On Markov Chains, Attractors, and Neural Nets

Diego R. Roqué*
D. R. Roqué & Associates,
5464 SW 142 PL,
Miami, Florida 33175

Time varying Markov chains are considered whose higher order level equilibrium properties are exploited in the context of point, cyclic, and strange attractors to demonstrate the plausible mechanics of how networks of neurons may register signatures and/or recognizable patterns in the human brain as a precursor to the development of intentional behavior.

1. Introduction

The work presented here relates specific aspects of the theory of Markov chains with some concepts that arise in the theory of complexity. This is done in order to support and help explain more recent hypotheses on how networks of neurons process information. The reader is referred to [1] where a good treatment of the subject of Markov chains is offered, and to [2] where a good introductory review on complexity is given.

In [3,4], a closed form solution was given to some simple time varying Markov chains defined on a binary state space. These stochastic processes were characterized by their convergence to a unique finite stationary cycle of probability distributions which is independent of time and independent of initial conditions. They were also shown to exhibit weak ergodicity in their distribution functions. Any one realization, except for the repetitive nature of the cycle, ultimately becomes independent of time and independent of initial conditions while yielding all relevant information about the long run behavior of the process.

In the projected extensions to this work, a particular type of time varying Markov chain was outlined that had the following interesting property: If one considered entropy in its probability context, then the processes could start with the highest levels of entropy possible as initial conditions (in the state probability distribution) and end up at cyclic stationarity with absolute zero entropy. An appeal was made to the work in [5] which was essential to show this. The hypothesis was then suggested that it is possible for neural pathways to process information in this manner. It is the purpose of this work, then, to expand on that notion.

*Electronic mail address: drroque@aol.com.

In the monumental piece of scholarly work presented in [6], the old paradigm that had served as the guidelines for philosophical development in the arena of action theory is demolished. In doing so, a suitable replacement model is constructed by treating the mental processes associated to human intentional behavior as a complex system. The development uses every possible bit of experimental evidence obtained up to our current time to establish the theoretical framework on firm footing. The work presented here is intended to refine some aspects of this theory as it could be made to fit in some corners of this extensive and in itself complex development.

The mathematical models presented here may not be an exact perfect match to the biological basis they intend to portray but they may uncover some principles that govern the manner in which networks of neurons represent (or even acquire) knowledge and information. The foundations for a theory of cognition could very well result from this development. These models are primarily used to establish an association between the theories of Markov chains and complex systems with the objective of expanding the theoretical framework that explains how neural pathways and networks of neurons process information.

One can borrow from military science to establish three levels at which one can look and explore the human brain. First is the strategic time horizon or scale that corresponds to millions of years of evolution. In this period one has a trial and error learning and adaptive process where paths develop and mature in response to both external and internal stimuli. This would be analogous to sunflowers evolving and developing their tropism. Second is the operational time horizon or scale corresponding to the time that it takes a human brain, more or less, to carry out activities such as completing a doctoral dissertation or proving a very difficult theorem. At this level and the next, one can look for autocatalytic and self organizing processes in the brain such as the ones that are incorporated into the new paradigm [6] as well as other processes of gradual development in response to stimuli such as those exhibited in post-natal development of the human brain [7]. In a sense, it could be said that at this level or scale the human brain has the capacity to mimic evolution itself but on a much shorter time span. Finally, there is the tactical time horizon that corresponds to the scale at which one observes the neurons actually firing along their pathways. In this paper attention is devoted primarily to the third level and some of the repercussions or consequences at the second level.

Emphasis is given to three important characteristics in terms of how one looks at the brain. First, traditional brain research has assumed that information inside the brain flows from a source (or sources) to a sink (or sinks). Information is viewed as a commodity that flows through the neural pathways. Here instead the view is taken that information is encoded into the paths from the sources to the sinks. It is the paths that contain and constitute the information. But this view does not preclude the possibility that in fact there may be some transmission, from one point to another inside the brain and in the nervous system in general, of information and instructions and other communication elements. It is generally accepted that this is indeed the case. It is also possible to envision the possibility of transmissions with a dual purpose or mission. These would be to constitute the paths as information as well as to carry or convey information or instructions to some other part of the brain. There is some evidence to support this view which is currently being interpreted as sparse coding and neural correlates in the brain [8].

