Proceedings of the International Conference New Trends in Cognitive Science, eds. Riegler, A. and Peschl, M.,
pp 139-146, Vienna: Austrian Society of Cognitive Science, 1997

Unguided Categorization, Direct and Symbolic Representation, and Evolution of Cognition in a Modified Connectionist Theory

L. Andrew Coward
7518 Cottonwood Court, Garland, Texas 75044, USA
email: andrew_coward@nt.com

Abstract

Severe constraints apply to the architecture of any system which uses large numbers of components to deliver complex functionality. If not taken into account, these constraints can invalidate conclusions drawn from simulation of cognitive subsystems. The pattern extraction hierarchy is a connectionist architecture which can satisfy these constraints. Within the architecture, functionality is partitioned at a high level in a manner which is consistent with the range of qualitative functionality available from neurons. A system with this architecture can perform unguided categorization resulting in internal representations experienced as mental images. These internal representations function to generate alternative behaviors and select the most appropriate. Functional descriptions of cognition can therefore be created which make use of grounded representations supported within a connectionist architecture. A scenario for the step by step development of human cognition is proposed on the basis that biological brains have the pattern extraction hierarchy architecture. This scenario demonstrates how the ability to generate behaviors appropriate for very complex situations uses a constant succession of mental images and has developed through the interaction of verbal and tool making skills.

Introduction

Symbolist theories (see Newell 1990) have drawn on the computer processing paradigm to argue that cognition can be modeled as the manipulation of sets of arbitrary symbols in accordance with rules operating only on the form of the symbols. Such theories have been criticized for a range of reasons. The most widely discussed issue has been how to account for the categorization which is fundamental to cognition (Harnad 1987) and in particular how to establish a relationship between mental objects which are manipulated and the external environment (the grounding problem, see Harnad 1992). Other problems are how an intelligent agent which views the environment from its own perspective can relate mental objects to behavior (the situatedness problem, see Cognitive Sciences 1993), and in which the environment is subject to change (the frame problem, see Janlert 1987). There are also operational problems such as how to provide non-sequential manipulation of mental objects which will not fail if small sets of individual processing elements are damaged.

Representationalist theories such as Dretske (1987) have explored how mental objects could relate to external objects and behavior, but descriptions are at a general level, and even at that level there is controversy around the capability of such models to account for real behavioral scenarios (see Melnyk 1996, Noordhof 1996, Dretske 1996).

Functionalist theories such as homuncular functionalism (Dennett 1978, Lycan 1987) attempt to break down cognitive systems into meaningful subsystems, but the relationship of such theories to neurophysiology is unclear.

Connectionist theories, because of their apparent resemblance to neurophysiological reality, have been offered as alternatives to overcome these problems. However, resolution is at best

Proceedings of the International Conference New Trends in Cognitive Science, eds. Riegler, A. and Peschl, M., pp 139-146, Vienna: Austrian Society of Cognitive Science, 1997

incomplete, and as Pfeifer (1996) has pointed out, even unsupervised learning schemes such as Kohonen (1988) can only define categories within carefully preselected sets of patterns.

If a system has a combination of high functional complexity and high component count, a requirement to test, modify and build copies of such a system places strong constraints on the system architecture. Specifically, there must be a relatively simple and flexible relationship between elements of functionality at the system level and elements of functionality at the component level. To achieve such a relationship, functionality must be partitioned into conceptually consistent elements at many levels of detail.

The universal dependence on the von Neumann architecture for complex electronic systems results from its ability to employ a consistent 'instruction' concept to partition functionality at all levels of detail from system to device. Coward (1990) argued that evolutionary pressures have applied similar constraints to biological brains, and have resulted in an architecture based on partitioning functionality into pattern extraction/action elements at all levels of detail. This architecture is labeled the Pattern Extraction Hierarchy (PEH) architecture.

In von Neumann systems, the requirement for a simple functional architecture generates an interlocking set of constraints between possible partitioning of high level functionality and possible device capabilities. Any attempt to design intermediate level functionality without reference to both high level and related device level constraints results in systems which are extremely difficult to build, test, or modify.

