The Borderline Between Subsymbolic and Symbolic Processing: A Developmental Cognitive Neuroscience Approach

J.G.Wallace (jgwallace@swin.edu.au)
Information Technology Institute, Swinburne University of Technology
PO Box 218, Hawthorn, Victoria Australia 3122

K.Bluff (kbluff@swin.edu.au)
School of Information Technology, Swinburne University of Technology
PO Box 218, Hawthorn, Victoria Australia 3122

Abstract

Ideas, empirical data and methodologies from a broad range of disciplines are deployed in exploring the functional borderline between subsymbolic and symbolic processing in human cognition. Initial clarification of functional relationships between the two forms of representation involves a brain monitoring study based on the concept of 'semantic transparency.' The search for further clarification focusses on two major issues, the ontogenetic and phylogenetic origins of local neural areas and processes underlying formation of distal associations between them. Pursuing these objectives has proved to be a challenging, interdisciplinary enterprise. A model of development of local neural areas is presented which assigns a critical role to astrocytes and their interaction with adjacent neurons. An extension to include the phylogenetic dimension, draws on the concept of 'cortical inheritance', a largely ignored aspect of genetic theory. An account of distal association formation involves co-option of hippocampal place fields for a new use.

Introduction

The goal of the work to be reported is exploration of the nature and development of the functional borderline between non or subsymbolic processing and symbolic processing in human cognition. Pursuing this objective has involved the combination of ideas, empirical data and research methodologies derived from a broad range of disciplines. Johnson's (1997) recent definition of 'developmental cognitive neuroscience' captures much of the territory involved. We do, however, range more widely across both domains and levels of explanation than he considers appropriate. This will become evident in the following outline of our work to date.

In seeking initial clarification of functional relationships between the two forms of representation we adopted Smolensky's (1988) concept of 'semantic transparency' as an analytic tool. For our purposes a suitable definition is provided by Clark (1989) in an extended discussion of Smolensky's views: "A system will be said to be 'semantically transparent' just in case it is possible to describe a neat mapping between a symbolic (conceptual level) semantic description of the system's behaviour and some 'projectible' semantic interpretation of the internally represented 'objects' of its formal computational activity." (p.18). This definition provides a theoretical basis for the formal description of a range of systems varying in their degree of semantic transparency. At one end of the spectrum providing an example of total semantic transparency is a classic symbolic system in which all internal processing proceeds by manipulation of symbols directly and constantly mappable onto features of the system's interaction with its environment. A total absence of semantic transparency would be exhibited by a subsymbolic PDP system in which internal processing proceeded via units devoid of constancy across contexts and any identifiable mapping onto environmental features. In using semantic transparency as a means of assessing the extent to which cognitive tasks draw upon subsymbolic or symbolic processing the critical variable is the content of conscious awareness. The degree of semantic transparency of a segment of cognitive performance is determined by the extent to which the course and/or results of internal processing can register in conscious awareness as a result of mapping onto linguistic symbols or non-linguistic symbols such as sensory images. The extent of mapping is assumed to reflect the degree of correspondence between the grain or segmentation of processing and the modularity of linguistic or sensory symbols.

A detailed account of our empirical investigation is provided elsewhere (Wallace and Bluff, 1994). Three quantification tasks were selected on the basis of subjects' introspective reports during their performance. The first, counting by ones, is characterised by a high degree of semantic transparency; the second, subitising, represents a very low degree of semantic transparency and the third, subitising and adding, falls between the first and second on the spectrum. Steady state visually evoked potential data (SSVEP) were acquired from adult subjects engaging in each type of quantification, (Silberstein et al., 1990 a, b). The results were consistent with the distinctive and interactive nature of the two forms of representation. Significant variations were detected in the activity level of the prefrontal cortex and the parieto/temporal region of the right cerebral hemisphere provided a further source of discrimination.

