{j}V_i+\left(\prod_{i=1}^{j}V_i\right)}{\prod_{i=1}^{j}V_i}\right]\right]^2}{\prod_{i=1}^{j}V_i} \tag{13}$$

Where, φ is defined as optimization parameter as the iterated function for value optimization for a given number of value actions (x_n), that derive as the final optimization equation for this model:

$$\frac{2*sin\left(\sum_{i=1}^{j}V_i\sum_{i=1}^{j}V_i\right)+\sum_{i=1}^{j}V_i*\left(+2*\left|\int_{e+1}^{e^{-1}}x^2(dx)+(\alpha_1+\alpha_2\dots x_n)\right|\right)}{\frac{\left[1+\left[\prod_{i=1}^{j}V_i+\frac{\prod_{i=1}^{j}V_i+\left(\prod_{i=1}^{j}V_i\right)}{\prod_{i=1}^{j}V_i}\right]\right]^2}{\prod_{i=1}^{j}V_i}}$$

$$+m_n \tag{14}$$

Where, $\left(+2*\left|\int_{e+1}^{e^{-1}}x^2(dx)+(\alpha_1+\alpha_2\dots x_n)\right|\right)$ may be modified as $\left(+2*\left|\int_{e-\frac{-\frac{1}{0}}{1}}^{e^{-\frac{-\frac{1}{0}}{1}+e}}x(dx)\right|+(\alpha_1+\alpha_2\dots x_n)\right)$.

Or,

$$2*sin\left(\sum_{i=1}^{j}V_i\sum_{i=1}^{j}V_i\right)+\sum_{i=1}^{j}V_i * \frac{\left(+2*\left|\int_{e-\frac{-\frac{1}{0}}{1}}^{e^{-\frac{-\frac{1}{0}}{1}+e}}x(dx)\right|+(\alpha_1+\alpha_2\dots x_n)\right)}{\prod_{i=1}^{j}V_i}+m_n$$

(15)

That gives value of 28 when $x_n = 87$ or for,$x_n = 50$, the system would optimize the value to 16, which is, somewhat rational given the complexity of the system. Some further optimization is possible when, $\int_{e_{+1}}^{e^{-1}} dx$. The equation (14) that we derive from equation (8), we may call it a general optimization value equation (GOVE) of a linear dynamical system which defines the probability factor by including a variance factor and does not allow for too large deviations in values, although, this is one of its flaws, since, it does not include any randomness in variables. Hence, I denote 'm_n' as the modulator function for vector parametrization that can be modified to the system's requirements as an integer function. From equation (15), it is possible to incorporate some amount of further linear optimization. There remains the question of how it would help identify the context of reinforcement learning model using the MDP method?

- Can it be possible to represent the problem of choice that the agents face when they take actions under uncertain reward functions?
- Can the outcomes of a learning function be determined in terms of agent productivity, or is it possible to know beforehand given a system's total efficiency using Reinforcement learning method?
- Another question to come into my mind: does R.L. enable or help enhance total ability in agents, considered *only when* ability across agents is not identical?

As far as this model specifies, I obtain several optimal values that are of real significance to determine the probability of unrealized state of value functions; both of patterned and unpatterned, suboptimal and optimal rewards as well as of the total rewards to be gained from a similar MDP system when outcomes are unknown. This paper in particular, tries to answer some queries upheld

(Novarese, 2009) in terms of rationality of choosing a specific policy or goal. Here, I consociate upon the uncertainty related to the question of the problem of valuation through selective choice of policy actions using the MDP model to understand what idiosyncrasies does any learning model or agents face, considering the fact that they have the limited capacity to use all the information present in a system, as also, to determine the efficiency of such policies undertaken. In conundrum, and in order to determine the efficiency of a specific goal pursued with predetermined actions and the real values associated with each policy actions, as well as to determine beforehand what would be the total realization in terms of reward or value had been those specific goals were pursued, would be, in terms of generalization, provide some deeper understanding of the behavior of policy actions and their effects on individual (groups) of agents. Agents would have been then, invariably following some definite (un)patterned policy actions (not) knowing their general outcomes in terms of reward realizations. It is of interest to cite that patterned processes are path dependent and selection of a particular path during a critical juncture period is marked by contingencies (Mahoney, 2000). It would become an entirely difficult endeavor to machinate mathematically to prove and establish with absolute precision of what path(s) an agent may follow which is, a temporal uncertainty based on behavioral heterogeneity, and the given finite nature of dimensionality of the finalities of path trajectory, where, the chaos theory seems to reconcile causation with contingencies by linking causally unpredictable outcomes to initial conditions (Ferguson, 1997, Tucker, 1999), which would have, otherwise reconciled to indeterminacy of these patterns.