A second concept is that of attractors which play such an important part in the modern theory of complexity. In the context of systems, according to [6]: "All attractors represent characteristic behaviors or states that tend to draw the system towards themselves. . . " Three types of attractors have been identified and they are used here as mathematical modeling tools. They are point, cycle, and strange attractors. The cited author makes extensive use of these and cites many other scientists that have studied and theorized about brain processes. These other authors have used this concept in contexts that are supportive of the overall development. One particularly illustrative quote comes from [9] where connectionist models of linguistic performance are constructed with results supporting the claim that artificial neural networks construct semantic attractors. According to [6], the work in [9] proposes that lexicons be viewed "as consisting of regions of state space within that system; the grammar consists of the dynamics (attractors and repellers) which constrain movement in that space. [This] approach entails representations that are highly context-sensitive, continuously varied and probabilistic (but, of course, 0.0 and 1.0 are also probabilistic), and in which the objects of mental representation are better thought of as trajectories through mental space rather than things constructed."

Entropy in its probability-information context [10] as defined in [11] and [12] provides the basis for the third concept. It is the view of the human brain as an entropy churning mechanism or device. Processes such as cognition, learning, and others could be interpreted in the context of a mechanism that reduces variety and multiple possibilities (very high entropy) to one well defined and established realization (zero entropy). An analogy that comes to mind is the game of Master Mind in either the plastic or computer version. Initially the player is faced with a panel hiding from view four differently colored pins selected at random from a pool of pins in six different colors. The objective for the player is to match the four pins in both color and position. The hidden setting could correspond to some external or even internal stimulus to the brain. The initial entropy is at its maximum (all permutations are equally likely). The player starts the first step by placing in front of the panel an arbitrarily selected set of four colors. The player is then given information on the number of color and position matches and on the number of only color (no position) matches. The player then proceeds to use the information gained and continues for several iterations more at each step receiving the same kind of feedback until a perfect match in color and position is found. At this point the entropy has been reduced to zero. Assuming a player that has developed a consistent set of repeatable optimal search rules, then the exact same stimulus or input will always produce the same search path or trajectory. In the context of the brain, the feedback could be transformed from a function of identifying to one of cataloging and identifying. In this game, the path carries at each step or iteration history and memory that defines the trajectory. In other words, the game is not markovian in character. Writing a computer program to play this game should prove an interesting exercise since part of the objective is to arrive at zero entropy in the minimum possible number of steps. One could conceive of a state space for this game that would incorporate to the permutation at each step all feedback information of the trajectory with their corresponding previous permutations as the states at each iteration or step along the way. This is essentially what the player sees as the game carries its history on its back. This leads to a much larger state space and the game is then markovian in nature but with origin-constrained trajectories.

A mathematical model is developed in section 3 that permits the incorporation and analytical verification of the characteristics and properties that are described in this introduction and are ascribed to the brain processes. The model and its variants consist of finite time varying (nonhomogeneous) Markov chains [5]. The requirements of such a model are a well defined discrete and finite set of states that the process or system may visit, a discrete parameter space indexing the epochs at which the system transitions from one state to the next so that the state visited or occupied at each epoch is a random variable defined on the state space with a well defined probability distribution, a well defined set of transition probabilities for all epochs in time, and the process meeting the criterion of satisfying the Markov property. The latter is a property that simply states that whatever state the process will move into depends only on where the process currently is and not on any prior history. The time index or parameter space is characterized as discrete in its most natural occurrence but it is also possible to consider discrete transition epochs embedded in a time continuum so long as there is a well defined and behaved sequence of transitions in this continuum. In the case of a time continuum, a natural extension is to consider continuous time Markov chain models or to consider their generalization to semi-Markov models that still utilize an embedded Markov chain

[13]. But if the time spent in each state is deterministic and always the same for all transitions in the time continuum, then the results for the embedded process (thought of as in the relative frequency domain or proportion of transitions) would coincide or have immediate identical interpretation to the continuous time process (thought of as in the time domain or proportion of time). Notwithstanding commentary for clarification, extensions and generalizations of this kind are not considered necessary here.

2. Biological basis of model

It would appear that in the evolution of scientific knowledge theory comes first followed by experiment to gather evidence in support or disapproval. In reality a symbiotic process develops whereby experiment feeds of new theory and new theory feeds of experiment. Such is the current state of affairs in brain research. A lot of research has been conducted during the past 25 years as the brain has gradually begun to yield its secrets.

The problem of understanding consciousness or awareness, in particular, has fascinated scientists and philosophers alike. A good account of this is given in [14]. Theoretical speculation abounds but some ideas seem to carry more weight than others. The work in [14] cites the opinion of Francis Crick and Christof Koch (in *Seminars in Neuroscience*, 1990) that "only by examining neurons and the interactions between them could scientists accumulate the kind of empirical, unambiguous knowledge that is required to create truly scientific models of consciousness, models analogous to those that explain transmission of genetic information by means of DNA." Another idea cited is that of Gerald Edelman who "contends that our sense of awareness stems from a process he calls neural darwinism, in which groups of neurons compete with one another to create an effective representation of the world." Then there are Penrose at Oxford who proposes "that the mysteries of the mind must be related to the mysteries of quantum mechanics," and Rasmussen at the Santa Fe Institute who suggests "that the mind may be an 'emergent,' that is unpredictable and irreducible, property of the brain's complex behavior." And there are many others. But following Crick and Koch and more advanced technology and methods, many scientists are beginning to gather substantial empirical evidence on the workings of the human brain.

