] Wang, Y. and Sengupta, B., Performance analysis of a feedback congestion control policy under nonnegligible propagation delay, *Proc. ACM SIGCOM* (1991).

Bekakos M. P., and Sambandham, M. (eds) Dynamic Publishers, Atlanta, Georgia, US

fo Lab, Dept. of General Sciences, Agricultural
iversity of Athens,
hens, Greece
ept. of Informatics, Athens University of Economics
l Business, Athens, Greece

Computational Methods and Neural Networks (2000) 235-264

EMBODIMENTS OF THEORIES OF MIND: A REVIEW AND A COMPARISON

PETROS A. M. GELEPITHIS[1]

This article reviews the *basic elements* of four major programmes of research on intelligent architectures and their underlying theories of mind, and briefly compares them (section five) against a set of design requirements for general intelligence.

The rationale for my choice is threefold. First, each and every programme draws upon at least two of the disciplines constituting cognitive science and aims at covering aspects of both human and machine cognition. Second, in complement to the first one, each programme draws, primarily, upon a different discipline. Specifically, in section one, Soar's development relies heavily on AI practice. In section two, ACT* is greatly influenced by psychological considerations and findings. In section three, Edelman's Theory of Neuronal Group Selection is based on evolutionary and neurophysiological findings and principles. Finally, in section four, Arkin's autonomous robot architecture, goes back to the cybernetic roots of psychological and neuroscientific studies and illustrates nicely their impact on the design of intelligent robots. Finally, the programmes, viewed collectively, illustrate well the two major contrasting weltanschaungs underlying the development of such theories and their implementations.

Of course, choice implies exclusion and I would like here to say a few words about two additional research programmes which had to be left out of this review. First, is *The Society of Mind*, an ambitious proposal by Marvin Minsky [43] to explain the workings of the human mind. His approach is synthetic assuming that an intelligent entity can be built "from many little parts, each mindless by itself." (ibid. p 17). These parts or particles or processes as sometimes Minsky refers to are called agents. Unfortunately, the numerous agents, which are rather briefly introduced have not really been integrated. This results into an interesting reading but not a unified theory of cognition and definitely not an embodied theory of mind. Second, it is Holland *et al.*'s [37] work which may be seen as a framework for the development of a theory of mind that would posit induction as its backbone.

1. SOAR AS A UNIFIED THEORY OF COGNITION

Soar is an architecture for general intelligence. It started being developed in the early 1980's supporting a considerable number of problem solving methods. As its repertoire was extended, the original methodological commitment of using human behaviour as a guideline for its development was turned into the objective of closely modelling human behaviour. Most recently it has been proposed (Newell [45]) as a candidate architecture embodying a unified theory of cognition. This proposal aims to provide a detailed theoretical framework for understanding cognition and its key notions are knowledge, representation, search, computation, symbols and architecture. Within that framework, the notion of 'intelligence' (and on that basis of AI itself) is defined in terms of the notion of the *knowledge level*. The rest of this section introduces the notion of architecture and Soar itself and briefly comments on the latter's candidacy as a unified theory of cognition. We start with Newell's definition of architecture intending to cover both human and computer architectures[1]:

> "An architecture is the fixed structure that realizes a symbol system." (Newell [45], p. 80).

Firstly, a clarification concerning the term 'fixed'. This must not be taken to mean immutable. As Newell repeatedly emphasised, fixed structure means changing relatively slowly. More importantly, "the fixed structure that realizes a symbol system" does not refer to all the structure that may realise a symbol system. Some parts of the overall structure have to change, even momentarily, for behaviour to occur. In Newell's terminology, "[such] momentarily varying structures are labeled the memory content." (ibid. p 78).

