

On Cognition, Structured Sequence Processing, and Adaptive Dynamical Systems

Karl Magnus Petersson ^a

^aMax Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

Abstract. Cognitive neuroscience approaches the brain as a cognitive system: a system that functionally is conceptualized in terms of information processing. We outline some aspects of this concept and consider a physical system to be an information processing device when a subclass of its physical states can be viewed as representational/cognitive and transitions between these can be conceptualized as a process operating on these states by implementing operations on the corresponding representational structures. We identify a generic and fundamental problem in cognition: sequentially organized structured processing. Structured sequence processing provides the brain, in an essential sense, with its processing logic. In an approach addressing this problem, we illustrate how to integrate levels of analysis within a framework of adaptive dynamical systems. We note that the dynamical system framework lends itself to a description of asynchronous event-driven devices, which is likely to be important in cognition because the brain appears to be an asynchronous processing system. We use the human language faculty and natural language processing as a concrete example through out.

Keywords: cognition, sequence processing, neural network, adaptive, dynamical systems, asynchronous processing.

PACS: 87.19.lf, 87.19.lj, 87.19.ls, 87.19.lw

CLASSICAL MODELS OF COGNITION

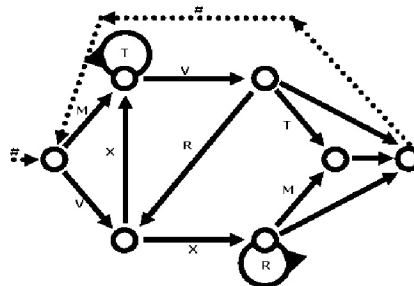


FIGURE 1. The Reber machine: an information processing device formulated within the Church-Turing framework.

A common framework for describing cognitive systems in cognitive neuroscience is Marr's three descriptive levels of analysis: the computational, the algorithmic, and the implementational [1]. Compared to classical cognitive science, cognitive neuroscience typically works with a modified perspective including simultaneously perspectives on cognition, learning, and development. However, the underlying ideas are essentially similar [2]. Within this framework, the three descriptive levels are generalized to: 1) the *cognitive level* - a formal theory of structured cognitive states and the cognitive transition system [3-5]; 2) the *dynamical system level* - given a formal cognitive theory, a state-space is specified and processing is formulated in terms of a dynamical system that embeds the cognitive level system specification [5, 6]; 3) the *implementation level* - given a dynamical system, this level specifies the physical hardware implementation of the dynamical system, for example a neural network [7-9]. The framework of classical cognitive science [10] assumes that information is coded by structured representations ("data structures") and that cognitive processing is accomplished by the execution of algorithmic operations ("rules") on the basic elements of compositionally structured representations [3, see also 4, 11]. This paradigm suggests that cognitive phenomena can be modeled within the framework of Church-Turing computability and effectively takes the view that isomorphic models of cognition can be found within this framework [4, 11]. The classical framework has served well for language modeling in linguistics/psycholinguistics and all common formal language models can be described within this framework [12, 13]. In the following, it is important to note that all information processing devices in the Church-Turing framework can be reformulated as discrete-time dynamical systems [5, 9]. A typical