The next phase of our work involved adoption of a developmental perspective. Newell (1990) attributes the existence of a physical symbol system to the fundamental restriction imposed on the amount of information represented and processed in a local neural area by the limited energy available. Symbol tokens overcome the restriction by providing distal access to another local area and further energy while maintaining the linkage necessary for integrated information processing. Critical to further specification of this account of the origins of symbolic processing is clarification of the nature and capabilities of subsymbolic processing proceeding within a local neural area. In seeking to contribute to this clarification we have focussed on two major issues, the ontogenetic and phylogenetic origins of local neural areas and the processes underlying formation of distal associations between them. Pursuing these objectives has proved to be a challenging, interdisciplinary enterprise.

Development of Local Neural Areas

Brain and, more specifically, cortical rhythmic activity provides a means of defining local neural processing. Local resonances are relatively high frequency rhythmic processes in the 30-80 Hz range with a spatial dimension in the millimetric range. A major distinction between local and lower frequency global/regional resonant modes is that the local resonant modes are not coherent over distances of more than a few millimeters. Human regional and global resonant modes, in contrast, are always coherent over a distance of centimeters. It is suggested that cognitive behavioral or perceptual states requiring a high degree of spatio-temporal specificity in neocortical activation are only consistent with local states, (Silberstein, 1995). We focus on the local cortical neural mechanisms at the millimetric level underlying the acquisition and operation of specific cognitive performance. Although local acquisition or learning depends on the functioning of distant areas such as the hippocampus and the striatum their role is relatively global and complements the spatio-temporal specifics that emerge in local areas, (Wickens, 1993).

Our approach deviates from the current norm by assigning a critical information-processing role to glia and, more specifically, astrocytes and their interactions with neurons. Although one of the two major cell types composing the brain, glia have been assigned an entirely auxiliary role in which they provide physical, trophic and metabolic support to neurons. Recently, however, it has been established that, although astrocytes show no evidence of the type of electrical excitability involved in signalling between neurons, they exhibit intracellular calcium dynamics that provides an alternative basis for signalling between them. Finkbeiner (1992, 1993), Smith (1992, 1994) and others have, accordingly, proposed that astrocyte networks might mediate slow modulations of neural function, like those underlying arousal, selective attention, motivational state, mood change, learning and memory.

Our model of the development of local neural areas assigns two critical roles to astrocytes: the formation of associations between astrocytes as a result of adjacent neural functioning defines the dimensions of a local neural area; astrocyte based mechanisms operate a boundary or barrier which compartmentalizes a local neural area in relation to adjacent cortical areas.

A detailed account of the model and the wide range of experimental evidence from which it is derived is presented elsewhere, (Wallace and Bluff, 1995 a, b; 1997). The multiplicity of roles played by calcium (Ca^{2+}) and, specifically, fluctuations in the concentration of free intracellular Ca^{2+} provide the framework of the model and offer a new approach to the construction and operation of complex, neural connectivity.

Astrocytes, unlike neurons, can only communicate directly with other astrocytes if they are immediately spatially proximal, as the communication occurs through the medium of a gap junction. The interaction between astrocytes and neural synaptic activity is simulated as a two step process. As neural activity proceeds the connection strengths of synapses are adjusted by a competitive learning rule. Occurrence of adjacent synaptic activity registers on specific locations on the membrane of astrocytes. Continuation results in release of Ca^{2+} from internal stores close to the initial or trigger location and initiation of Ca^{2+} waves traversing the intracellular space. The wave direction is determined by the relative levels of free Ca^{2+} concentration elsewhere in the astrocyte cytoplasm. Wave propagation beyond the initiating cell to adjoining astrocytes depends on co-occurrence of raised free Ca^{2+} levels on both sides of a gap junction.

The considerable difference in time scales of neural and astrocytic mechanisms is reflected in adoption of a time window representing the number of cycles of neural activation (and thus synaptic strength modification) which constitute a cycle of astrocytic activity. Over time consistent associations between astrocytes are established by repeated waves. The spatial direction of association chains is represented by increases in the base levels of appropriately situated Ca^{2+} stores adjacent to trigger locations and established gap junctions. The effect of the formation of astrocytic association chains cycles back on neurons via their gap junctions. This results in a rise in the intraneuronal free Ca^{2+} level and may produce interneuronal Ca^{2+} waves. When the Ca^{2+} level in a neuron reaches a threshold the plasticity of its current synaptic connections is reduced to a level consistent with long-term stabilization. The processes of astrocytic and neural association formation via gap junctions and synaptic connection stabilization interactively define the characteristics of functionally significant local neural areas.