Newell's conception of the human cognitive architecture is based on three more or less reasonably assumed constraints. First, human cognitive architecture supports mind-like behaviour. This follows from combining the Physical Symbol System Hypothesis (PSSH) with the above definition of architecture. Second, human cognitive architecture is realised in "neural technology" (i.e., neural circuits operating at approximately 10ms). Third, given that the minimum time

[1] To be noted that this is not a universally accepted conception, but it might well become. Alternatives include Pylyshyn ([49], p. 191) defines the notion of 'cognitive architecture' as the unique organisation level at which the states being processed receive a cognitive interpretation. For Simon ([54], p. 38) "Architecture is simply a high-level description of the brain in terms of its information-processing properties." Changeux and Dehaene ([17], p. 66) view the notion of human architecture differently. It is equated with the notion of a physical structure which, along with its associated functions, constitute a particular level of organisation in biological systems.

required for cognitive behaviour to occur is of the rough order of seconds, minimal cognitive behaviour should be realised approximately within 100x10ms; this assumed constraint has been variously called "the 100 program-step limit" (Feldman and Ballard [24]), or the "real-time constraint on cognition (Newell [45], p. 130).

Soar's overall architecture may be seen in Figure 1. As with all complex systems

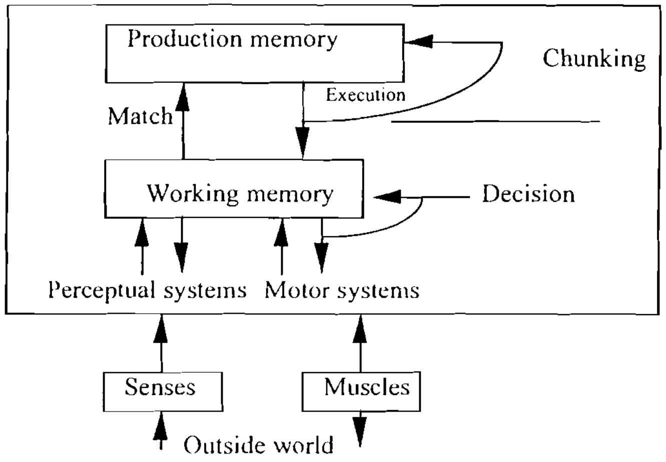

Figure 1. Overview of the Soar cognitive architecture (Newell, *et al.* [46], p. 111)

Soar can be described at several levels. The ones so far described are the problem space and symbol levels (Laird *et al.* [40]), and the knowledge level (Rosenbloom *et al.* [51]). Here we only use elements of these descriptions to the extent that they contribute to our stated goals. It should be noticed that Soar's perceptual and action subsystems are still nascent. In what follows we concentrate on Soar's central cognition. There are six main characteristics.

First, in Soar the overall representational framework is provided by problem spaces. Each and every cognitive task is represented by a problem space; this uniformity has been termed the *problem space hypothesis* (Newell [45], p. 163). To accomplish a given task Soar formulates a problem space which has to be appropriately searched. This search requires two familiar types of search: problem-space search, and knowledge search. The former, ubiquitously used in AI from its very beginning, consists of applying operators to a state to create new states. The latter, come to fore with the advent of the knowledge-based approach to AI, consists of finding the appropriate knowledge required to be used and its relevance is determined by the current state and problem-space created. Knowledge search is more widely known as heuristic search (although some may argue that the latter is a special case of the former). The Soar architecture is built to accomplish four primitive functions: (i) selection of appropriate problem space; (ii) selection of a state from those available; (iii) selection of an operator; and (iv) application of the operator to obtain new state.

Soar's second feature, the *memory uniformity assumption*, "is that all of its long-term memory is constructed as a single production system." (ibid. p. 164). All knowledge, both for guiding the problem solving and for implementing the operators resides in the system's production (or recognition) memory (see Figure 1). This is a highly contentious hypothesis; see, for example, ACT*'s memory hypotheses in Section 2 and Tulving [57]. In contrast to the usual characteristics of a production system, Soar: (i) cannot modify the elements of its working memory or take any other actions; and (ii) has no conflict resolution strategy. This makes Soar very similar to a content-addressed memory system.

Its third main characteristic concerns the *representation mechanism* used. It is the oldest and most widely used AI representational scheme: O-A-V triplets. In Soar both the Attributes and Values may be other Objects, but there are no default values or automatic inheritance procedures.