Based on the value of actions where agents acquire reward while following individual policies, they can be classified according to the policy patterns that they follow, and their behavior distinctiveness may be ascertained, i.e., whether agents are risk averse or risk loving, since under experimental conditions, agents would invariably go for the best patterns by selecting optimal policies that would maximize their reward functions. Some inferential computed values

derived from the rewards may be summarized in table I(See Appendix). As like with any other models, there are best performers and worst performers among agents guided by policy directions. In this model, the agents simply unlocked values of actions by maximizing on what they were given initially and gained rewards while following certain path oriented policies. The model specification of MDP thus established the efficacy of generating reward and gain value using Reinforcement learning method applying Markovian Decision Process (MDP). It is evident that from the model that patterns among policies is one of the prime determinant of value maximization strategies, or in other words, value maximizing policies follow not random order but patterned decision actions. However although, there are major limitations in this model as regard to the probabilities that an agent will choose a particular policy, the risk appetite of the agents, the choices between competing alternatives and the simplicity of the decision rules. I have deliberately omitted backward propagation loops among states that would have generated more number of policies given the values and would have given several new entangled policies, since, all the policies that the system may have is not explored, neither are their reward functions. Hence, I devised mathematical formulations to provide some proximity to the finalities of total value functions the system has, as policy finalities of each agents (Minsk, Farley and Clark, 1954). The efficacy lies in the fact that agents are able to exploit and explore to unlock value potential from a system given decisional choice (Novarese et. al. 2007) as a framework for decision making and, by following certain patterns of policies through trial-and-error or search and reward procedure. The MDP model also determines the optimality or sub-optimality of decision choices and policy actions.

In nature, the dynamical behavior of agents is likely to be influenced by their interaction with the environment (other agents). Having options allow human agents to select precisely among variables (decision rules) which determines the direction of a path trajectory while solving complex problems. However, the technique of thinking that determines human reasoning which helps to ascertain

the '*patterns of thinking*' require establishing the empirical universality of a process or knowledge which is quite complex and bewildering.

V. Discussion

The above model simplifies the value determination problem of a learning system as also, mathematically, proves the computational derivations that were obtained from the 22 policy functions. However, in all total 15 more policy functions may be conceived out of this model making a total of 37 (not reported). The equations' probability determination capability needs particular mention- computing the probability of a complex system's behavior when some of its variables are unknown, which is, in-fact, not in exactitude, but in approximation, relatively comprehensible. When a system becomes too complex and some of its variables remain undefined, it may be represented mathematically to study the distributive patterns of both of its known and unknown variables under uncertainty. This rationalization in terms of behavioral neurobiology of decision and choice modeling simulations for general abstraction in problem solving characterizes cognitive capabilities of human agents who are thus, able to apply a varied array of decision rules and strategies when solving complex decisional problems. This also characterizes the essential features of motivated behavior which is intentional, voluntary, and purposively goal directed where agents have expectancies and incentive factors related to the nature of tasks that they undertake. They develop and apply models for drawing inferences about the nature and complexity of the problems that they face and this creates the existence of diversity in decision rules present in a population. Comparative analogy can be drawn from Southwood (1981) who suggests that it may be due to apostatic selection in maintaining aspect diversity that defines the variations in genetic constituent of agents in defining the theory of the dynamics of biological populations. In similar anthropomorphic vein, cognition plays an essential role in the analysis of motivation and emotion as in the mind's capacity to deal with information, including its reception, storage, processing

and retrieval that require energy. Since agents have certain expectancies, motivational energy is required to activate the internal states to meet those expectancies. This theory is derived from Atkinson-McClelland model of expectancy and value which underlines the nature of cognitive processes involved in achievement motivation. Agents in general, require internal ability or effort on a given task. Since ability across agents is relatively stable, the level of effort fluctuates. In a path-dependency process, success as value or reward stems from ability or the quantum of effort that agents put in. In a learning system such as reinforcement learning, agents uncover the true value of a path or policy through trial-and-error; where, repeated failure leads to success. This stems from the extra efforts that agents are required to take in order to examine attributions of the causes of performance by identifying the set of alternative responses and consequences of each response.