In seeking to extend our work to include the phylogenetic dimension and, more specifically the intragenerational learning represented in local neural area phenotypes we have taken inspiration from a controversial and largely ignored aspect of genetic theory. Sapp (1987) provides the background to the concept of 'cortical inheritance'; in this context 'cortex' refers to the complex structures that make up the cell surface (skeleton or ectoplasm) of ciliated protozoa. Genetic studies (Sonneborn, 1970) indicate that the basis for the structures seems to be contained in the cell cortex which carries information for its gross organization and transmits it to progeny independently of the genes. Structural elements of the ciliate cortex provide the 'scaffolding' for the insertion of new organelles. Ciliate geneticists generally conclude that the location of the cortical parts is not random but, at the time of their development, forces exterior to the organelle itself dictate their location, orientation and number. There is no consensus on the nature of the forces but Hershey (1970) links them to ectoplasmic spatial principles underlying cell polarity.

The inheritance of cell organization through structural information maintained and transmitted by supra molecular structures clearly represents a challenge to the claim that biological evolution is solely the evolution of nucleotide sequences. The relevance of the idea of cortical inheritance for our work arises from the nature of organelles. Peripheral organelles consist of endoplasmic reticulum and some mitochondria, both of which act as releasable Ca^{2+} stores. We, thus, have a basis for linking local neural area phenotypes to a genotypic mechanism by means of the astrocytic Ca^{2+} stores spatially adjacent to trigger locations and established gap junctions since they provide an architectural trace of local neural areas arising from intragenerational learning.

In seeking to incorporate the idea of cortical inheritance into our modelling we have been influenced by the dominant neuroanatomical view that the majority of cerebral cortical tissue is largely equipotent early in epigenesis. Evidence, such as that reviewed by Johnson (1997), however, suggests a more probabilistic epigenetic view in which certain cortical areas have a detailed architecture slightly different from the basic neural structure common to the cortex. This makes them the most efficient at processing certain types of input. It is possible for other areas to carry out similar processing but not so efficiently. We take this view as our point of departure.

In seeking to translate astrocytic Ca^{2+} store traces of local neural area architecture into a genotypic form suitable as a basis for the next generation of a neural network we are currently exploring the effectiveness of a topological rather than biological approach. Discrete point positions distributed in space can be characterised using a representational network called a minimal spanning tree (MST). The MST technique aims to quantify spatial dot patterns by revealing hidden nearest-neighbour correlations. It has recently been applied in a broad range of research areas but has received relatively little attention in the biosciences, (Jones et al., 1996). In our case the locations of the Ca^{2+} stores defining local areas provide the spatial dot pattern from which genotypes are derived reflecting intragenerational learning.

An increasing amount of biological evidence supports the plausibility of the type of close relationship proposed between astrocytes and neurons in the initial wiring of the cortex. Astrocytes play a critical role in determining the intra-neuronal Ca^{2+} concentration level controlling the direction and continuity of neural growth in the pre-, peri- and post-natal period, (Kater and Shibara, 1994; Kater and Lipton, 1995). Studies involving Ca^{2+} imaging of cortical slices from embryonic and immediately post-natal rats and mice reveal spontaneous waves of increased Ca^{2+} spreading through clusters of neurons (Charles et al., 1996; Yuste et al., 1995). Waves did not propagate to distant neurons or to neighbouring neuronal clusters, even when there were clearly visible processes between the two clusters. In some cases neural Ca^{2+} waves were associated with intercellular waves in neighbouring glial cells. Clusters showed a single, consistent trigger cell or site of initiation of intercellular Ca^{2+} waves which propagated over groups of 10 to 200 cells.