Soar's fourth characteristic, and third uniform feature, is called the *decision cycle*. It consists of two phases: elaboration and decision. During a given elaboration phase all knowledge that is available about the current situation is acquired. At the end of it a set of preferences[2] have been accumulated in Soar's working memory. As the reader may already have anticipated, (see feature one above), the elaboration phase reaches a quiescence only after *all* productions whose conditions are satisfied have been fired and their corresponding working-memory elements have been added into the working memory. After quiescence has occurred the decision procedure starts operating on the set of accumulated preferences and is based on the following preference language: accept/reject, better/indifferent/worse. This operation takes place within a particular context[3] stack consisting of a goal-subgoal hierarchy starting with G_0. It should be noticed that the decision procedure: (i) uses *all* preferences available in the working memory, regardless of the time they were produced by an elaboration phase, and *none* of the non preference elements present in it.

[2] Term used in Soar; it refers to what action is to be taken.

[3] A context is a quadruple specifying a goal, a problem space, a state, and an operator; symbol (G, P, S, O)

If an unequivocal choice can be made then that choice is being made and Soar begins a new decision cycle. Nevertheless, more often than not, given that the set of preferences produced is arbitrary (it may be even empty), such a choice can not be made. Each and every case that produces no clear choice is termed an *impasse*. There are four possible types: (i) a tie impasse, whenever a discrimination can not be made from among a set of alternatives; (ii) a no-change impasse, whenever no choices are available ("Either they have all been rejected or nothing was ever declared to be acceptable." Newell [45], p. 175); (iii) a reject impasse, whenever rejection of a decision already made occurs as a preference and, therefore, a what-to-do next query arises; and (iv) a conflict impasse, whenever conflicting preferences are produced by the productions fired (e.g., $\text{preference}_1 : O_1$ better than O_2; and $\text{preference}_2 : O_2$ better than O_1).

To resolve an impasse, and then only, Soar creates a subgoal. That is a goal context consisting of a goal and slots for a problem space, state, and operator, i.e., (G, -, -, -). Given a new subgoal and the set of preferences in working memory Soar

> "might first choose a problem space (g, p, -, -), then an initial state (g, p, s, -), then an operator (g, p, s, o). Installing the operator releases the knowledge that implements the operator and produces a new state (s') in working memory, which might then be chosen on the next decision cycle to be the next state (g, p, s', -). Then the next operator (g, p, s', o') is chosen, then the next state and so on, to move through the problem space to attain the desired state." (ibid. p 173).

Of course at any point several operators, instead of one, may be proposed, a tie impasse be created, and a second problem space set up to make a selection. Although an operator can now be selected, it may be possible that not enough is known for the required productions to fire within the elaboration phase. This leads to a third problem space, and possibly into a hierarchy of problem spaces bottoming out when there is sufficient knowledge in a space to enable it to act.

The final main feature of Soar's architecture is called *chunking* and, according to Newell, is learning from experience; it may also be seen as a form of explanation-based learning. The basic idea is that whenever some results have been obtained an 'appropriate' production P is created and added to Soar's LTM. Assuming P: $C_1, C_2, \ldots, C_m \dashrightarrow A_1, \ldots, A_n$. Then, P is appropriate if and only if:

$A_1, \ldots, A_n$ -- obtained results.

$C_1, C_2, \ldots, C_m$ = the working memory elements existed *before* problem solving started and used to produce the obtained results. An appropriate chunking production is called a chunk. Chunking may drastically reduce search whenever Soar faces *exactly* the same situation again. This exactitude requirement needs a bit of clarification. Newell (ibid. p 189) writes:

> "But the *exact same situation* means that all the elements in working memory at the time of the left-hand solid vertical [i.e., an impassse] are the same. Yet, only elements that were relevant to producing the result were picked out in building the chunk. All the rest are irrelevant. The chunk will not test for any of these"

This of course means that, in our example chunk, any situation containing the elements (A, B, C), will evoke the chunk and therefore may lead to over generalisation. This may happen during human learning too and there seems to be no general solution to it.

Chunking, usually, improves the performance of Soar primarily by reducing the number of steps required. Nevertheless, this reduction is not always sufficient because of the existence of expensive chunks, i.e., "chunks that require a large amount of effort in accessing them from the knowledge base." (Tamble *et al.* [56]). They have shown that, by restricting the expressive power of Soar's production system, language expensive chunks disappear.