Although biological brains have minimal resemblance to such electronic systems, Coward (1990) argued that an analogous set of constraints have had a clear impact on the physiology and psychology of biological brains, resulting in such brains having the PEH architecture, and that this architecture therefore provides the basis for understanding cognition in terms of physiology.

Classical connectionism in general has focused on intermediate level cognitive functionality, and therefore requires significant modification to be applicable within the PEH architecture. The purpose of this paper is to demonstrate that this modified connectionism can resolve the categorization and grounding issues, identify and address the weaknesses in models such as Dretske's, and provide an underlying physiological model for a modified version of homuncular functionalism.

Modifications to Connectionism

Classical connectionist theories (for a good critical discussion see Sterelny 1990) model the brain as a network of interconnected nodes. In general the theories share the following properties: 1. Nodes have a range of states (not-firing, firing at different rates) which can be communicated to the rest of the network via connections to other nodes. 2. Connections to other nodes vary in type (inhibitory or stimulatory) and strength. 3. The state of a node is determined by its immediately preceding state and by the sum of the states of the nodes which input to it multiplied by their rate of firing and type of input. The effect of these factors on the state of the node depends on locally determined parameters including its threshold, and is not influenced by global system properties. 4. The information content of the network is coded in the weights of inputs, and learning proceeds by slight modification of weights. 5. There is a layering of nodes, with a layer receiving external outputs separated from a layer generating external outputs by one or more 'hidden' layers. A layer receives the bulk of its inputs from the immediately preceding layer.

The PEH architecture differs in that the functionality of neurons is determined by the requirement for functional separation at high level, subject to the available range of neuron functionality. A neuron can be regarded as being programmed with a pattern, and the firing of the neuron indicates to the system that enough of the programmed pattern is present to require attention in the current circumstances. Rates of firing and modulation of firing rates can also communicate information to the system (such as degree of attention required and correlation between a set of extracted patterns). The extraction of a pattern contributes to recommending an action, which action recommendation may or may not eventually be accepted.

The qualitative functionality of a neuron is defined by the parameters which manage the pattern extracted. One type of parameter is the source of the information which controls changes to extracted pattern and can include: feedback from comparison with an expected result; feedback from pleasure or pain; past firing of the neuron itself; correlated firing of neurons in neighborhood; correlated firing between the neuron and another neuron; or correlated firing within a separate functional group of neurons. A second type of parameter is the nature of the changes to the extracted pattern produced and can include: assignment or removal of a neuron; addition or deletion of an input; correlated addition or deletion of a set of inputs; changes in relative strength of inputs; correlated changes in the strength of a set of inputs; general change in effective input strengths (i.e. threshold change); and changes in sensitivity to other parameters. A third type of parameter is the permanence of changes to the extracted pattern which can include: change only at the time the source of information is present; change for a limited time following the source of information being present; and change for long period following source of information being present

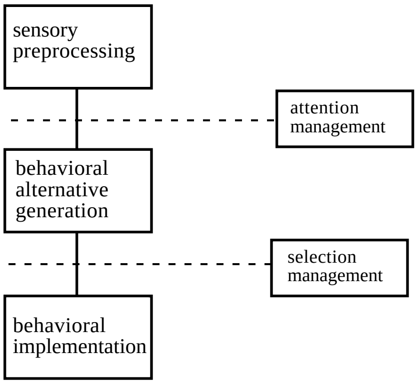

Figure 1. The Pattern Extraction Hierarchy Functional Architecture

Figure 1 shows the major functional separations in the PEH architecture. Sensory preprocessing extracts patterns from raw input. These patterns are stable with respect to external objects (an example would be object color independent of illumination). Attention management allows the set of stable patterns associated with a single object to enter behavioral alternative generation, which generates a set of alternative responses to the object. Selection management chooses one behavior to enter behavioral implementation, where a portfolio of muscle movements is managed to produce a detailed response.