Results support a model in which a domain starts by the activation of one or a few trigger cells that subsequently activate the rest of the cells through gap junctions. Spontaneous coactivation of neurons through gap junctions is increasingly recognised as a general property of developing circuits in the mammalian CNS. Nonsynaptic interactions through gap junctions may be a more reliable mechanism for interactions among developing cortical neurons than conventional synaptic transmission. Critical gap junctions are primarily located on dendrites. Spatial cell coupling patterns via gap junctions often resemble future synaptic connection patterns leading to the hypothesis that groups of gap junction coupled cells represents functional cell assemblies that guide formation of synaptic connections and serve as a scaffold or blueprint for columnar structure during development, (Kandler and Katz, 1995). Finally, it has been proposed that the restriction of the spread of coactivation which provides domains with distinct borders may be attributable to filtering of the activation signal when passing through gap junctions.

Learning of Distal Cortical Associations

The development and operation of distal cortical connections have stimulated varied approaches. Shastri and Ajjanagadde (1993), for example, rely on synchronous oscillations as an operational means of achieving integration. Learning is triggered by synchronous activation rather than the co-activation familiar in connectionist networks. This highlights the fundamental technical problem of distinguishing spurious co-variance of widely separated local neural areas from genuine, shared relevance to a specific context in cognitive functioning.

Our approach to establishing distal associations between local cortical areas has much in common with Singer (1994). Synchronization of responses of spatially separated local cortical areas with zero-phase lag is adopted as the definition of association. Simulation studies confirm that synchrony can be established without phase lag by reciprocal connections even if they have slow and variable conduction velocities. This possibility does not exclude common input to local cortical areas as a contributor to the establishment of synchronization. It seems highly probable that common input from bifurcating cortico-cortical projections or diverging sub-cortical-cortical projections will be necessary for the development of co-ordinated temporal patterns of response between widely separated local areas. We adopt synchronous modulation of the excitability of local areas as a means of influencing the probability of them engaging in synchronous firing. The source of the modulation is a proposed mechanism for resolving the problems posed by cortical separation of local areas and the determination of their shared, specific contextual relevance. The mechanism is based on cortico-hippocampal interaction and involves both cortico-cortical and sub-cortical-cortical connections.

Study of the hippocampus has yielded a range of views on its function, (McClelland, 1995). Our mechanism adds to the range by requiring prior recruitment of hippocampal neural structures to new functions through evolution. Burgess et al. (1994) provide an account of the structures concerned and their function in the rat. Extracellular recordings of pyramidal cells in freely moving rats show that most cells in hippocampal regions CA1 and CA3 are 'place' cells which only fire when the rat is in a particular portion of its environment. Burgess at al. (1994) consider place cells in terms of their receptive fields and suggest how place cell firing fields or 'place fields' are constructed. Region CA1 provides many place fields each restricted to a portion of the environment. Extracellular recording data indicate that less than 5% of place cells have multiple place fields and that place fields are single, peaked, smooth functions. Place cells are controlled by sensory cues from different modalities. It is believed that the sensory inputs come from the entorhinal cortex which has access to multimodal sensory information.

Our mechanism involves the assumption that, in human cognition, local cortical areas can be substituted for sensory input from portions of the environment and via entorhinal cortex become represented as place fields in the hippocampus. It is indisputable that co-option of pre-existing features into new features has occurred during the course of evolution, (Raff, 1996). Our proposal conforms to the definition of 'co-option' in that preexisting structures which perform a particular function are presumed to be enlisted for a new use that is selectively advantageous in evolutionary terms.

The nature of the representations of cortical structure available to the hippocampal system is problematic. Consensus favours a form of attenuated representation. In our model, sparse connections from cortical local regions, via layers 2 and 3, connect to entorhinal cortex in a functionally similar manner to environmental, sensory inputs. These connections are reciprocal and impermanent since the flexibility required for the development of new environment representations is, also, necessary for construction of a sequence of distal cortical associations.