With chunking introduced we have completed our description of Soar's central cognitive architecture. For a system to be capable of exhibiting intelligence in the real world though, perceptual and motor processing is a must. The next couple of paragraphs briefly introduce the basic theoretical, (not, presently, operational), ideas of Soar's perceptual and motor systems, as well as extensions to some of its basic ideas already introduced.

The first idea concerns Soar's working memory. In the "total cognitive system" (Newell's term for the perceptual, motor, and central cognitive subsystems) its role is augmented to include that of buffer memory, providing thus the cognitive system with the time required for its processing. Although some buffering should almost certainly exist the necessity of this assumption ("buffering must occur" Newell [45], p. 195) is almost certainly mistaken.

The second basic idea concerns productions. There are two additional sets of productions: the first, called encoding productions, puts "the elements into a form to be considered by central cognition."; the second, called decoding productions, "provide the expansion and translation of the motor commands produced by the cognitive system into the form used by the motor system." (ibid. p. 197). The only difference is that for the encoding and decoding productions there is no

such thing as the decision cycle; they are entirely free of the goal context stack. Whatever their function *all* productions arise from chunking. This is the second central assumption of the total cognitive system. Its importance, if true, would provide an integrated basis for perceptual and motor learning and their relation to deliberate cognitive learning.

2. ACT* AS A COGNITIVE ARCHITECTURE

This section provides a brief introduction to the work of J. R. Anderson and his team on cognitive architectures and, in particular, ACT*. Specifically, we present the fundamental assumptions of Anderson's theory of mind and his general framework for ACT*, introduce the basic mechanisms of the latter, and indicate its applicability. ACT* started being developed back in the early seventies (Anderson and Bower [6]), and is the basis for a number of theories, remedial systems, and extensions proposed within the so called ACT framework (e.g., ACTE (Anderson [1]); PUPS (Anderson and Thompson [7]); ACT-R Anderson [5]).

ACT* (Anderson [2], p. *ix*) "is a theory of *cognitive architecture*-that is, a theory of the basic principles of operation built into the cognitive system." It should be noticed that Anderson's later work on human cognition put emphasis on a different methodology ("a different level of analysis") based on his principle of rationality and considered the notion of a cognitive architecture as, essentially, subservient to that of a rational analysis (Anderson [3, 4]). He nevertheless, hoped that ACT* and rational analysis would share some general presuppositions concerning the workings of the human mind and that an update of ACT* compatible with his rational analysis approach to human cognition would be possible. This indeed happened in the ACT-R architecture (Anderson [5]). Although "[T]he ACT-R theory is the best specified of the ACT theories.", the majority of its assumptions are identical or slight generalisations of the ACT* theory. On that basis and the fact that ACT* is still more widely known we have confined out attention to it rather than ACT-R.[4]

Two assumptions are at the core of his theory. First, that higher-level cognition (what Anderson seems to equate with thought) constitutes a unitary human system, i.e., all higher-level cognitive functions are explainable by a single set of

[4] The reader is referred to Anderson [5] for a detailed discussion of ACT-R's theoretical assumptions, (including a comparison with ACT*), and a discussion on issues related to rule-based systems (rules being claimed to constitute the building blocks of cognitive skills).

principles[5]. Second, that production systems provide the appropriate computational architecture for modelling, at least aspects of, human thought.

The general framework for the ACT production system consists of three types of memory: declarative, procedural and working and six processes: encoding and performance which link the working memory, and hence the overall system, to the outside world; storage and retrieval relating working with declarative memory; and match and execution linking working with procedural memory.

Working memory and the processes associated with it constitute the core of the system. ACT's working memory contain the currently accessible information of the system. This information may be: (i) information retrieved from the system's declarative memory; (ii) information deposited through the system's encoding process; and (iii) information deposited into the working memory through the execution process. All working memory information is characterised by two things. First, it is in a active state, and second, it is in declarative form (either permanent or temporary). As Anderson ([2], p. 20) writes:

> "The *storage* process can create permanent records in declarative memory of the contents of working memory and can increase the strength of existing records in declarative memory."

Specifically,

> "When a temporary cognitive unit is created and there is not a permanent copy of it, there is probability p that a permanent copy will be created. If there is a permanent copy, its strength will be increased one unit." (ibid. p. 22).

The rest of the processes are self-explanatory or standard productions systems terminology.