Proceedings of the International Conference New Trends in Cognitive Science, eds. Riegler, A. and Peschl, M.,
pp 139-146, Vienna: Austrian Society of Cognitive Science, 1997

Neuron Type	**Changes to Programmed Pattern**	**Source of Information Controlling Changes**	**Permanence of Changes**
Perceptron	correlated changes in the strength of a set of inputs	feedback from comparison with an expected result	long term
Hebbian	addition of an input; changes in strength of an input	correlated firing between neuron and another neuron	long term
Adaptive Resonance	correlated changes in the strength of a set of inputs	feedback from comparison with an expected result; correlated firing of neurons in neighborhood	long term
Behavioral Alternative Selection	assignment of neuron; set sensitivity to arousal factor for assigned region	correlated firing of neurons in neighborhood	long term
	addition of input	correlated firng of input and neurons in neighborhood	medium and long term
	deletion of inactive inputs (i.e. imprinting)	correlated firing of neurons in neighborhood	long term
	threshold reduction	correlated firing of separate functional group of neurons	short term
	threshold reduction	firing of the neuron itself	short term
Sensory Preprocessing	elimination of neuron	firing history of neuron itself - no firing after initial imprinting	long term
as for BAG neuron without threshold reductions and plus	threshold increase	firing of neuron itself	short term
Behavioral Alternative Generation	addition of inputs	imprinting in BAG function	long term
	change in strength of recently active inputs	feedback from pleasure or pain	long term
	threshold reduction	correlated firing of separate functional group of neurons	short term
	threshold increase	correlated firing of neurons in neighborhood	short term

Table 1. Different Neuron Types

Table 1 summarizes the parameters for the neurons which play the key role in three of the major PEH functional separations. The corresponding parameters for perceptron learning (Rosenblatt 1961), Hebbian learning (Hebb 1949), and adaptive resonance (Carpenter and Grossberg 1988) are also shown. The important point is that in PEH the high level functional partitioning and device functionality are mutually and consistently constrained by the requirement for a simple functional architecture.

The characteristics of the primary neuron required for the behavioral alternative generation function are: 1. Some learning proceeds by neurons instantaneously acquiring a set of inputs, with those inputs and their relative weights remaining fixed from thereon. This imprinting occurs by some neurons being configured with a large number of provisional inputs. At the instant of learning, the subset of inputs which are firing become permanent, and all others are tagged for permanent weakening or deletion. 2. The firing of some neurons can exert a global influence, by modifying thresholds throughout a defined region of the network. 3. There is a clear functional partitioning between a region which generates alternative behavioral recommendations and in which learning occurs by imprinting and a region in which a behavior is selected from the alternatives and in which learning is by gradual adjustment to weights. 4. There are well defined routes for feedback from specific layers to specific earlier layers.

Electronic simulation has demonstrated that a PEH architecture can organize its experience of objects into categories without restrictions, guidance, or feedback, and that pleasure and pain feedback can function to link the categories with appropriate behavior (Coward 1996).

Cognitive Modeling within the Pattern Extraction Hierarchy Architecture

Neurons are arranged in a hierarchy of pattern extraction, with each level in the hierarchy receiving inputs mainly from the preceding level, and the patterns at each level therefore being combinations of patterns at the previous level. Patterns close to sensory input can be interpreted as a sensory characteristics of objects, while deeper in the hierarchy patterns acquire the function of behavioral recommendations ('*this* is present, therefore do *that*'). Note that this hierarchical arrangement, as illustrated in figure 2, is not a functional architecture. It is rather an information or data architecture indicating how information from the environment is partitioned within the system. This data architecture must of course be consistent with the functional architecture.

The perception of an object generates an assembly of neuron firing linking patterns close to sensory input with patterns controlling action. Such an assembly extends across a region generating alternative behavioral recommendations and a region selecting a behavior from the alternatives. In the behavioral generating region, the assembly is made up of parallel cascades of neuron firing through domains generating different types of action recommendation. Thus a cascade through one domain corresponds with an aggressive behavioral recommendation, the cascade through another domain with a food seeking behavioral recommendation etc. These cascades exit the generation region and compete in the selection region for control of behavior.

Some behavioral recommendations, if accepted, act back on a domain of the generation region to lower neuron thresholds and increase the probability of a strong behavioral recommendation from that domain for any object. Thus anger lowers the thresholds of neurons in the aggressive domain, hunger the thresholds in the food seeking domain etc.