The work in [8] gives a clear explanation of how neurons communicate. A very brief initial summary states that "a neuron that has been excited conveys information to other neurons by generating impulses known as action potentials. These signals propagate like waves down the length of the cell's single axon and are converted to chemical signals at synapses, the contact points between neurons." A more detailed explanation follows that indicates there is a momentary reversal of a membrane potential that is instrumental in generating the transfer of the pulse and excitation from one cell to the next at a synapse. The epochs at which the next cell (or group of cells) qualifies (or qualify) as "now being excited" are likely candidates for a discrete parameter space whether in a natural scale or embedded in a time continuum. Certainly a synchronization of neuronal firings would be a requirement and insofar as the Markov property is concerned, clearly the next level of excitation would depend only on the current one.

As early as the 1960s, the *Van Nostrand Scientific Encyclopedia* [15] reported a hypothesis on the workings of the brain, albeit simplified, that already contained some rudiments of the model being established here. According to the encyclopedia: "It is postulated that when a sensory impression reaches the brain, it stimulates a nerve cell which, in turn, stimulates another cell. A third cell is then stimulated and so on until a circle has been completed, and the last cell restimulates the first one. The circuit continues to reverberate, thus retaining the impression which set it off. It is further postulated that these reverberating circuits hold the impressions so that they can be recalled later, or compared with other impressions. It is believed that a cell may participate in more than one circuit, thus accounting for various associations of sensory and muscular activity." The "impression" in a closed circular path already suggests information content in the path. The "reverberation" suggests pulses, waves, cycling, oscillation, and synchronization. But an open path from source to sink may pulsate and/or have information content as well. A mesh of paths, open or closed, may act accordingly as well. There may be some elaborate architecture and circuitry and then, there are different kinds of neurons, functions, and specializations.

Crick and Koch (in [8]), studying the problem of consciousness, have addressed the role in visual awareness of a 40 cycle per second oscillation in firing rate that is observed throughout the cortex. The same reference reports that the oscillations were discovered by Wolf J. Singer and his colleagues at the Max Planck Institute for brain research in Frankfurt. It also states that the oscillations may synchronize the firing of neurons that respond to different components of a perceptual scene and hence may be a direct neural correlate of awareness. One may add that if this is the case then the overall conglomerate of neurons, with different sectors responding to different components at somewhat different frequencies, should exhibit neither perfect total synchronization nor total lack of correlation of the rates of firing. Enough correlation may be there, however, to indicate the neurons are broadly tuned.

In [8] it is also reported that there are neurons in monkey visual systems that respond to faces but not to other visual stimuli. Recent experiments show face cells to be broadly tuned, responding to faces with similar features rather than to one face alone. The number of neurons that must be activated before recognition emerges is not known, but the data are consistent with a sparse coding rather than global or diffuse activation. It appears that individual motor cortex neurons are broadly tuned as well. Experiments indicate that the vector obtained by summing the firing frequencies of many neurons is better correlated with the direction of movement than is the activity of any individual cell.

[7] on the other hand, reports on path architecture, offering the following quote: "Visual pathways in the adult demonstrate the segregation of axons. Neighboring retinal ganglion cells in each eye send their axons to neighboring neurons in the lateral geniculate nucleus. Similarly, the neurons of the geniculate nucleus map their axons onto the visual cortex. The system forms a topographically ordered pattern that in part accounts for such characteristics as binocular vision (Dana Burns-Pizer)." This suggests that the definition of the state space along the transitioning (step-wise) paths may very well be architectural or topographical in nature, even down to the level of which and/or how many synapses and branches of dendrites are involved in the conduction of the impulses. The chemical exchanges at the synapses may very well be involved in this process as well.

[7] also reports that Corey Goodman at Berkeley and Thomas Jessel of Columbia University have demonstrated that in most instances, axons immediately recognize and grow along the correct pathway and select the correct target in a highly precise manner: "A kind of 'molecular sensing' is thought to guide growing axons. The axons have specialized tips, called growth cones, that can recognize the proper pathways. They do so by sensing a variety of specific molecules laid out on the surface of, or even released from, cells located along the pathway. The target itself may also release the necessary molecular cues." The distinction is made, however, between target selection and selection of address within a target. There appears to be variation and adjustment in the latter which is currently being interpreted as error and error correction as part of the development and maturing process of the brain. It is not quite clear what the context of "error" is; whether it is visual or architectural or some other kind of deficiency. The following clarifying remark is offered in [7]: "The eventual emergence of discretely functioning neural domains (such as the layers and ocular dominance columns) indicates that axons do manage to correct their mistakes during address selection. The selection process itself depends on the branching pattern of individual axons." It is very likely that pathway selection, branching patterns, and chemical and molecular exchanges are all involved in the definition of the state space and the transition probabilities throughout the paths that in fact help set up the self organizing state attractors discussed in [6].