Declarative information (Anderson calls it knowledge, without distinguishing between the two), comes into chunks called cognitive units. Following Wallach and Averbach [59], Anderson assumes three different representational types for these cognitive units: abstract proposition, encoding meaning; spatial image, encoding spatial configuration; and temporal string encoding the order of a set of items. For example, a cognitive unit may represent a proposition like (love, Jane, John); a spatial image like (a circle above a triangle); or a temporal string like(one, two, three, four). Cognitive units cannot contain more than five elements.

[5] The alternative, the faculty approach, rejects the assumption that uniform principles of growth or learning provide the basis of brain's development (see, e.g., Chomsky [18])

For example, (a small circle above a triangle) is not a cognitive unit. One can nevertheless use cognitive units as elements of other cognitive units, thus creating complex hierarchical structures. Cognitive units or elements are seen as nodes and their connections as links of a network. It is *assumed* that:

> "At any time t, any cognitive unit or element i has a nonnegative level of activation $a_i(t)$ associated with it."

and, expectedly, each cognitive unit or element has a strength, s_i, on the basis of which the relative strength of a link is defined (see Anderson [2], p. 22 for the defining function, and pp 169-170 for calculating the activation of a pattern node).

This notion of activation level actually represents the likelihood that a particular piece of information will be useful at a particular time t, and hence it is different from the neurophysiological conceptualisation of it. Equally, it is reminiscent of, but distinct from, the physical notion of field; as Anderson (ibid. p. 89) put it: "Activation flows from a source and sets up levels of activation throughout the associative network." There are three sources of activation:

(i) an encoded environmental stimulus (e.g., the encoded form of an image, or word);
(ii) the information structures determined by the execution of a production rule; and
(iii) the focused elements of working memory information structure.

The state (not the behaviour) of the system over time is given by a postulated differential equation describing the change in activation level in terms of a quantity positively proportional to the amount of input and negatively proportional to the current level of activation. Activation is arguably the most important single process in ACT theories. It defines ACT*'s working memory and its change determines the central operations of the cognitive architecture. To these operations and notions we turn next.

The central control mechanism of ACT* is an activation-based, data-flow pattern matcher (ADPM). It is a synthesis of Forgy's [25] data-flow PM and McClelland and Rumelhart's interactive activation model (McClelland and Rumelhart [42]; Rumelhart and McClelland [52]). This mechanism actually realises the following five conflict resolution principles characterising the ACT theory: (i) degree of match; (ii) production strength; (iii) data refractoriness; (iv) specificity; and (v) goal-dominance. Such a mechanism sets ACT* apart from, probably the majority of, other production systems which apply conflict resolution after pattern matching.

ACT's pattern matcher is capable of both full and partial matching and designed to prefer the former over the latter. Production strength reflects the frequency with which a production has been successfully applied in the past.

Data refractoriness refers to the idea that the same data cannot serve in two patterns simultaneously. For example, the Necker cube lines responsible for producing the two distinct representations of the cube. Consequently, data matched to one production cannot be matched to another one. Specificity is a standard AI option in conflict resolution; if two productions match the same data preference is given to the one with the most 'specific' condition. In ACT a rule A is considered more specific than a rule B if A's condition matches in a proper subset of the situations where B's condition would match.".

Goal-dominance is straight forward; productions referring to the current goal (it can only be one) take precedence over other applicable productions. If more than one goal-directed rules are applicable (i.e., match the current active goal) combination of principles (i), (ii), and (iv) select the one that will apply. Goal-dominance as a conflict resolution principle is sharply contrasted with Soar's complete lack of it (cf. Section 1).

The ACT theory has been tested on three domains: memory, learning, and language. With respect to memory, ACT's theory of memory follows directly from the knowledge representation scheme employed, and addresses successfully a number of issues concerning the encoding, retention, and retrieval of short stories and discourse. Within procedural learning ACT addresses phenomena of cognitive skills like computer programming and decision making. Tasks within these domains are particularly well suited to be represented by production rules. A fact that has led Anderson ([5], p. 4) to claim that "a cognitive skill is composed of production rules." With respect to language ACT's performance was rather restricted. It involved a couple of simulations of child language acquisition producing linguistic output in the form of a set of productions.