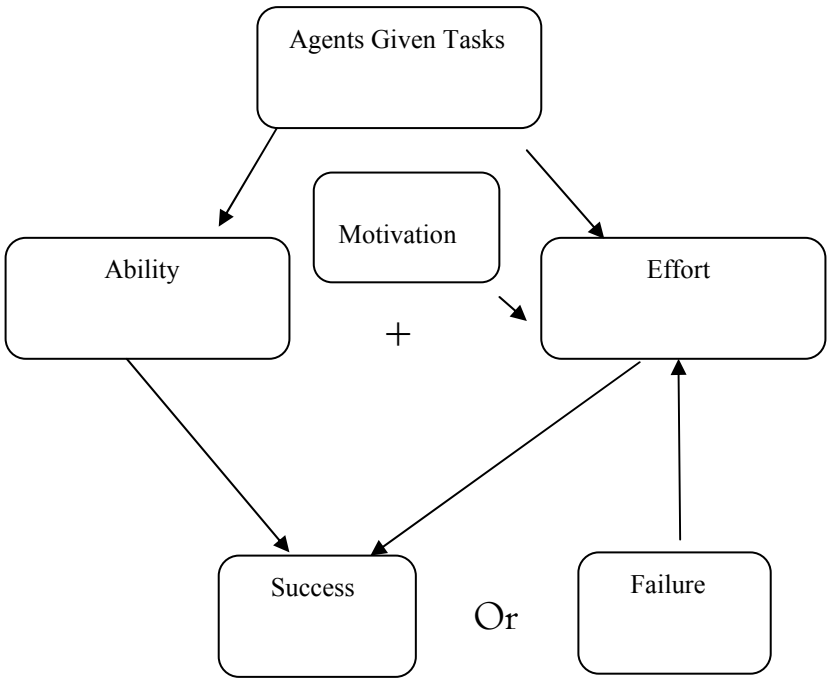


Fig.2The Ability-Effort-Motivation Cycle

Roget (1834), came closer in defining cognition as a materialistic representation of higher brain function way back in the nineteenth century through his following quotation- *'the brain is the material instrument by which*

we retrace and combine ideas, and by which we remember, we reason, we invent'. The above MDP model somewhat mimics neural network of a human brain that throw some light on the cognitive aspects of decision process analogous to computer models based on ANN which is based on pattern recognition, feature extraction, matching and extrapolation. Modern new learning machines are being developed based on the concept of adaptive neural network where machines have the programmed ability to learn new behavioral patterns, receive and compare new stimuli with previously stored information, retain this new information and modify its behavior when situation demands-a feature, that provides some degree of cognitive ability. In the field of biomechanics, robots are being devised to mimic human capabilities using these ANN to achieve neuroeconomic efficiency in operation due to better motor coordination using Reinforcement learning techniques and the theory of backward-propagation. Relatively new contributions to this literature of cognitive neuroeconomics states that under some experimental conditions, memory do not affect outcomes and fails to explain individual performance (Novarese, 2009) while in-fact, it is the conceptual understanding and perceptual representation of states that often alter experimental outcomes. Since brain's storage capacity for all humans are architectonically same, learning through search-trial-and-error based model of R.L. may enhance the brain's network utilization capacity and thus, may increase long-term memory. For some agents, the process of learning is random at the beginning and more stable as time passes by when new information is acquired that helps to develop stable associations between sequences and responses. Functional complexity levels of environmental states and interrelationships among them may induce the brain to utilize more of its unused network by establishing new connecting patterns of neural network as a basis for controlled programming of neural adaptive behavior. This may be important in the context that patients with cognitive disabilities (Beauchamp et.al., 2008) and amnesia induced by neurodegenerative diseases like Parkinson's and Alzheimer's may benefit from *therapeutic learning*- a concept considered to be beneficial in limiting further memory loss, perhaps, by expanding the network transmission

capacity through re-establishing *lost connections* among neuroglial cells in the brain when pharmacologic agents fails to alleviate the pathology.

Neurobiological Implications:

Knowledge about how the brain interacts with its external environment to produce economic and rational behavior will further aid economic decision makers and social scientists to understand the variation in individual decision making, and to make better choices under uncertainty. Here, I would consider some deeper understanding related to the subjective matter of neuroscience, without which, I believe, it will not endure such justification to this new domain of neuroeconomics. The human brain is a complex organic computer composed of 100 billion neurons with 100 trillion connections among them. On average, about 50,000 neurons die or atrophy each day between the ages 20-75, shrinking the size of the brain by 10% by the time one reaches the age of 75. Studies indicate that neurogenesis causes less anxiety as well as enhanced learning and memory formation in adult rodents. The functional domains of human and rat brains are somewhat similar giving the advantage of performing detailed analytic study using fMRI or functional magnetic resonance imaging that simulate much of the higher functions of human brain. There are certain areas in the brain that are associated with the cognitive, reasoning, spatial learning, reward and pleasure functions. The hippocampal region is associated with spatial learning while the dorsolateral prefrontal cortex relates to reasoning, the ventral striatum associated with reward and pleasure, and the orbital frontal cortex being associated with processing of emotions. The association cortex in the human brain is responsible for high level of cognition. Of particular importance is the role of basal ganglia, an old mental structure also found in reptiles associated with special task performance providing the mental strength necessary to reduce the quantum of information processed in the human brain. Human cognitive functions can thus be functionally classified into; complex pattern recognition, reasoning and higher level of problem solving skills.

It is important to establish the factuality of the theory of whether glial cell formation depends on the ability of the human nervous system to recognize pattern. Higher pattern recognition or cognitive functions stimulate glia cell proliferation with the combined influence of neural factors, biochemical, electrical and physiological events which are both exogenous and endogenous. Determinants of new glial cell formations are stimulated by Ca influx, nerve impulse generation, quantum of action potential, adenosine release, and information processing which are modulated by the temporal factors in learning, the nerve cell and synapse populations as well as by memory formation. The neuro-physiological basis of pattern recognition is a function of nerve cell conduction frequency, Ca influx, adenosine release (Guthrie, 1999), the mechanisms which are disrupted in some of the diseases related to demylineating and neurodegenerative amnesic pathogenesis like Multiple Sclerosis, Alzheimer disease (Forman et.al., 2004), Parkinson's and Myasthenia Gravis, or, in mitochondrial and calcium ion transport disorders that severely impair information processing capacity of the brain and leads to nerve conduction disorders. Memory retention and pattern recognition ability of the brain also depend on the population of synapses and neurons which also stimulate glial cell formation, perhaps by some other unknown mechanisms yet fully not understood. Peter T. Lansbury (1992) showed in culture Petri-dishes that excessive build-up of an entangled protein plaque similar to A-Beta molecule carved out of the APP(amyloid-beta precursor protein) are toxic to neurons which interfere with processes critical to learning and memory formation in rats and may thus be the reason behind cognitive decline in patients with Alzheimer's disease. In Parkinson's disease, destruction of neurotransmitter *dopamine* secretion neurons in the substantia-nigra region of the brain causes involuntary tremors and impaired motor coordination and balance which are among the disease's hallmarks (Lang and Lozano, 1998). Using modern scientific tools, researchers are now able to characterize the neurochemistry of

altered mental states whereby, analyzing the causes of abnormal behavior and impaired cognition have become possible.

Higher analytical reasoning function like Pattern recognition capability may also be determined by the constant population of neuron and synapses as well as by the fixed amount of memory and information processing capacity of the brain at any given time. It may be theoretically assumed that higher pattern recognition capacity would lead to more glial cell formation in the brain. In the adult brain, the number of glial cell outnumbers that of neurons by a ratio of 9:1. However, it is not clear since at constant glial cell population, it may be mathematically represented that the synaptic and neuron population remains constant as a function of

$$f(x) = \mathrm{G_c} + \sum_{n=1}^{\infty} \left[\mathrm{a_n(S_y)}x + \mathrm{b_n(O_s)}\right] x,k \qquad (15)$$

Leaning induces and stimulates glial cell proliferation which further aid in controlled programming of neural adaptive behavior. The theory of neural control of behavior was explained vividly by Bently and Konishi (1978) There is a definite relationship between glial cell proliferation and memory formation. Thus, higher cognitive analytical functions like pattern recognition seem to be more simulative and directly related to glial cell formation, but not to new neurons, since, neurons once damaged do not seem to regenerate. However, pattern recognition capability is enhanced due to glial cell development; learning and memory, which have mutually neural, electrical and biochemical properties as well incorporates patho-physiological principles. Then the question remains whether physiological, electrical and biochemical properties induce glial cell formation? What other determinants help better pattern recognition adaptability of human brain?