3. Markov chain models

Consider the following transition probability matrix for an ergodic Markov chain defined on the state space $E = \{0, 1\}$. Let

$$P = \begin{pmatrix} 1 & 0 \\ x & x \end{pmatrix},$$

where x represents any nonzero entry in the matrix.

The matrix P is regular and has steady-state or stationary solution given by the vector $\pi = (1, 0)$. If one considers the transformation $\pi_{k+1} = \pi_k P$, then π is a fixed point of this transformation [16]. Of course, all steady-state solutions of Markov chains are fixed points of their respective transition matrices viewed as transformations. In these cases, all orbits will lead towards the stationary solutions. Ordinarily, however, the latter are usually probability distributions that at most indicate the probability of finding the system in any particular state. Only in a case like the one given by the matrix P above, where one has one absorption state, can one say that the fixed point is also a point attractor. Whatever trajectory the system takes, it is invariably and unavoidably absorbed into the single state from which it can never leave. A fixed point attractor then acquires the characteristics of a state point attractor as well.

In a similar manner, time varying Markov chains give rise to state cyclic attractors. Consider the pair of alternating transition probability matrices given by

$$\cdots \begin{pmatrix} 0 & 1 \\ x & x \end{pmatrix} \begin{pmatrix} x & x \\ 1 & 0 \end{pmatrix} \cdots.$$

As can be verified from the closed form solution, the stationary cycle for this process is given by the vector pair $(1, 0)$ and $(0, 1)$ which alternate with one another. The resulting state frequency vector is $(1/2, 1/2)$ at maximum frequency entropy. This deterministic oscillation between two states was discussed in [4] where the author described its higher order level of equilibrium consisting of a zero entropy stationary cycle and weak ergodicity in its distribution function. Obvious extensions to larger state spaces were also discussed. For these, however, one had to appeal to the work in [5] to verify their existence. It was then pointed out the significant importance of these Markov chains that as point and cyclic attractors exhibit zero entropy in their cyclic equilibrium distributions.

3.1 Definitions

For the Markov chain processes considered here, one may assume that the Hajnal qualification holds [4]. This qualification imposes a constraint on the matrices comprising the structural cycle of the process. It requires that the cycle, consisting of a finite sequence of factor matrices, begin with a regular and scrambling matrix. This guarantees regularity throughout the required products of matrices. A scrambling matrix is any regular stochastic matrix such that for any two of its rows there always is at least one column with nonzero entries for both rows. This result is due to Hajnal in [17].

The transition matrices to be considered are all regular square stochastic matrices that correspond respectively to finite Markov chains all defined on the same state space. One knows, for example, that stochasticity of the matrices is a property that is closed under matrix multiplication.

For convenience, one may assume at this time that the $n \times n$ matrix M is defined in the state space $E = \{1, 2, 3, \ldots, n\}$ and that $m(i, j)$ represents the entry in the ith row that corresponds to "from" state i and the jth column that corresponds to "to" state j of the matrix M. The following proposition is of interest.

Proposition 1. In the matrix product $M = AB$, if $a(i, k) = 1$ and $b(k, j) = 1$, then $m(i, j) = 1$.

Proof. The proof of this proposition is very simple and is omitted here. ■

Proposition 1 is all that is needed to be able to understand how stationary cycles of zero entropy arise naturally in the context of finite time varying Markov chains.

Consider a finite sequence of k factor matrices and suppose that their product exhibits one and only one closed path connected by a sequence of transition probabilities (all equal to 1) linked as in Proposition 1. One may label such a closed path as follows:

$$(i_1, i_2)(i_2, i_3)(i_3, i_4) \ldots (i_{k-1}, i_k)(i_k, i_1),$$

where (i_j, i_{j+1}), which is equal to 1, represents the entry in the matrix product's jth matrix that corresponds to row i_j and column i_{j+1}.

With this label, the following may be observed:

$$\begin{array}{lll} (i_1, i_2)(i_2, i_3) & \text{yields} & (i_1, i_3) = 1 \\ (i_1, i_3)(i_3, i_4) & \text{yields} & (i_1, i_4) = 1 \end{array}$$

and so forth, until

$$(i_1, i_k)(i_k, i_1) \quad \text{yields} \quad (i_1, i_1) = 1$$

for the entire product of matrices.