Competition between cascades occurs within the behavioral alternative selection management function by cross inhibition. Pleasure and pain affect the connection strengths of recently active connections in this region, and affect the ability of similar cascades to gain control of behavior in the future.

The specific behavioral recommendation generated within a domain depends on the category to which the perceived object is assigned as discussed in the next section. In one domain, speech behavioral recommendations are developed, associating spoken words with activation of the neurons which most often fire when the object is perceived. Speech has thus created the possibility of feedback, and an internalized version of this feedback route drives a constant succession of mental images used to extend the range of memory searched for guidance for current behavior as discussed in sections 5 and 6.

Creation of Categories and Competition for Control of Behavior

A simplified version of the process for creation of categories in the PEH architecture can be understood by reference to figure 2. The inputs represent object characteristics which have been derived from sensory systems to be object specific (e.g. object color independent of illumination, object size independent of distance etc.). Such characteristics can themselves be extracted from raw sensory input by the same category creation process to be described. The presence of a characteristic in a perceived object is represented by the activation (firing) of the corresponding input.

Initially all nodes have large numbers of randomly selected inputs. a-layer nodes have inputs from characteristics, b-layer nodes from a-layer nodes etc. All inputs have equal weight, variable weight can be modeled by duplication. Figure 1 represents a single module, many such modules are available to learn different categories.

Suppose an object is perceived which is unlike any previous objects. A module is assigned to learn a new category of which the object will be the first member. The thresholds of all the nodes in the module are lowered until some nodes are firing in the output c-layer. All the nodes which fire in the module are imprinted with the input configuration which fired them: inputs which did not fire are tagged for deletion, and the node threshold is set at or slightly below the new total number of inputs. If the same characteristics were to be extracted from an object in the future, the same nodes would fire again, a permanent memory trace has been instantaneously created. The imprinted nodes can be regarded as patterns which have been extracted from the object, which may repeat in the future.

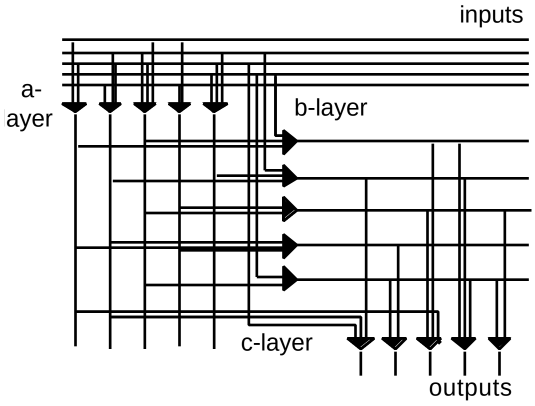

Figure 2. A Simplified Category Module

Once a module has been imprinted, its mode of operation changes. Each module has an internal 'similarity criterion' which is a percentage of the total number of imprinted nodes in the b-layer. When another object is perceived, its characteristics are available to all imprinted modules, and can therefore generate node firing within those modules. If the number of firing nodes in each module is below the similarity criterion for that module, the object is perceived to be different from any previously perceived object, and a new category module is imprinted. Otherwise, the module within which the ratio of firing nodes to similarity criterion is largest is deemed to be the category to which the object belongs. If there is no firing in the output layer, nodes are imprinted until an output results. Any nodes imprinted in the middle layer of course affect the similarity criterion.

Actual categorization algorithms have additional features to reduce category duplication, merge duplicates, etc. and can effectively create internal categories (Coward 1996). The critical restriction is that the algorithms can only use information derived from sensory input.

Any output from a category is a signal that the perceived object belongs to the category, and in general the specific combination of output nodes which fire is unique to the object and occasion on which the object was perceived.

To express this process in pattern extraction terms, a category is created by extracting and imprinting a set of patterns from a single object of a hitherto unfamiliar type. Objects are assigned to an existing category if they contain a significant subset of the patterns extracted from objects previously assigned to the category, and patterns extracted from the new object are added to the category definition. Because patterns are imprinted from supersets, actual patterns can include a wide range of characteristics which are not directly associated with the object, such as body conditions and environmental conditions which may or may not prove useful in generating behavior then or later. Hence some patterns in the output layer are unique to such other conditions and allow the selection of behavior to be influenced accordingly. In addition, the cross connection to support associative memory is created.