example of a transition function T in the Church-Turing framework is given in **Figure 1**. T corresponds to a specific finite-state architecture (e.g., the control of a Turing machine) that specifies the internal states (nodes) and valid transitions between states (arrows). Cognition is equated with internal information processing in interaction with its (computational) environment (cf., **Figure 2**) and is specified by the following components: the space of internal states Ω ($s \in \Omega$); a set of dynamical principles $T: \Omega \times \Sigma \rightarrow \Omega \times \Lambda$ that govern information processing: given an internal state $s \in \Omega$ and input u (transformed according to $i = f(u) \in \Sigma$ by some transduction mechanism), T specifies a new internal state $T(s, i) \in \Omega$, while output $\lambda \in \Lambda$ is generated optionally according to a suitable output transformation $\lambda = g(s, i)$, which can be taken to be a component of T . Thus at time t , the system receives input $i(t)$ in state $s(t)$, and the system changes state into $s(t+\Delta t)$ and outputs $\lambda(t+\Delta t)$ according to: $[s(t+\Delta t), \lambda(t+\Delta t)] = T[s(t), i(t)]$ (1) which is the description of a forced discrete-time dynamical system and we emphasize that (1) is one manifestation of *sequentially organized, structured processing* or *hierarchically structured sequence processing*. When we recognize that there is a one-to-one correspondence between *discrete-time dynamical system* of the form (1), operating on discretized (quantized) values, *formal grammars*, and *computational mechanisms of Church-Turing type*, we immediately realize that, in an essential sense, the specifics of (1) provides the dynamical system with its *processing logic*. We note that the dynamical system formulation (1) of Church-Turing devices also is suitable for asynchronous event-driven devices in which logical time (dependent sequence order) rather than physical time is the crucial abstraction. In these systems physical time implements logical time and this is the fundamental reason for the qualification that only a *subclass* of the available physical states is representational, while the totality of accessible physical states basically serves as a necessary part of the implementation medium. Here, we have not explicitly specified the memory organization of the system. It is important to note that the characteristics of the memory organization, in particular whether the storage capacity is finite or infinite, is a crucial determinant of the productive/parsing power of the architecture [9]. However, it seems natural to assume that the brain is finite with respect to its memory organization and if one assumes that the brain implements a classical model of cognition, it follows that this can be represented as a finite-state architecture (note that a finite-state machine will behave as a Turing machine as long as the memory limitations are not encountered). However, it is possible to view classical models as approximate abstract descriptions of cognitive brain functions and a complementary perspective is offered by neural network models. It can be shown that cognition, viewed as a neurobiological system with its inherent limitations (i.e., if one assumes that the brain sustains some level of noise or does not utilize infinite precision processing; either of these assumptions seems necessary for physical realizability) is well-approximated by finite-state behavior [9].

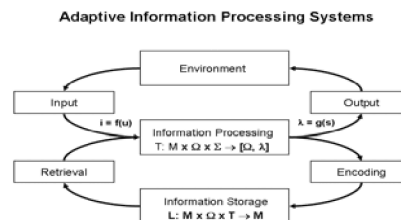


FIGURE 2. A general outline of adaptive information processing systems.

In order to incorporate learning/development in this picture, the processing dynamics, T , has to be extended with an adaptive learning dynamics, L (**Figure 2**). For simplicity, let us focus on some particular cognitive system/module C , which is fundamental in the sense that all normal individuals develop cognitive capacities related to C . Given C , we conceptualize developmental/learning processes as a trajectory in C 's accessible configuration space $M = [m \mid \text{the model } m \text{ can be instantiated by } C]$. The developmental/learning process is driven by C 's interaction with its environment in conjunction with innately specified developmental processes. Thus, at any point in time C is in a given configuration $m_C(t)$ (i.e., C instantiates the model $m_C(t)$). If we assume that C incorporates innately specified prior structure, we can capture this with the notion of a structured initial state, $m_C(t_0)$. As C develops, it traces out a trajectory in M determined by its adaptive dynamics L according to: $m_C(t+\Delta t) = L(m_C(t), s(t), t)$ (2) where the explicit dependence on time in L captures the idea of an innately specified developmental process ("maturation") and possible dependence on the developmental history. If C and L are such that $m_C(t)$ converges on a final model $m_C[F]$, characterizing the end-state of the developmental process reached after time Δt_F : $m_C(t_0 + \Delta t_F) \approx m_C[F]$. The configuration m_C determines the transition function T in the following sense: T is parameterized by m_C , $T = T[m_C]$ (or $T: M \times \Omega \times \Sigma \rightarrow \Omega \times \Lambda$), and (1) is modified according to: $[s(t+\Delta t), \lambda(t+\Delta t)] = T[m_C, s(t), i(t)]$ (1')

The adaptive framework outlined here is essentially similar to Chomsky's well-known hypothesis concerning language acquisition [14] where the module C is taken to be the human faculty of language, L the language acquisition device, and the configuration space M is determined by an innate universal grammar. Different aspects of

the universal grammar are captured by M , L , and the initial state $m_C(t_0)$. Chomsky and others have argued extensively that the inherent properties of M , L , and $m_C(t_0)$ are determined by innately specified (genetic) factors. From this point of view, language acquisition is the result of an interaction between two sources of information: 1) an *innate prior structure*, which is likely to be both language specific and general non-language specific in nature; and 2) the *environment*, both the linguistic and the extra-linguistic (which can be viewed as an interactive boundary condition [2]). The type of learning that characterizes language acquisition is largely implicit, self-organized in nature, as opposed to the acquisition of reading/writing skills, which has an explicit supervised component [15, 16].