Consider, for example, the case of three matrices in an alternating product sequence as follows: $\ldots ABCABCABCABC \ldots$, where A starts the cycle. Let the state space be given by $E = \{1, 2, 3, 4\}$ and let the single closed path be $(1,2)(2,4)(4,1)$. One may then notice the following matrix product relationships:

$$\begin{array}{lll} (1,2)(2,4)(4,1) & \text{yields} & (1,1) = 1, \\ (2,4)(4,1)(1,2) & \text{yields} & (2,2) = 1, \text{ and} \\ (4,1)(1,2)(2,4) & \text{yields} & (4,4) = 1. \end{array}$$

Following the developments in chapter 9 (p. 519) of [5], one has that $(1,1) = 1$ means that state 1 is absorbing in the matrix resulting from the product ABC and hence $(1,0,0,0)$ is the first vector in the stationary cycle. Similarly, $(2,2) = 1$ means that state 2 is absorbing in the matrix resulting from the product BCA and hence $(0,1,0,0)$ is the second vector in the stationary cycle. Finally, $(4,4) = 1$ means that state 4 is absorbing in the matrix resulting from the product CAB and hence $(0,0,0,1)$ is the third and final vector in the stationary cycle with zero entropy. In this case, once cyclic stationarity is reached or if the process begins at cyclic stationarity, state 3 vanishes from view. Thus, in the context of the time varying Markov chain, state 3 turns out to be a transient state. The equilibrium state frequency vector in this case is $(1/3, 1/3, 0, 1/3)$.

3.2 Representing information

The particular stationary cycles of probability distributions discussed so far, all have in common the distinction of being zero entropy state cyclic attractors. The processes could start at the highest levels of entropy and reach this kind of stationarity which is independent of both time and initial conditions. By a suitable interpretation of the states it is then possible to represent (or even cognize, or associate) information in (to) these processes. To represent binary numbers, for example, one would need a minimum of two states. For convenience, one may use a third state as a delimiter to separate the words. To represent large and small decimal numbers one would need a minimum of 10 states or 11 if a delimiter is used. The following is an example of how one could represent the binary form of the number 5 (101 in binary representation): Let state 0 represent the digit 0, and let state 1 represent the digit 1. State 2 will be used as a delimiter state (D). Then the following time varying Markov chain will yield a zero entropy stationary cycle containing this number:

$$\cdots \begin{pmatrix} x & x & x \\ x & x & x \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x & x & x \\ 1 & 0 & 0 \\ x & x & x \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ x & x & x \\ x & x & x \end{pmatrix} \begin{pmatrix} x & x & x \\ 0 & 0 & 1 \\ x & x & x \end{pmatrix} \cdots.$$

Here it is the case that

$$\begin{array}{lll} (2,1)(1,0)(0,1)(1,2) & \text{yields} & (2,2) = 1, \\ (1,0)(0,1)(1,2)(2,1) & \text{yields} & (1,1) = 1, \\ (0,1)(1,2)(2,1)(1,0) & \text{yields} & (0,0) = 1, \text{ and} \\ (1,2)(2,1)(1,0)(0,1) & \text{yields} & (1,1) = 1. \end{array}$$

Hence the stationary cycle is given by the vectors $(0,0,1)$, $(0,1,0)$, $(1,0,0)$, and $(0,1,0)$ which represents the sequence

$$\ldots D101D101D101D101\ldots.$$

The equilibrium state frequency vector is given by $(1/4, 1/2, 1/4)$.

Similarly, one could express the number 15 in its binary representation (1111) by the following time varying Markov chain:

$$\cdots\begin{pmatrix} x & x & x \\ x & x & x \\ 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} x & x & x \\ 0 & 1 & 0 \\ x & x & x \end{pmatrix}\begin{pmatrix} x & x & x \\ 0 & 1 & 0 \\ x & x & x \end{pmatrix}\begin{pmatrix} x & x & x \\ 0 & 1 & 0 \\ x & x & x \end{pmatrix}\begin{pmatrix} x & x & x \\ 0 & 0 & 1 \\ x & x & x \end{pmatrix}\cdots$$

which results in the following rotated closed paths:

$$\begin{array}{lll} (2,1)(1,1)(1,1)(1,1)(1,2) & \text{yields} & (2,2) = 1, \\ (1,1)(1,1)(1,1)(1,2)(2,1) & \text{yields} & (1,1) = 1, \\ (1,1)(1,1)(1,2)(2,1)(1,1) & \text{yields} & (1,1) = 1, \\ (1,1)(1,2)(2,1)(1,1)(1,1) & \text{yields} & (1,1) = 1, \text{ and} \\ (1,2)(2,1)(1,1)(1,1)(1,1) & \text{yields} & (1,1) = 1. \end{array}$$

This corresponds to the state sequence

$$\ldots D1111D1111D1111D1111D1111\ldots.$$

It is obvious that numbers in any base and words and sentences can be represented as desired. But this is not necessarily the way the brain represents this information in a literal sense. One resorts to these because they represent the foundation of the most important way we can externally relate to knowledge and information. The coding mechanics in the abstract, however, is entirely arbitrary. Responses and registration of all sensory inputs, for example, may be accounted for in this scheme. Self organizing and autocatalytic structures could also evolve dynamically along this type of pattern formation and recognition [6].