The way in which this information is used can be understood by reference to figure 3. An intelligent agent has domains of behavior defined genetically and by experience. Examples are a fearful domain which generates behavior appropriate if the agent is weak, an aggressive domain appropriate if the agent needs to defend self or property, a friendly domain appropriate if the agent needs to gain cooperation. The presence of the need condition can itself be extracted from sensory input, including input of internal body conditions. Categories for all different types of objects are configured in all domains, although categories will receive neuron resources for recording additional objects in proportion to the experienced relevance of the category to the behavior domain. The system meaning of an output from the 'dog' category in the friendly domain is an action recommendation to be friendly towards the perceived dog. Note that although the 'dog' categories are superficially duplicated, the categories have different functional roles, and may contain different pattern elements. This category 'duplication' is the reason local damage does not remove event memory, but may cause personality shift, as discussed in Coward 1990. The specific set of firing nodes which make up that output represent system meanings such as 'pat the dog', 'smile at the dog', or 'greet the dog'. The action alternatives from the different domains are directed to the action selection structure. There, the action alternatives compete by cross inhibition, and a successful output from, say, the aggressive pipe allows the specific node firing from the dog category in the aggressive domain to proceed to motor control. Nodes in the action selection structure have input weights which vary in response to pleasure and pain if the input is active in the pipe which generated the output. Interpipe connections are inhibitive and are strengthened by pain and weakened by pleasure. Intrapipe connections are stimulative and are strengthened by pleasure and weakened by pain. A perceived need to defend self or property generates anger, which lowers the threshold of all nodes in the aggressive generation domain, thus increasing the chance of a strong, accepted action recommendation from that domain.

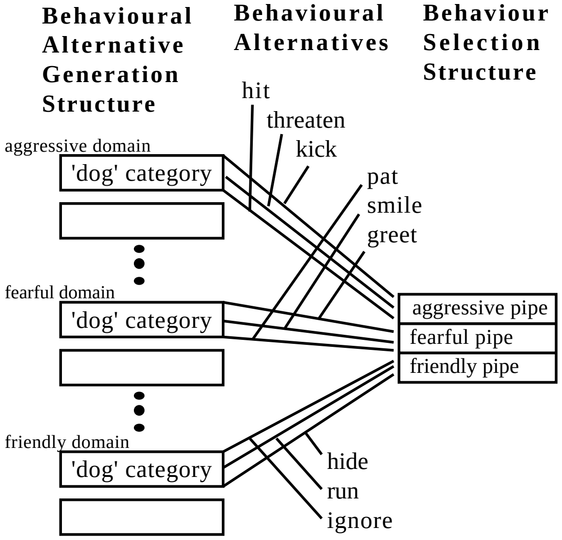

Figure 3. Behavior Alternative Generation and Selection

A perceived object is thus represented mentally by an assembly of activated patterns extending across descriptively equivalent categories in different behavioral domains. The activated patterns range from simple sensory combinations to complex combinations of combinations of combinations etc. which from a system point of view represent behavioral recommendations. For example, a particular combination of dog characteristics *plus* body motion (tail wagging) *plus* my feeling in need of exercise recommends being friendly to the dog.

The 'dog' example is a physical object category. Other categories record sets of patterns extracted from relationships between objects and generate behavioral recommendations. Such categories could use output from physical output categories as part of their input. In general a hierarchy of categories develops heuristically, with output from a category directed both to other categories within the generation region and externally to the selection region.

Dretske (1988) has proposed a functional theory of behavior in which beliefs play a central role. In Dretske's model, a belief C causes an action M. C functions to represent Fs in the world, and the presence of Fs will therefore cause C to be tokened, causing M. The reason the tokening of C causes M rather than N is that the system containing C is rewarded for producing M when and only when there are Fs present. C tokenings represent Fs as present, and are therefore present when and only when Fs are present. The system can therefore recruit C tokenings to cause M.