B. Role of Glial Cells

One primary question is, whether pattern recognition capacity is directly related to the amount of G_c or glial cell population in the brain? Model of higher brain functions reveal the role of glia that influence the formation of synapses and thus help determine strength of neural connections and neural algorithms. Intercommunication mechanism among neurons and glial cells is the theory behind cognitive learning and storing long-term memory. The memory recall process of human brain is similar to retrieving records that match a pattern like a batch file or registry function in computers. Glial cells are typically of two types; astrocytes and oligodendroctyes or Schwann cells. It is presumed that glial cells contribute to information processing in the brain through detection of signaling among glial cells. In the human brain, the glial cells outnumber neurons by a ratio of 9:1. Previously, glial cells were thought to be associated with the maintenance role of bringing nutrients from blood vessels to neurons as also, in preserving the ionic balance in the brain. But glial cells lack the membrane properties required to actually propagate their own action potentials for which, neurons are best suited for. Electrical impulses called action potentials induce neuronal cells to release neurotransmitters (acetylcholine, dopamine, serotonin, 5-HT etc.) across synapses. Earlier work on the hypothesis that calcium influx into the glial cells led to stimulation resulted in the development of a method called calcium imaging to test whether glial cells are sensitive to stimulations (Smith 1990, Kater 1996). Analysis of voltage-sensitive ion channels in glia also reveals that glia cells sense similar electrical signals in axons. However, glia relies on chemical messengers (signals) instead of electrical ones to convey messages. Glial cells usually detect neuronal activity through a variety of receptors on their membranes through which they communicate with neurons and each other. It is also interesting to note that glia influences synapse formations and also alter signals at the synaptic gaps between neurons.

The mechanism of information processing in the brain depends on some wide underlying physiological phenomenon related to nerve cell conduction and impulse transduction of action potentials, neurotransmitter release and synaptic network in the brain. New synapse formation may be related to learning and memory development and need for analytic reasoning capabilities of the human brain which may be directly related to glial cell proliferation while in neural coding, neuro-spatial conduction of nerve impulse threshold of differential frequency do not alter synapse formation. It has been established by (Stevens et.al., 2002) that a neurotransmitter adenosine (from breakdown of ATP) release from astrocytes is one of the factors in new synapse formation as well as in myelination. Then, it is highly probable that certain enzymes stimulate or inhibit new synapse formation by activation of new genes that regulate synapse formation. In some diseases, synapses are destroyed by specific intracellular mechanism by the action of proteases that leads to abnormal nerve impulse conduction syndromes. So, the brain or some intracellular mechanisms determines the optimal level of glial cell requirement for building synaptic network(neural network) required for higher analytical functions related to higher order pattern recognition and reasoning. This function is induced by triggering on specific genes and enzyme activation within the cell nucleus as to determine how many more synapses are to be required for neural coding of analytical functions and forming neural network (glia-axon) with the existing neurons by increasing the number of synaptic junctions. These are analogous to logic gates in computer architecture for information processing of nerve impulse conduction across the, association cortex, thalamus, hippocampus, SI area, PSSC, that greatly increases the ability of neural network for information processing functions. Then, what would be the effect of rapid neuron or glial cell depletion on cognitive efficiency? The answer perhaps lies in some proteins which behave faultily in the human body, or, there may have other causes i.e. stress and environmental factors that induce genetic mutations giving rise to bad proteins. Thus, on the nature of the evolutionary thought of human decision-making integrating computational theories of mind, where, conceptual issues

related to cognitive sciences and the problem of choice can be dealt more interestingly when we learn further about how our brain functions. Perhaps in time to come, advanced technology may develop novel machines and biomedical interventions to deal with the Quantum Brain Hypothesis(Kuljis, 2010) and other interesting topics like digital nootrophins (artificial digital memory enhancers?) to enhance the power of human cognitive dimensions, both in normal and disordered cognition in humans. More research is needed hence in the field of "Molecular Neuroeconomics" for a more interdisciplinary integration toward a coherent understanding of human behavior and human decision-making to solve some unresolved dilemmas.

VI. Conclusion

The general conclusion that can be drawn from this study is the efficiency factor of reinforcement learning, and how agent-based modeling applying MDP method may aid in better decision choice and actions taken by such agents. This study also re-establishes the importance of pattern recognition among policy options and in-efficiency of random actions for reward accumulation. This method may be further enhanced by inducting a better model incorporating unknown variables as values that would help identify specificity in agent behavior as well, decipher the risk aversion and risk appetite of agents under action. Application domains can be expanded to other interdisciplinary fields like predicting the price trend of crude oil as well as reserve capacity accounting when given possible states having diversity of actions and value choices. On this frontier, I have thus undertaken an interdisciplinary approach involving, although, in greater aspects, the "neural" part of economic decision making by reinforcing the pillars of the subjective domain of Neuroeconomics.□

References:

- Alchian, A. (1950) 'Uncertainty, Evolution and Economic Theory', Journal of Political Economy, vol. 58: 211-221.
- Allais, M. (1952), "Le comportament de l'homme rationnel devant le risque: Critique des postulats de l'école Américaine", Econometrica, 21, pp. 503-46.
- Arthur, W. Brian. (2005). Paper prepared for Handbook of Computational Economics, Vol. 2: Agent-Based Computational Economics, K. Judd and L. Tesfatsion, eds, ELSEVIER/North-Holland.
- Arthur, W. Brian., Durlauf, Steven, and Lane, David. (1997). The Economy as an Evolving Complex System II. Introduction to the volume.
- Atkinson, J.W. (1957). 'Motivational determinants of risk-taking behavior'. Psychological Review, 64; 359-372.
- Bayer, C- Ralph, and Renou, Ludovic. (2011). Cognitive abilities and behavior in strategic-form games.University of Leicester, Department of economics, Working Paper no. 11/16.
- Beauchamp M.H., Dagher A., Panisset, M. and Doyon, J. (2008). Behavioural Correlates of Cognitive Skill Learning in Parkinson's Disease, The Open Behavioral Science Journal, 2008, 2, 1-12.
- Bem, D.J., and Allen, A.(1974) On predicting some of the people some of the time. Psychological Review, 81, 506-520.
- Bently, David., Konishi, Masakazu.(1978). Neural control of behavior. *Ann. Rev. Neurosci.* 1:35(59).
- Bracha, Anat, Brown J. Donald; (2010) Affective Decision Making: A Theory of Optimism Bias, Fed. Res. Bank Boston Working Paper no. 10-16.
- Clint Ballinger. 2008. Classifying Contingency in the Social Sciences: Diachronic, Synchronic, and Deterministic Contingency, Cambridge University Paper.
- D.J. McFarland. (1970). Behavioral aspects of homeostasis'. Advances in the study of behavior, vol. 3, 1-26.
- David P. (1997) Path - Dependence and the Quest for Historical Economics: One More Chorus of the Ballad of QWERTY. Discussion Paper in Economic and Social History. Oxford: University of Oxford.
- Dosi, G., Faillo, Marco., Marengo, Luigi., Moschella, D. (2011). Modeling routines and organizational learning. A discussion of the state-of-art. Laboratory of economics and

management, Sant' Anna School of Advanced Studies, Italy, Working Paper Series, 2011/04.

- Egidi M (2002), "Biases in Organizational Behavior", in M Augier and J.J. March (eds), The Economics of Choice, Change and Organization: Essays in Memory of Richard M. Cyert, Aldershot, Edward Elgar.
- Ferguson, Niall. 1997. Virtual history: Towards a 'chaotic' theory of the past. In Virtual history: Alternatives and counterfactuals, Niall Ferguson ed. London: Picador.
- Fields, R.D., Stevens-Graham, B. 2002. New Insights into neuron-glia communication, *Sience:* Vol. 298, 556-562.
- Forman, S. Mark, Trojanowski, Q. John, Lee, M-Y Virginia, 2004. Neurodegenerative Diseases: A decade of discoveries paves the way for therapeutic break-through. Nature Medicine, Vol. 10, 1055-1063.
- Friedman M. (1953), Essays in Positive Economics, The University Press, Chicago.
- Hall, A.D., and R.E. Fagen (1956), "Definitions of a System", in L. von Bertalanffy and A. Rapoport (eds.), General Systems: Volume I. Ann Arbor: University of Michigan Press.
- Hayek, F.A. (1952). The Sensory Order. An Inquiry into the Foundations of Theoretical Psychology, London: Routledge & Kegan Paul.
- Heider, F. (1944). Social perception and phenomenon causality. Psychological Review, 51: 358-373.
- Holland J.H., K. Holyoak. (1986), R. Nisbett, P. Thagard, Induction, MIT Press.
- Houser, Daniel, Keane, Michael, and McCabe Kevin (2004); Behavior in a dynamic decision problem: An analysis of experimental evidence using a bayesian type classification algorithm. Econometrica: Vol. 72, 781-822.
- Kahneman, Daniel and Tversky, Amos. (1979). Prospect theory: An analysis of decision under risk. Econometrica, Vol.47, No. 2, 263-292.
- Kuljiš O. Rodrigo:(2010) Integrative Understanding of Emergent Brain Properties, Quantum Brain Hypotheses, and Connectome Alterations in Dementia are Key Challenges to Conquer Alzheimer's Disease. Front Neurol. 2010; 1: 15. PMCID: PMC3008926
- Lang A.E., Lozano, A.M. 1998. Parkinson's disease, Parts 1and 2. New England Journal of Medicine, Vol.339, 1044-1053 and 1130-1143.