In traditional computer engineering, one codes in a higher order language and uses a compiler to help the machine translate this code into machine or assembly language. It appears that here one may be helping explain the machine or assembly language of the brain and what seems to be presented in [6] is an explanation of how a reverse compiler operates. In other words, how the brain is autonomously capable of reaching higher levels of order and complexity in its functions. It could well be that the very foundation of cognition and of the acquisition or generation of meaning to oneself and others lies in this reversal of roles of the brain as opposed to the computer.

3.3 Cycles with random path selection

There is one characteristic that is common to all finite time varying Markov chains exhibiting cyclic stationarity where the stationary cycles have zero entropy. After stationarity is reached (or if the process starts at cyclic stationarity) all trajectories or sample paths are identical. This is what turns the process into a state cyclic attractor.

Next one may consider a cyclic process where there are exactly two possible paths at cyclic stationarity. This is illustrated *via* a simple example consisting of a three step cycle in four states. At the beginning of each cycle a fair coin is tossed and either one of two possible paths is selected depending on the outcome of the coin toss.

The finite time varying Markov chain is defined in the state space $E = \{0, 1, 2, 3\}$, and the three step cycle is defined as follows:

$$\cdots \begin{pmatrix} 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 \\ x & x & x & x \\ 0 & 0 & 0 & 1 \\ x & x & x & x \end{pmatrix} \begin{pmatrix} x & x & x & x \\ 1 & 0 & 0 & 0 \\ x & x & x & x \\ 0 & 0 & 1 & 0 \end{pmatrix} \cdots .$$

This process is characterized by a stationary cycle given by the vectors $(1/2, 0, 1/2, 0)$, $(0, 1/2, 0, 1/2)$, and $(1/2, 0, 1/2, 0)$ with a state frequency vector given by $(1/3, 1/6, 1/3, 1/6)$. The two possible state paths at each cycle are $\{0, 1, 0\}$ and $\{2, 3, 2\}$. At each cycle, either path is selected with probability 1/2. It may be pointed out that the coin need not be fair. One may adjust the probabilities of each path by changing columns one and three in the first matrix of the cycle. States 1 and 3 are transient in the first matrix of the cycle yet they are positive recurrent in the context of the time varying process.

The stationary cycle of this process does not exhibit zero entropy and hence the process does not qualify as a state cyclic attractor. It does possess enough structural and behavioral characteristics, however, to show some of the qualities of a miniature state strange attractor.

Any two independent trajectories of duration one period (or cycle) are identical with probability 1/2. In general, the probability that any two independent trajectories of duration n periods are identical is $(1/2)^n$. Thus, it is easy to see that the probability that any two trajectories or sample paths are identical goes to zero very quickly. Not only is there an underlying equilibrium behavior (in terms of probabilities), however, but also enough cohesion and definition in the pattern of visits to the states to know that some states are grouped together independent of any random phenomena.

One can use the random number generator in an HP 11C calculator to simulate five independent trajectories of duration five periods (cycles) each. For convenience, one may label state 0 as Red, state 1 as Blue, state 2 as Green, and state 3 as White.

1. GWGRBRRBRRBRGWG...
2. GWGRBRRBRRBRRBR...
3. RBRGWGGWGGWGGWG...
4. GWGRBRGWGRBRRBR...
5. RBRRBRRBRRBRRBR...

To pick out which states are grouped together and in which order and which are not from just viewing a trajectory requires a mental process of association which, in a very rudimentary way, is akin to a "context sensitive constraining" that may emerge in the brain from the dynamic interaction of patterns and attractors [6]. This exercise, for the reader, is a bit more difficult if the view does not exhibit a starting point for the trajectory.