Figure 4 gives the functional description at an equivalent level of detail in the PEH architecture. Object F(i) of type F activates representations in categories C^x, C^y, and C^z in the behavioral domains x, y and z. A behavioral recommendation $C^x(i)$ of type x but specific to object F(i) is generated by category C^x and so on. The three behavioral recommendations compete for control of action, and the winner $C^y(i)$ drives a behavior which is specified by the information contained in $C^y(i)$.

The successful or unsuccessful outcome of the behavior acts on the behavioral selection structure through pleasure and pain to increase or decrease the ability of behavioral recommendations with similar content to $C^y(i)$ to gain control of behavior in the future. Dretske's model has difficulties in providing a plausible account of some scenarios, and Dretske himself comments (1996) that "the emergence of a novel (but nonetheless appropriate) response out of pre-existing control structures [is] a mystery". Melryk (1996) discusses the scenario of his imagined first encounter with a three-legged dog, provoking him to say to his companion "Look, a three-legged dog !" If he had never before had dealings with a three-legged dog, he found it difficult to see how a belief could have been recruited to cause speech. In PEH, the three-legged dog activated representations in a range of 'dog' categories, including one which generates 'talk about dog' type recommendations. Three-leggedness will be a novel element requiring pattern imprinting to get an output in the 'dog' categories. Hence a 'talk about dog' recommendation contains substantial information referring to three-leggedness which if the recommendation is accepted will be reflected in the actual words spoken.

External and Internal Symbolic Representation

The patterns activated when dogs are seen form a hierarchy from simple to complex. For useful communication, the spoken word "dog" must generate a dog-representation in the mind of the hearer. Learning the word dog is a process of hearing the sound at the same time as having a sensory driven representation activated. A Hebbian mechanism results in the output of a category 'dog spoken' activating the intermediate level patterns which fire most frequently when a sensory driven dog representation is activated. Thus the spoken word "dog" can generate a similar range of action recommendations to visual perception of a dog.

A possibility for feedback has been created. Seeing a dog can generate a recommendation 'say "dog" ' which if accepted could maintain a dog representation active in the speaker's mind. This mechanism keeps an unknown telephone number in mind long enough to dial. The feedback loop can become internalized, leading to a self sustaining sequence of images. Suppose I see a cat, and while the patterns extracted from the cat are still active I see a picture of a dog. The simultaneous activation of dog and cat patterns leads to activation of some of the more complex patterns which were imprinted when watching a dog chase a cat. These patterns by feedback generate a more extensive set of dog chasing cat patterns at the less complex level, including patterns referring to the cat escaping up a tree. By feedback a more extensive set of tree patterns are activated, and so on. As argued in Coward 1990, this constant succession of images driven by feedback has the function of extending the range of recorded individual experience which can be incorporated in generating behavioral alternatives, and is experienced as consciousness.

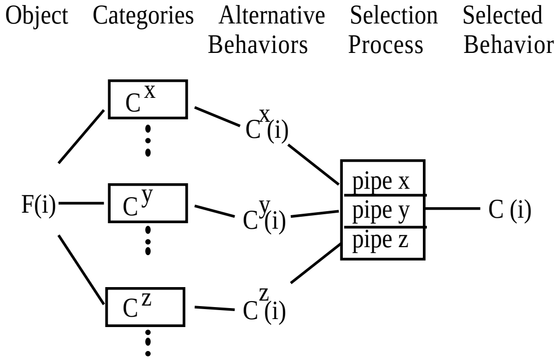

Figure 4. Functional Model for Behavior Generation

Dennett (1978) and Lycan (1987) argue that general, flexible intelligence can be broken down into lesser and more specialized intelligences which they label homunculi, with each homunculus being an ensemble of more specialized homunculi. Dennett (1991) goes on to argue that consciousness derives from speech homunculi. The categories which develop in the PEH architecture functionally resemble the homunculi of Dennett and Lycan, and the categories which generate symbolic representations resemble Dennett's speech homunculi, with the difference being the feedback mechanism which activates a subset of the representation of an actual sensory experience.