- Mahoney, James. 2000. Path dependence in historical sociology. Theory and Society 29:507-548
- Marengo, L., G. Dosi, P. Legrenzi and C. Pasquali (2000), "The Structure of Problem-Solving Knowledge and the Structure of Organizations", Industrial and Corporate Change, 9, 757-788.
- Marshall, A. (1867-8), Ye Machine, in Raffaelli (1994).
- Marshall, A. (1890), Principles of Economics, London: MacMillan.
- Mongin, Philippe; Expected Utility Theory Handbook of Economic Methodology (J. Davis, W. Hands, and U. Maki, eds. London, Edward Elgar, 1997, p. 342-350).
- Novarese, Marco and Lanteri, Alessandro. 26 April 2007. Individual learning: theory formation, and feedback in a complex task Centre for Cognitive Economics, Universit`a Amedeo Avogadro: *MPRA Paper No.3049*.
- Novarese, Marco, Rizzello, Salvatore. (2004) The Intermingling Between Cognitive Economics and Experimental Economics: a Few Remarks on History, Methodology and Applications. . Dipartimento di Economia "S. Cognetti de Martiis, Centro di Studi sulla Storia e i Metodi dell'Economia Politica "Claudio Napoleoni" (CESMEP)Working paper No. 06.
- Novarese, Marco. Is bounded rationality a capacity, enabling learning? *EERI Research Paper Series* No 12/2009
- Oeffner, Marc:(2008) Agent based Keynesian Macroeconomics-An evolutionary model embedded in an agent-based computer simulation. Inaugural Dissertation.
- Rizzello, Salvatore, Egidi, Massimo. (2003). Cognitive Economics: Foundations and Historical Evolution. Dipartimento di Economia "S. Cognetti de Martiis, Centro di Studi sulla Storia e i Metodi dell'Economia Politica "Claudio Napoleoni" (CESMEP) Working paper No. 04.
- Rizzello, Salvatore. (2003). Towards a cognitive evolutionary economics. Dipartimento di Economia "S. Cognetti de Martiis, Centro di Studi sulla Storia e i Metodi dell'Economia Politica "Claudio Napoleoni" (CESMEP) Working paper No. 03.
- Rizzello, Salvatore. 2004. Knowledge as Path-dependence Process, *Journal of Bioeconomics*, 6, 255 – 274.
- Roget, P.M. (1834)-"Animal and vegetable psychology considered with refernce to natural theology," 2 vols., Pickering, London.

- Simon A. Herbert. (1978). Rational decision-making in business organizations. Nobel Memorial Lecture, 8 December, 1978. Carnegie-Mellon University, Pittsburgh, Pennsylvania, USA.
- Simon A. Herbert; (2000), Bounded Rationality in Social Science: Today and Tomorrow, Mind & Society, Vol 1, n. 1, 25-40
- Simon, A. Herbert (1976). "From Substantive to Procedural Rationality", in: S. Latsis (ed.) Method and Appraisal in Economics. Cambridge: Cambridge University Press. pp.129-148.
- Simon, A. Herbert. (1999). The many shapes of knowledge. Revue d'économie industrielle. Vol. 88. 2e trimestre 1999. pp. 23-39.
- Stevens, B., Porta S., Haak, L.L., Gallo V., Fields, R.D. 2002. Adenosine: A neuron-glial transmitter promoting myelination in the CNS in response to action potential. Neuron, Vol. 36, No.5, 855-868.
- Sul Jai-Yoon, Orosz George, Givens S. Richard, Haydon G. Philip. 2004. Astrocyte connectivity in the hippocampus. Neuron Glia Biology, Vol.1, 3-11.
- Thorndike, Edward Lee (1911), *Animal Intelligence: Experimental Studies*, New York, Macmillan.