An even more difficult exercise will result if one were to flip a fair coin on every other step of the cycle. A four step cycle on four states follows for which this is the case:

$$\cdots \begin{pmatrix} 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 \\ x & x & x & x \\ 0 & 0 & 0 & 1 \\ x & x & x & x \end{pmatrix} \begin{pmatrix} 0 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} x & x & x & x \\ 1 & 0 & 0 & 0 \\ x & x & x & x \\ 0 & 0 & 1 & 0 \end{pmatrix} \cdots.$$

The stationary cycle for this process is given by the vectors $(1/2, 0, 1/2, 0)$, $(0, 1/2, 0, 1/2)$, $(0, 1/2, 0, 1/2)$, and $(1/2, 0, 1/2, 0)$. This results in an equilibrium frequency vector for the states given by $(1/4, 1/4, 1/4, 1/4)$. Any cycle can result in any one of four equally likely state paths given by $\{0, 1, 1, 0\}$, $\{0, 1, 3, 2\}$, $\{2, 3, 1, 0\}$, and $\{2, 3, 3, 2\}$.

Using the same color code, one can simulate five independent trajectories of duration five periods (cycles) each.

1. RBWGGWWGRBBRGWBRRBBR…
2. GWBRRBWGGWBRGWBRRBBR…
3. GWWGRBWGGWWGGWBRGWBR…
4. GWBRRBWGRBBRRBBRRBBR…
5. GWWGGWWGRBBRRBBRGWBR…

From viewing the resulting trajectories it appears easy to tell that Red and Blue as well as White and Green must be associated to each other somehow but picking out the four equally likely four step paths may not be so trivial.

In [6], citing [18], it is pointed out that the key to this problem lies in *context sensitive redundancy*, which is a kind of constraint that establishes divergence from independence: "By correlating and coordinating previously aggregated parts into a more complex, differentiated, systematic whole, contextual constraints enlarge the variety of states the system as a whole can access."

The two Markov chains described in this section exhibit characteristics of miniature state strange attractors. To help explain how this is so, one may quote [6] which vividly describes strange attractors in general as they are placed in the context of the new paradigm using complexity to model the processes of the human brain.

> In the last twenty-five years or so, a third type of attractor has been identified. So-called strange or complex attractors describe

> patterns of behavior so intricate that it is difficult to discern an overarching order amid the disorder they allow. All attractors represent characteristic behaviors or states that tend to draw the system towards themselves, but strange attractors are "thick," allowing individual behaviors to fluctuate so wildly that even though captured by the attractor's basin they appear unique. The width and convoluted shape of strange attractors imply that the overall pathway they describe is multiply realizable. Strange attractors describe ordered global patterns with such a high degree of local fluctuation, that is, that individual trajectories appear random, never quite exactly repeating the way the pendulum or chemical wave of the B-Z reaction does. The strange attractors of seemingly "chaotic" phenomena are therefore often not chaotic at all. Such intricate behavior patterns are evidence of highly complex, context dependent dynamic organizations.

Finally, in the context of time varying Markov chains, one may observe that the first of the two processes considered in this section consists of a three matrix product cycle in four states. If the cycle were to consist instead of the last two matrices in this triple, there would be a problem. There would be two separate closed paths yielding products that are nonregular or decomposable. Hence there would not be a stationary cycle. But neither matrix is a scrambling matrix (although both are regular) and the Hajnal qualification would be violated in this instance (see [4] and/or [17]). This qualification is sufficient to guarantee regular products but it is not necessary.

3.4 Irregular sample paths

The number of processes that can register signatures or recognizable patterns in the brain is much larger than the class of time varying Markov chains considered thus far. One other possibility, for example, is that of stochastic processes with irregular sample paths. A discrete time stochastic process defined on a discrete state space E is said to have irregular sample paths if at periodic epochs in time, the process either picks up and/or drops at least one state. One can easily construct such a time varying Markov chain with irregular sample paths. Consider the following three matrix product cycle:

$$\cdots\begin{pmatrix}0 & 1\\ x & x\end{pmatrix}\begin{pmatrix}x & x & x\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}x & x\\ x & x\\ 1 & 0\end{pmatrix}\cdots.$$

This process will produce the cyclic attractor zero entropy stationary cycle given by the vectors $(1,0)$, $(0,1)$, and $(0,0,1)$. Obviously one state is added before and then dropped after the third step of the cycle in every cycle even before reaching cyclic stationarity. Whether this process meets the criterion for weak ergodicity in the distribution function can be questioned under the circumstances. The lack of presence by state 2 in steps one and two of the cycle, however, could be construed as that state having probability zero in those steps.

4. Summary and conclusions

The work presented has intended to associate and relate aspects of the theory of time varying Markov chains, dynamic entities themselves, with concepts in the modern theory of complexity. It supports the developments in action theory brought forth in [6]. The common thread all throughout is the modeling of brain processes and their characteristics using concepts that are common to both theories. The new paradigm and more contemporary experimental evidence has helped organize one's thinking on the business of thinking. The following anonymous quote comes to mind "I think I think, therefore I am who I am."