The nervous system is naturally viewed as a biophysical dynamical system and the temporal evolution of a given brain system, determined by its underlying neuronal dynamics, corresponds to transitions between physical states, which we here have conceptualize as information processing in terms of transitions between cognitive states. The brain's recurrent network architecture is a prerequisite for integration of functionally specialized brain regions. This set of facts can serve as a basis for a general dynamical systems perspective on cognitive brain functions and suggests a form of functional modularity, *dynamic functional modularity*, which naturally lends itself to a description in terms of Marr's three modified levels of analysis [2].

NON-CLASSICAL DYNAMICAL MODELS OF COGNITION

A developmental learning system can be conceptualized as a system of coupled dynamics: 1) a dynamics for processing of information in interaction with 2) a dynamics for learning, memory, and development; for example, **(1')** and **(2)** might correspond to the perception-cognition-action cycle and the encoding-storage-retrieval cycle of **Figure 2** [2, 16]. We thus arrive at a dynamic conceptualization of a cognitive learning system $D = \langle \text{functional architecture, representational dynamics, learning dynamics} \rangle$ in terms of: 1) *functional architecture*: a specification of the structural organization of the system (e.g., a network architecture); 2) *representational dynamics*: a specification of a state-space Ω carrying/representing information, and dynamical principles, $T: M \times \Omega \times \Sigma \rightarrow \Omega \times \Lambda$, governing the active processing of information (e.g., evolution of membrane potentials and action potential generation); and 3) *learning dynamics*: a configuration space M , which includes a specification of learning (adaptive) variables/parameters, m , for memory formation and information storage in a general sense (e.g., synaptic parameters), as well as dynamical principles, a "learning algorithm", $L: M \times \Omega \rightarrow M$, that govern the temporal evolution of the learning variables in configuration space [e.g., Hebbian learning; 7, 17]. The temporal evolution of the adaptive parameters depends on the active processing of information and the learning dynamics is typically, but not necessarily, conceived of as taking place on a slower (longer) characteristic time-scale than that of the representational dynamics. This formulation is directly reflected in the interaction between the two cycles of **Figure 2** which correspond to active information processing in interaction with learning/memory. To be more explicit, this can for example be formulated within the framework of coupled multivariate stochastic differential/difference equation with additive noise processes $\xi(t)$ and $\eta(t)$ [18]: $ds = T(m,s,i)dt + d\xi(t)$ and $dm = L(m,s)dt + d\eta(t)$ **(3)** where $i = f(u)$ is the input representation, and $\lambda = g(s,i)$ is the output representation. Note that the representational dynamics corresponds to the first part of **(3)** while the learning dynamics corresponds to the second, and **(3)** correspond directly to the classical picture given by **(1')** and **(2)**. More general formulations are also possible (e.g., in terms of semi-groups of operators). As in the case of **(1)**, or **(1')** and **(2)**, we emphasize that **(3)** is a manifestation of the concept of sequentially organized, structured processing, which also in this case provides the dynamical system with its processing logic, here understood in the generalized terms of state-space dynamics. Again, this formulation offers a descriptive framework for *asynchronous event-driven devices* and also in this case, *logical time* and *representational subclass* are important abstractions, where physical time and the totality of available physical states serve as implementation media. We note, however, general dynamical system theory [19] is in a sense obviously too rich a framework; for example, it turns out that for any given state-space one can find a universal dynamical system whose traces will generate any dynamics on that state-space [19]. Thus, it is crucial to specify the relevant constraints and processing principles from cognition and the underlying neurobiological networks [2]. In a very important sense, most of the empirical and theoretical work necessary for a deeper understanding of cognitive brain functions remains in order to specify detailed and integrated models of cognition, development, and learning.

INTEGRATION OF THE LEVELS OF DESCRIPTION

The point of the discussion so far has been to specify an explicit form for cognitive models formulated as adaptive dynamical systems, both classical and non-classical, and we have emphasized the close similarity in form between the two. This provides a context for how cognitive states, viewed as physical states, and operations on these can be empirically characterized with functional neuroimaging techniques (PET/FMRI/EEG/MEG/TMS). These methods allow direct or indirect quantitative measurements of aspects of the physical states of the brain. The problem of how the brain processes structured sequences of information is a generic and fundamental problem in cognition.