3. A BIOPHYSICALLY BASED THEORY OF MIND

A theory of mind based on evolutionary and neurophysiological findings and principles is currently under development by Edelman (e.g., [22, 23]. This theory is known as the Theory of Neuronal Group Selection (TNGS), and incorporates a new approach, synthetic neural modelling, to the construction of intelligent entities (see, for example, Reeke *et al.* [50]). The aim of this section is to

provide an outline of TNGS and Darwin III, a complex automaton built on the basis of its basic postulates.[6]

We start with introducing the three fundamental tenets of the TNGS: developmental selection, experiential selection and reentrant mapping. Developmental selection is concerned with how the brain anatomy "is first set up during development". It is assumed that the combined effect of three major types of events: molecular regulation, growth factor signalling, and selective cell death, result into a variety of anatomical networks called primary repertoires[7]. The selective nature of primary repertoire creation is assumed to be due to the topobiological competition of neuronal populations. Of course, developmental selection is constrained by the genetic code but the latter does not determine the specific connections of the primary repertoires. It is this nature-nurture interplay which provides the richness of individuality within the human species.

Experiential selection is the name given to the mechanism responsible for creating secondary repertoires through the selective strengthening or weakening of synaptic connections. It should be noted that developmental and experiential selections are intermixed. The primary reason is that synaptic modification or creation may well occur even in the developed brain as, for example, when new neuronal processes sprout out and form additional synapses.

The selectional mechanisms described above constitute the elements out of which the third mechanism- reentrant mapping or reentry- is being built. This process is postulated to explain how brain maps[8] interact and aims to provide (in combination with memory) at least some of the sought bridging laws between psychology and physiology. Not surprisingly, Edelman ([23], p. 85) considers this tenet to be

> "perhaps the most important of all the proposals of the theory, for it underlies how the brain areas that emerge in evolution coordinate with each other to yield new functions."

Reentrant mapping is characterised by three properties. First, map formation through primary and secondary repertoires action. Second, formed maps "are

---

[6] For a detailed overall description of Darwin III and its functional subsystems the reader is referred to Reeke *et al* [50] In the same article a brief comparison of the design principles of Darwin III with connectionist, AI and other neurobiological models is being made.

[7] A repertoire may de defined as a neuronal group able to respond to any of a wide class of inputs (see Reeke *et al* [50], Edelman [23]) Repertoires are distinguished into primary (arising by processes of somatic selection) and secondary (arising by selective modification of synapses).

[8] A map may be defined as a collection or class of reuronal groups

connected by massively parallel and reciprocal connections". Finally, connected maps interact through "reentrant signalling", that is, "as groups of neurons are selected in a map, other groups in reentrantly connected but different maps may also be selected at the same time." (ibid.).

Reentry is related to the notion of feedback but is fundamentally distinct from it. Reeke *et al.* ([50], p. 1501) specify five differences. First, "Reentry is inherently parallel and involves populations of interconnected units, whereas feedback involves the recursion of a single scalar variable." These populations are neuronal groups within brain maps. Second, reentry is 'distributed' in the sense that each area or map "simultaneously reenters to many other" areas or maps. Third, reentrant connections are not all used at all times[9]. Fourth, reentrant mapping "is used more for correlation than for error correction or gain control." Finally, as Edelman's quote above indicates reentry can give rise to the creation of novel functions.

This concludes our brief introduction to the three basic postulates of the TNGS and we proceed to describe its true heart consisting of the mechanisms proposed for perceptual categorisation, memory, and concept formation.

Perceptual categorisation may be defined as "the selective discrimination of an object or event from their objects or events for adaptive purposes." (Edelman [23], p. 87). The basic structure for perceptual categorisation is called global mapping. It contains a number of neural maps interacting both among themselves and with nonmapped parts of the brain like the basal ganglia and the hippocampus. Interactions within a global mapping allow

> "selectional events occurring in its *local* maps ... to be connected to the animal's motor behavior, to new sensory samplings of the world, and to further successive reentry events." (ibid. p. 89).

This complex, higher-order, dynamic structure provides the basis of perceptual categorisation through 'appropriate' selection of neuronal groups within its maps. According to the TNGS, 'appropriateness' refers to internal criteria of value which constrain the domains in which they occur without determining specific categorisations.