Concerning the association between the theories of Markov chains underlying the mathematical modeling and the theory of complexity, [6] brings in a word of warning in the following commentary: "... As a result, unlike the near-equilibrium processes of traditional thermodynamics, complex systems do not forget their initial conditions: they 'carry their history on their backs' (Prigogine, Spring 1995, U. S. Naval Academy). Their origin constrains their trajectory."

Of course, equilibrium in traditional thermodynamics is akin to the kind of equilibrium offered by temporaly homogeneous Markov chains. Most of the homogeneous chains used to construct the time varying processes are far from their own equilibrium while forming part of a higher order level of equilibrium in cyclic stationarity and weak ergodicity in their distribution functions. The fact remains, however, that all of the processes discussed are Markov chains. This simply means they do not carry their history with them even though they are time varying or dynamic in nature. These processes may either constitute a rare exception or possibly may not exhibit a high enough degree of complexity to qualify as full fledged complex systems. Yet they do exhibit characteristics associated to the latter. Reconciling the two concepts might prove a nontrivial task.

In the context of time varying Markov chains, when conditions for equilibrium exist, the time independence of the stationary cycle is best understood by contrasting the steady-cycle of distributions to transient cyclical behavior (time-dependent) as the process converges towards cyclic stationarity. It also helps to conceptualize an easily formulated lattice of temporary cycles each holding for long enough to approach its own stationary cycle but ultimately taking a quantum leap to the next distinct structural cycle with a different stationary cycle at some point during the trajectory. Such a process would be completely time dependent and far from any kind of equilibrium. It may be necessary as well to reconcile complexity with this particular concept of equilibrium and nonequilibrium as it relates to Markov chains and their role as attractors. To a certain extent, this may have already been accomplished in [6]. It is suggested that stimuli to the brain simply generate origins that constrain the structure of the cycle rather than the trajectory directly. Once a structural cycle is selected, the brain emulates equilibrium conditions.

It appears that to the inexorable march of thermal entropy to infinity corresponds an inexorable march of information entropy to zero and that the price of temporary reversals or increases in information entropy is temporary reversals or decreases in thermal entropy.

In short, we have provided here mathematical models that illustrate the mechanics of processes ascribed to the brain in virtue of exhibited characteristics and by the backing of biological knowledge and some more recent experimental evidence. These processes can register signatures and recognizable patterns in the brain.

References

[1] S. M. Ross, *Introduction to Probability Models* (Academic Press, Orlando, FL, 1985).

[2] M. M. Waldrop, *Complexity* (Simon & Schuster, New York, 1992).

[3] D. R. Roqué, "Queueing Networks Structured via Interacting Overflow Lines," Doctoral dissertation, George Washington University, University Microfilms #8302633, Ann Arbor, MI, 1982.

[4] D. R. Roqué, "Rainfall in Little Havana," *Proceedings of the Association for the Study of the Cuban Economy (ASCE)*, **8** (1998) 142–149, available at http://www.ascecuba.org.

[5] R. A. Howard, *Dynamic Probabilistic Systems*, volume 1 (John Wiley, New York, 1971).

[6] A. Juarrero, *Dynamics in Action* (The MIT Press, Cambridge, MA, 1999).

[7] C. J. Shatz, "The Developing Brain," in *The Scientific American Book of the Brain* (The Lyons Press, New York, 1999).

[8] G. D. Fischbach, "Mind and Brain," *Scientific American*, Special Report SCA45001, 1992.

[9] J. Elman, "Language as a Dynamical System," in *Mind as Motion*, edited by R. Port and T. van Gelder (The MIT Press, Cambridge, MA, 1995).

[10] Van Court Hare, Jr., *Systems Analysis: A Diagnostic Approach* (Harcourt, Brace & World, Inc., New York, 1967).

[11] R. V. L. Hartley, "Transmission of Information," *Bell Systems Technical Journal*, 7 (1928) 535–563.

[12] C. E. Shannon, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, 1949).

[13] S. M. Ross, *Applied Probability Models with Optimization Applications* (Holden-Day, San Francisco, CA, 1970).

[14] J. Horgan, "Can Science Explain Consciousness?" in *The Scientific American Book of the Brain* (The Lyons Press, New York, 1999).

[15] *Scientific Encyclopedia*, fifth edition (Van Nostrand, New York, 1968).

[16] R. L. Devaney, *A First Course in Chaotic Dynamical Systems* (Addison-Wesley, Reading, MA, 1992).

[17] J. Hajnal, "Weak Ergodicity in Non-Homogeneous Markov Chains," *Proceedings Cambridge Philosophical Society*, 54 (1958) 233–246.

[18] L. Gatlin, *Information and the Living System* (Columbia University Press, New York, 1972).