Human language is of course the cardinal example and in a series of empirical and theoretical studies, which we review briefly, we have approached this problem from the point of view of implicit artificial grammar learning (AGL). A typical implicit AGL experiment includes an acquisition phase followed by a classification test [20]. During implicit acquisition, participants are engaged in a short-term memory task, without performance feedback, using an acquisition sample of symbol sequences generated from an artificial grammar (**Figure 1**). Subsequently, subjects are asked to classify new items as grammatical or not based on their immediate impression (guessing based on "gut feeling") and perform reliably above chance (**Figure 3**). Thus, participants have acquired knowledge about relevant aspects of the underlying generative mechanism and subjects can typically provide little (if any) sufficient reasons to motivate their classification decisions [15, 21, 22]. An alternative way of assessing the implicit acquisition of a grammar is the structural mere exposure version of AGL with preference classification [15]. The mere exposure effect refers to the finding that repeated exposure to a stimulus induces an increased preference for that stimulus compared to novel stimuli. Folia et al. [15] used the Reber grammar represented in **Figure 1** and characterized a novel implicit structural mere exposure AGL paradigm using preference classification in a 2x2 factorial design with local substrings familiarity (high/low) relative to the acquisition string set and grammaticality (G/NG) as factors. The results showed that preference and grammaticality classification are qualitatively equivalent in terms of behavior and strongly support the notion that humans implicitly acquire knowledge about a complex system of interacting rules by mere exposure to the acquisition material (**Figure 3**). Recent fMRI investigations of similar AGL paradigms (**Figure 3**) have shown that grammaticality (G vs. NG) correlates with caudate nucleus activity while syntactic violations (NG vs. G) correlate with activation in the left inferior frontal region [BA 44/45; 20, 22]. These results are consistent with analogous findings in natural language processing [23] and lend further support to the suggestion that the left inferior frontal region has a specific role in the processing structural regularities.

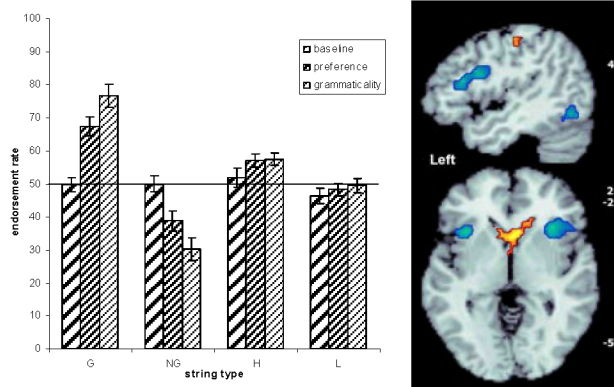


FIGURE 3. Implicit AGL. Left: Behavioral results (cf., [15]). Right: fMRI results (cf., [22]).

In order to get a handle on implicit grammar acquisition and how sequentially organized, structured processing of structured sequences, might take place in a recurrent neural network, we proceeded with a series of computer experiments using predictive networks based on the simple recurrent network (SRN) architecture [gradient-descent error back-propagation; adaptable learning rate and momentum; for details see 24].

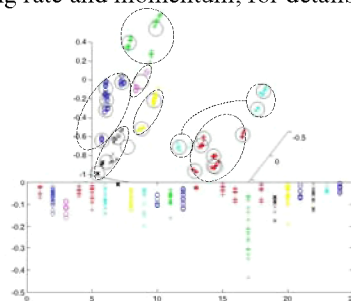


FIGURE 4. The activation of the hidden nodes plotted along the first three principal axes obtained by PCA. The dots correspond to internal states of the Reber machine (**Figure 1**) and cluster in the state-space of the SRN (cf., [24]).

In brief, the discrete-time SRN can be viewed as a simple network analogue of the finite state architecture [9] specified by a coupled dynamics of the form (3) without explicit noise terms. Our analysis of the state-space dynamics of the network reproduces the dynamics of the Reber machine (**Figure 4**). In other words, we observe an "isomorphic" model of a classical cognitive processing device being embedded in the non-classical network dynamics. Grammar acquisition, whether natural or artificial, is commonly conceptualized in terms of a structure-based ("rule") acquisition mechanism or a statistical learning mechanism. Our analysis suggests that the recurrent

network learned to represent the rules of the grammar in its state-space dynamics [24]. Since the learning dynamics we used is commonly viewed as a "statistical learning algorithm", the two perspectives on grammar acquisition are complementary. The combined approach of analyzing the state-space behavior of the network with a competitive network in combination with principle components analysis (PCA) provides an approach to grammar extraction from the state-space dynamics, which is closely related to the symbolic dynamics approach used in the analysis of non-linear dynamical systems [25].