Appendix

The Reward Functions:

1. 0 →1 →7 →9 →11 →14
2. 0 →6→8 →13 →12 →10 →14
3. 0 →4 →5 →14
4. 0 →1 →4 →5 →14
5. 0 →6 →4 →5 →14
6. 0 →6 →8 →2 →3 →12 →10 →14
7. 0 →6 →8 →2 →5 →14
8. 0 →6 →8 →13 →3 →14
9. 0 →6 →4 →7 →9 →11 →14
10. 0 →4 →7 →9 →11 →14
11. 0 →6 →8 →2 →3 →14
12. 0 →1 →4 →7 →9 →11 →14
13. 0 →6 →8 →13 →3 →12 →10 →14
14. 0 →4 →2 →3 →14
15. 0 →4 →2 →5 →14
16. 0 →4 →2 →3 →12 →10 →14
17. 0 →1 →4 →2 →5 →14
18. 0 →1 →4 →2 →3 →14
19. 0 →1 →4 →2 →3 →12 →10 →14
20. 0 →6 →4 →2 →5 →14
21. 0 →6 →4 →2 →3 →14
22. 0 →6 →4 →2 →3 →12 →10 →14

Path trajectory of decisions and reward function based on policies:

1. 0 →1 →7 →9 →11 →14=1+3+1+5+1=11
2. 0 →6→8 →13 →12 →10 →14=1+1+12-100+40+10=-36(b_1)
3. 0 →4 →5 →14=5+2+8=15
4. 0 →1 →4 →5 →14 =1-5+2+8=6(c)
5. 0 →6 →4 →5 →14=1+6+2+8=17
6. 0 →6 →8 →2 →3 →12 →10 →14=1+1+6+15+2+40+10=75(d_1)
7. 0 →6 →8 →2 →5 →14=1+1+6-20+8=-4 (d_2)
8. 0 →6 →8 →13 →3 →14=1+1+12+1-10=5(b_2)
9. 0 →6 →4 →7 →9 →11 →14=1+6+100+1+5+1=114
10. 0 →4 →7 →9 →11 →14=5+100+1+5+1=112
11. 0 →6 →8 →2 →3 →14=1+1+6+15-10=13(d_3)
12. 0 →1 →4 →7 →9 →11 →14=1-5+100+1+5+1=103(c_1)
13. 0 →6 →8 →13 →3 →12 →10 →14=1+1+12+1+2+40+10=67(b_3)
14. 0 →4 →2 →3 →14=5+2+15-10= 12(a_1)
15. 0 →4 →2 →5 →14=5+2-20+8=-5
16. 0 →4 →2 →3 →12 →10 →14= 5+2+15+2+40+10=74(a_2)
17. 0 →1 →4 →2 →5 →14=1-5+2-20+8=-12(e_1)
18. 0 →1 →4 →2 →3 →14=1-5+2+15-10=3(e_2)
19. 0 →1 →4 →2 →3 →12 →10 →14=1-5+2+15+2+40+10=65(e_3)
20. 0 →6 →4 →2 →5 →14=1+6+2-20+8=-3(f_1)
21. 0 →6 →4 →2 →3 →14=1+6+2+15-10=13(f_2)
22. 0 →6 →4 →2 →3 →12 →10 →14=1+6+2+15+2+40+10=76(f_3)

Table1. The Value Function Table of Policy Actions

Total Initial Value of the system before the program:	*85*
Total value unlocked or rewards derived by the agents:	*921*
Total No. of Policies:	*22*
Mean (avg.) Reward per agent:	*41.86*
Efficiency Rate:	*49%*
Reward gained by following PP's: (excluding. Top 3)	*563*
Reward Value of Top three Policies:	*35.72 (329)*
Reward Value of Top three Policies as a %:	*35.72%*
Reward Value gained by top 14 patterns per agent:	*40.21*
Efficiency of patterned policy choices:	*61.12%*
Mean reward gained by following random policies:	*6.25*
Total rewards gained following non-patterned policies as a percentage:	*4.3%*
Total Reward Value gained by following PP's as a %:	*96.85*

Text Box.1 Overall Efficiency from MDP model

FIELD (Policy Groups)	N	MEAN	STD	SEM	MIN	MAX	SUM
A (a_1+a_2)	2	43.00	43.84	31.00	12	74	86
B (b_1+b_2+b_3)	3	12.00	51.86	29.94	-36	67	36
D (d_1+d_2+d_3)	3	28.00	41.58	24.01	-4	75	84
E (e_1+e_2+e_3)	3	18.67	40.82	23.57	-12	65	56
F (f_1+f_2+f_3)	3	28.67	41.77	24.11	-3	76	86
G (g_1+g_2+g_3)	3	109.50	6.36	4.50	103	114	329

Textbox 2. Summary Statistics of policy groups and reward value functions

Some useful web-guides related to Reinforcement Learning.:

1. Reinforcement Learning Repository at http://www-anw.cs.umass.edu/rlr
2. University of Alberta on the history of Reinforcement learning :http://webdocs.cs.ualberta.ca/~sutton/book/ebook/node1.html