The Function and Evolution of Cognition

The following simplified account of how human cognitive capabilities could have developed is intended to demonstrate how the PEH architecture makes it possible to identify a series of small, plausible physiological changes, each associated with a behavioral advantage which would result in selection pressure, and that such a series leads to human cognition. The simplified account is based on capabilities defined within a systematic theory of cognition based on the PEH architecture (Coward 1997) and is strongly influenced at the behavioral level by the ideas of Jaynes (1976) and Johnson (1987).

There are three types of physiological capability which in the PEH architecture underlie the unique human cognitive system. The first capability is the development of a domain in which the behavioral recommendation of a category is to speak a sound which in the brain of a hearer will activate a subset of the patterns which are inputs to an equivalent category to the source category in the brain of the speaker. This is the verbal capability, which developed into speech as the physiological capability developed to support a hierarchy of interacting categories. The second capability is the development of a domain in which the behavioral recommendation of a category is to activate a subset of the patterns that are inputs to the category. The physiological development of this capability was the detailed information content of the category output controlling the distance feedback reached towards sensory input, with the result that an active mental representation of an object could generate a greater detail representation of a component of the object. For example, a dog could be represented mainly by relatively complex, abstract patterns, but feedback activating patterns at a relatively simple sensory level could generate a set of action recommendations from feedback categories representing tail of dog, head of dog etc. Accepting one of these recommendations would activate a strong representation of the dog 'component'. The third capability is the maintenance of more than one independent assembly in a relatively active state at the same time, so that secondary assemblies are generated from overlap between several assemblies. This overlap required a physiological solution to avoiding confusion between different assemblies.

The functions which have developed on top of these capabilities include repeating words aloud to self to keep a representation active in the mind of the speaker; the development of categories which generate representations of self which can be used to develop additional action recommendations (Jaynes 1976); and the development of categories which extract consistency or inconsistency between other categories. The improved behavioral capabilities which have provided the evolutionary selection pressure towards increased cognitive capability are in the areas of tool making; accepting commands; generating appropriate behavior for complex social situations; and developing organized, consistent belief systems to guide behavior.

Tool making requires the generation of behavioral recommendations to chip, flake and carve appropriately. A mental representation of the tool is required to generate such behavioral recommendations. Early tool making depended on visual input from an existing model to generate the required mental representation. The development of words, probably originally as warnings or indications to generate a mental representation and trigger behavior, opened the possibility of a tool maker repeating the name of a tool aloud to keep the mental representation active. Internal feedback has the advantage of less susceptibility to interruption. Once controlling the level of detail at which feedback operated became possible, more detailed control of mental representations of object components made it possible to create more sophisticated tools, and in due course carved or painted figures.

The ability to accept a verbal command requires the sound to generate a representation which in turn generates appropriate behavioral recommendations. To make it possible to continue to carry out a commanded task over an extended period of time, the representation must be reactivated if interrupted. Repeating a command to self extended the time a command could be effective, and internal feedback extended it further. Maintaining assemblies for a time in a partially active state allows a chance sensory input to fully reactivate the assembly, 'reminding' the person of their task.

As the complexity of social interaction increased, the range of individual specific memory which held possibly relevant patterns for generating a behavior appropriate to a situation increased. The development of self representation categories plus extended assembly overlap allowed more extensive searches of memory. For example, consider a situation in which a person X is interacting with three other people A, B, and C. A and B are people familiar to X, and C is a familiar type, a tax collector. At the behavioral level, suppose that A asks X for a favor. X has performed such favors for A in the past, and has been pleased with the results, but has refused such favors to B. X is there to collect tax from A, B, and C. A simple behavioral response based on past pleasure would be to agree. A more satisfactory but complex response would be to refuse, claiming no resources and recalling the refusal to B, but focusing the reply on tax collector C. Later, the refusal could be explained to A by reference to C. A quarrel with B is also avoided.