Each cortical local region (CLR) currently represented in the hippocampal system is connected to a cell (EC) in the entorhinal layer. The layer of place cells (PCs) in hippocampal area CA1 is arranged in place fields (PFs). Each EC has a reciprocal connection with PCs from 0.5 PFs. Excitation of CLRs produces competitive learning in PFs. Acceptable resolution of competitive learning is defined as emergence of a minimum of two active PCs.

Our approach to establishing shared, specific contextual relevance involves further connections to PCs in PFs. These represent input to the hippocampal system from motivational/emotional activity in the brainstem. The structure of the hippocampal system as a whole features a functional subdivision from input to output in a large number of lamellae. This is exemplified in the mossy fibre connections between the dentate gyrus and CA3 pyramidal cells and the Schaffer collaterals connecting the pyramidal cells of CA3 with those of CA1. In our model individual lamellae between the dentate gyrus and CA3 represent motivational/emotional elements (MEs). The architecture and operation of connectivity between the ME layer and CA3 place cells organised in ME fields (MEFs) is identical to the EC-PF arrangements already described. The same competitive learning process operates in MEFs.

Feedforward connections from the CA3 layer to CA1 place cells represent resolved MEFs. Each MEF has connections with PCs from 0.5 PFs. Critical to establishing shared contextual relevance is the requirement that acceptable resolution of competitive learning in any PF involves inclusion of at least one MEF connected PC in the active group.

Cortical activity and motivational/emotional activity proceed on different time scales. Detection of appropriate associations in our model is achieved by adoption of the same interactive process between astrocytes and neurons involved in the identification and stabilization of local neural areas. This produces local aggregation of the effect of EC-PC input to bring this aspect of PF competitive learning resolution into temporal line with the neuromodulation rate of ME processes. Resolution results in a 200Hz burst which reaches appropriate cortical local regions via reciprocal PC-EC and EC - local region connections (Ylinen et al., 1995). Arrival of a 200 Hz burst provides synchronous modulation of the excitability of appropriate local areas and advances the process of establishing reciprocal connections across the space separating them. The performance characteristics of this architecture are being explored through implementation of a computational model.

Conclusion

To date our attempts to explain the nature of the functional relationship between subsymbolic and symbolic processing in human cognition have involved work on a variety of levels ranging from complex cognitive performance to intracellular mechanisms. We have employed a diverse range of disciplines including neuropsychology, neuroanatomy, neurophysiology, molecular biochemistry and genetics. A range of methodologies and techniques has been used as heterogeneous as brain monitoring, computer modelling and confocal scanning microscopy.

In general, our results support the view of Norman (1991) on the nature of the borderline between subsymbolic and symbolic processing. The complexity and information processing power represented in the intracellular mechanisms of astrocytes and neurons and in neural-glial interaction in local areas suggests that much specific cognitive processing can occur without the need for symbol tokens to provide distal access to other local areas. The principle that the units in the physical symbol system refer to relatively elaborate local subsymbolic processing areas may assist in explaining the effectiveness of symbolic processing in spite of the constraints of slowness, seriality and limited memory size imposed by its dependence on neural networks. Our current work is focussed on proximal and distal relations between local areas and the emergence of more aggregated representations underlying development of non-linguistic symbols such as sensory images and linguistic symbols.

References

Burgess, N., Reece, M. and O'Keefe, J. (1994). A model of hippocampal function. *Neural Networks*, **7**, 1065-1082.

Clark, A., (1989). *Microcognition.* Cambridge: MIT Press.

Finkbeiner, S., (1992). Calcium waves in astrocytes-filling in the gaps. *Neuron.* **8**, 1101-1108.

Finkbeiner, S.M., (1993). Glial Calcium. *Glia.* **9** 83-104.

Hershey, A.D., (1970). Genes and Hereditary Characteristics. *Nature.* 226:697-700.

Johnson, M.H. (1997). *Developmental Cognitive Neuroscience*, Blackwell Publishers Inc., Cambridge, Mass.

Jones, C.L., Lonergan, G.T., and Mainwaring, D.E. (1996). Minimal spanning tree analysis of fungal spore spatial patterns. *Bioimages.* 4 (3), 91-98.