CONCLUSIONS

Cognitive neuroscience approaches the brain as an information processing device and here we have outlined some relevant aspects of this concept. In particular, we have identified a problem in cognition, which is both generic and fundamental, the problem of *structured sequence processing*. We argue that the natural generalization of the concept *processing logic* in Church-Turing type devices is *state-space dynamics* in dynamical system models of information processing. In an approach addressing the problem of structured sequence processing we illustrate how to integrate levels of analysis within the framework of adaptive dynamical systems. We argue that dynamical systems offer a descriptive framework for asynchronous event-driven devices. This is likely to be of central importance in cognition because the brain appears to be an asynchronous processing system. Finally, it is clear that our work on implicit AGL is in its infancy with respect to a more profound understanding of sequentially organized, structured processing in the brain. We have outlined a series of empirical and theoretical approaches to the problem. The integration of these approaches is hinted at by the common conceptual basis outlined in the first part of the paper. Important issues to address in future work include determining what might be a relevant state-space, with associated dynamics, in order to understand active information processing in general. Moreover, issues related to what might serve as relevant prior constraints and the characteristics of the adaptive machinery that allows all normal human individuals to acquire a native language robustly and in a relatively short period of time need to be addressed. These are essentially questions about the nature of the neural code, the dynamical variables that serve as information carriers, and its adaptive characteristics, which in the end allow all humans, through acquired cognitive skills, to actively participate and contribute to the process of cultural transmission.

REFERENCES

1. D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: W. H. Freeman and Company, 1982.
2. K. M. Petersson, *Learning and Memory in the Human Brain*. Stockholm: Karolinska University Press, 2005.
3. A. Wells, *Rethinking Cognitive Computation: Turing and the Science of the Mind*. Palgrave Macmillan, 2005.
4. M. Davis, R. Sigal, and E. J. Weyuker, *Computability, Complexity, and Languages: Fundamentals of Theoretical Computer Science* (2nd ed.). San Diego, CA: Academic Press, 1994.
5. H. T. Siegelmann, *Neural Networks and Analog Computation: Beyond the Turing Limit*. Basel: Birkhäuser, 1999.
6. H. T. Siegelmann and S. Fishman, *Physica D*, 120: 214-235.
7. M. A. Arbib, *The Handbook of Brain Theory and Neural Networks* (2nd ed.). Cambridge, MA: MIT Press, 2003.
8. F. C. Hoppensteadt and E. M. Izhikevich, *Weakly Connected Neural Networks*. New York: Springer, 1997.
9. K. M. Petersson, *Neurocomputing*, 65-66: 825-832, 2005.
10. A. Newell and H. Simon, *Commun. Assoc. Comput. Machinery*, 19: 111-126, 1976.
11. N. J. Cutland, *Computability: An Introduction to Recursive Function Theory*. Cambridge, UK: CUP, 1980.
12. R. Jackendoff, *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford, UK: OUP, 2002.
13. B. H. Partee, A. Ter Meulen, R. Wall, *Mathematical Methods in Linguistics*. Dordrecht: Kluwer Academic Publishers, 1990.
14. N. Chomsky, *Knowledge of Language*. New York: Praeger, 1986.
15. V. Folia, J. Uddén, C. Forkstam, M. Ingvar, P. Hagoort, and K. M. Petersson, *Implicit learning and dyslexia*, 2007.
16. K. M. Petersson, M. Ingvar, and A. Reis, "Language and literacy from a cognitive neuroscience perspective" in *Cambridge Handbook of Literacy*, edited by D. R. Olson & N. Torrance (Eds.). Cambridge, UK: CUP, 2008, forthcoming.
17. C. Koch, *Biophysics of Computation: Information Processing in Single Neurons*. Oxford, UK: OUP, 1999.
18. B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications* (5th ed.). Berlin: Springer, 2000.
19. A. Lasota and M. C. Mackey, *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*. New York: Springer, 1994.
20. K. M. Petersson, C. Forkstam, and M. Ingvar, *Cognitive Science*, 28, 383-407, 2004.
21. C. Forkstam, P. Hagoort, G. Fernandez, M. Ingvar, and K. M. Petersson, *NeuroImage*, 32, 956-967, 2006.
22. C. Forkstam and K. M. Petersson, *Proceedings of the Cognitive Science Society*, 696-701, 2005.
23. S. Bookheimer, *Annual Review of Neuroscience*, 25: 151-188, 2002.
24. K. M. Petersson, P. Grenholm, and C. Forkstam, *Proceedings of the Cognitive Science Society*, 1726-1731, 2005.
25. D. Lind and B. Marcus, *Symbolic Dynamics and Coding*. Cambridge, UK: CUP, 1999.