The description of this scenario in the PEH architecture terms is as follows. Assemblies representing A, B, and C are active to some degree in the brain of X. The request from A initially makes the A representation the currently active assembly, then the verbal request activates an assembly representing the requested action taking place. Overlap of the A assembly and the requested action assembly activates patterns imprinted in the course of generating the earlier accepted behavioral recommendation to grant the favor. Feedback generates an active representation of A being granted the request by self. Overlap between the representation of self granting the request and the assembly representing C activates patterns imprinted on previous occasions when tax collectors saw evidence of resources, followed by pain. The new representation therefore inhibits the ability of the 'grant the favor' recommendation to gain control of action. Patterns imprinted when B was refused a favor combine with the recommendation to grant the favor to generate output from an inconsistency module. In total, by additional imprinting, an action recommendation is synthesized incorporating patterns imprinted when claiming no resources, refusing a favor to B, and feeling good towards A.

The enhancement of internal feedback, multiple active assemblies, and consistency categories under the pressure for better behavioral responses to complex situations reached a critical point at which the development of organized, consistent belief systems such as science and philosophy became possible about 2500 years ago, and the classical Greeks were the first to exploit the capability in this manner.

Conclusions

As Harnad (1987) has pointed out, the "problem of how categories are learned and represented is ... central to contemporary cognitive science". PEH architecture provides a detailed model for how the categorization works, including "how ...[to] arrive at successful generalizations from finite samples of instances" and "the relation between our words and the things they describe"

On the basis of this modified connectionist model for unguided categorization, effective functional models for a wide range of behavioral phenomena can be developed, including a speculative but plausible process by which human cognition evolved.

References

Carpenter, G.A. and Grossberg, S. (1988), The ART of Adaptive Pattern Recognition by a Self-Organizing Neural Network, *IEEE Computer,* **3**, 77-88.

Coward, L. A. (1990), *Pattern Thinking: A Physiologically Based Brain Architecture*, New York: Praeger.

Coward, L.A. (1996), 'Understanding of Consciousness through Application of Techniques for Design of Extremely Complex Electronic Systems' *Towards a Science of Consciousnes*s, Tucson, Arizona.

Coward, L.A. (1997), 'An Integrated Theory of the Phenomenal Content of Consciousness' to be published.

Dennett, D.C. (1978), *Brainstorms: philisophical essays on mind and psychology*, Montgomery: Bradford Books.

Dennett, D.C. (1991), *Consciousness Explained*, Boston: Little John.

Dretske, F. (1988), *Explaining Behavior: Reasons in a World of Causes*, Cambridge: MIT Press.

Dretske, F. (1996), How Reasons Explain Behavior: Reply to Melnyk and Noordhof, *Mind and Language* **11**, 2, 223-29.

Harnad, S. (1987), Category Induction and Representation, in *Categorical Perception: The Groundwork of Cognition* (ed. Harnad, S), New York: Cambridge University Press.

Harnad, S. (1992), Connecting Object And Symbol In Modeling Cognition, in *Connectionism in Context* (ed. Clark, A and Lutz, R.), Berlin: Springer-Verlag.

Hebb, D. C. (1949) *The Organization of Behaviour*, New York: Wiley.

Janlert, L.E. (1987), Modeling Change - The Frame Problem, in *The Robot's Dilemma: The Frame Problem in Artificial Intelligence* (ed. Pylyshyn, Z.W.), Norwood: Ablex.

Jaynes, J (1976), *The Origin of Consciousness in the Breakdown of the Bicameral Mind,* Boston: Harvard.

Johnson, D.M. (1987), The Greek Origins of Belief, *American Philosophical Quarterly* **24**, 4, 319-27.

Kohonen, T. (1988), *Self Organization and Associative Memory*, Berlin: Springer-Verlag.

Lycan (1981) Towards a Homuncular Theory of Believing, *Cognition and Brain Theory*, 4.

Melnyk, A. (1996), The Prospects for Dretske's Account of the Explanatory Role of Belief, *Mind and Language* **11**, 2, 203-15.

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

Noordhof, P. (1996), Accidental Associations, Local Potency, and a Dilemma for Dretske, *Mind and Language* **11**, 2, 216-22.

Pfeifer (1996), Symbols, Patterns and Behavior: Beyond the Information Processing Metaphor, in *Encyclopedia of Microcomputers*, 17, 253-75, New York: Marcel Dekker.

Rosenblatt, F. (1961) '*Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*' Washington: Spartan.

Sterelny, K (1990), *The Representational Theory of Mind*, Oxford: Basil Blackwell.