Kandler, K., and Katz, L.C. (1995). Neuronal coupling and uncoupling in the developing nervous system. *Current Opinion in Neurobiology.*, 5, 98-105.

Kater, S.B., and Shibara, A. (1994). The unique and shared properties of neuronal growth cones that enable navigation and specific pathfinding. *Journal of Physiology.* 88, 155-163.

Kater, S.B., and Lipton, S.A. (1995). Neurontransmitter regulation of neuronal outgrowth, plasticity and survival in the year 2001. *Trends Neurosci.* 18, 71-72.

McClelland, J.L., McNaughton, B.L., and O'Reilly, R.C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the success and failures of connectionist models of learning and memory. *Psychological Review* **102(3)**, 419-457.

Newell, A., (1990). *Unified Theories of Cognition*, Harvard University Press, Cambridge, Mass.

Norman, D.A., (1991). Approaches to the study of intelligence. *Artificial Intelligence*, **47**, 327-346.

Raff, R.A. (1996). *The Shape of Life.* University of Chicago Press, Chicago.

Sapp, J. (1987). *Beyond the Gene.* Oxford University Press, New York.

Shastri, L., and Ajjanagadde, V. (1993). From simple associations to systematic reasoning - a connectionist representation of rules, variables and dynamic bindings using temporal synchrony. *Behavioral & Brain Sciences.* **16(3)**, 417-451.

Silberstein, R.B., Schier, M.A., Pipingas, A., Ciorciari, J. & Wood, S. (1990a). Dynamics of the steady state visually evoked potential topography in a visual vigilance task. *Brain Topography*, **3**, 272-273.

Silberstein, R.B., Schier, M.A., Pipingas, A., Ciorciari, J., Wood, S. & Simpson, D.G. (1990b). Steady state visually evoked potential topography associated with a visual vigilance task. *Brain Topography*, **3**, 337-347.

Silberstein, R.B., (1995). Neuromodulation of Neocortical Dynamics. In P.L. Nunez (Ed.), *Neocortical Dynamics and Human EEG Rhythms.* Oxford University Press.

Singer, W. (1994). The Role of Synchrony and Neocortical Processing and Synaptic Plasticity. In E. Domany, J.L. van Hemmen, K. Schulten (Eds.), *Models of Neural Networks II.* Springer-Verlag, New York.

Smith, S.J., (1992). Do astrocytes process neural information? *Progress in Brain Research.* **94**, 119-136.

Smith, S.J., (1994). Neuromodulatory astrocytes. *Current Biology.* **4**, 807-810.

Smolensky, P. (1988). On the proper treatment of connectionism. Behavioural and Brain Sciences, **1**, 1-74.

Wallace, J.G., Silberstein, R.B., Bluff, K. & Pipingas, A., (1994) Semantic transparency, brain monitoring and evaluation of hybrid cognitive architectures. *Connection Science*, **6**, No.1, 43-58.

Wallace, J.G., and Bluff, K. (1995a). Neurons, glia and the borderline between subsymbolic and symbolic processing. In C. Pinto-Ferreira and N.J. Mamede (Eds.). *Progress in Artificial Intelligence.* Berlin, Springer, 201-211.

Wallace, J.G., and Bluff, K. (1995b). Should ANN be ANGN? In J. Mira and F. Sandoval (Eds.). *From Natural to Artificial Neural Computation.* Berlin, Springer, 53-60.

Wallace, J.G., and and Bluff, K. (1997). A role for the hippocampal system in the learning of distal cortical associations. *Proceedings of the International Conference on Neural Networks (ICNN'97),* IEEE Press, Piscataway, NJ, 554-557.

Wickens, J., (1993). *A Theory of the Striatum*; Pergamon Press, Oxford.

Ylinen, A., Bragin, A., Nádasdy, Z., Jandó, Szabó, I., Sik, A., and Buzsáki, G. (1995). Sharp wave-associated high-frequency oscillation (200 Hz) in the intact hippocampus: network and intracellular mechanisms. *The Journal of Neuroscience* **15(1)**, 30-45.