To illustrate the above description let us briefly consider Darwin III, a complex automaton of the Darwin series, designed to demonstrate both the feasibility of perceptual categorisation on the basis of global mappings and the constructability of devices which exhibit sensorimotor capabilities "without the use of a

[9] This property is taken as evidence for the statistical nature of reentry.

priori definitions of categories, codes, or information-processing algorithms. (Reeke *et al.* [50], p. 1501).

Darwin III consists of a simulated nervous system, a simulated movable eye, and a simulated four-jointed arm[10]. Its nervous system consists of a set of interconnected repertoires. In a particular version of the system, its major subsystem, categorisation, is supplanted with oculomotor, reach, and tactile subsystems of neuronal groups.

A simple form of 'appropriate' behaviour as value-constrained categorisation occurs in Darwin III by having visual value circuits built on the basis of, for example, "Value = "light is better than no light". In cases like these the interaction of the resulting value circuits with Darwin III's categorising reentrant systems result into selective synaptic strengthening. In Edelman's words:

> "the action of these value circuits enhances the probability that synapses active when such circuits are engaged will be strengthened in preference to competing synapses. The net result is that with selection and experience the eye of the automaton tracks signals from lit objects." (ibid., p. 92)

It should be stressed, nevertheless, that neither somatic selectional systems nor value-based circuitry are on their own adequate for behaviour. They are rather singly necessary and jointly sufficient conditions.

Regarding memory and, especially, concept formation Edelman's proposals are considerably less developed. Nevertheless, his approach is equally sound and, with respect to some of the findings of most of the other disciplines comprising cognitive science, his views are significantly more encompassing and accommodating than those of, say, Allen Newell. On the other hand, the reader is left, not infréquently, with the impression that some major theories developed in disciplines other than the neurosciences, are treated less convincingly than one would like to see. A case in point is Chomsky's research programme and associated theories (see, for example, Chomsky [18, 19]. Similarly, psychological theories of memory (e.g., Craik and Lockhart [21]; Gathercole and Baddeley [26]; Schank [53]) are not at all considered. Still his goal of putting the study of the mind on firm biophysical bases is commendable and a needed antidote to the exclusively computational, or psychological theories of mind.

[10] A robot called NOMAD (Neurally Organized Multiply Adaptive Device) and based on Darwin III's principles is currently under construction; it consists of the same three types of subsystems but with a more complex nervous system.

His starting point, the definition of memory as "the ability to repeat a performance" is questionable, in particular since he does not use 'performance' in any special technical sense. Furthermore, he does not provide any sort of bridge between this high level and very general description and his subsequent description of memory in terms of the basic principles of selection and reentry. Edelman's ([23], p. 102) views of memory, in terms of the theory of neuronal group selection may be best summarised by the following excerpt:

> "The TNGS proposes instead that memory is the specific enhancement of a previously established ability to categorize. This kind of memory emerges as a population property from continual dynamic changes in the synaptic populations within global mappings - changes that allow a categorization to occur in the first place. Alterations in the synaptic strengths of groups in a global mapping provide the biochemical basis of memory."

The questions then are: What *sort* "of specific enhancement"? How *exactly* did this specific enhancement come about? On these questions Edelman is silent, he does nevertheless make three quite interesting, if not important, points on memory.

The first point to be made is that the proposed kernel of a theory of memory makes a clear distinction between synaptic modifications, the physiological basis of memory, and memory itself which is seen as a systemic property emerging from dynamic interactions at the neuronal group level. The second point is a straightforward consequence concerning recall. Since an animal is a dynamic entity it finds itself in continually changing contexts, such differing contexts contribute to the creation of similar but distinct perceptual categories corresponding to similar but distinct neural formations. On such a basis, recall activates "the prevailing neural formations" (Gelepithis [27, 28]), or as somewhat differently Edelman ([23], p. 102) put it:

> "Recall involves the activation of some, but not necessarily all, of the previously facilitated portions of global mappings."

The above features also give rise to the property of inexactceness and the great generalisation capability characterising human memory. Of course, major questions concerning the real mechanisms still remain but even mere presentation of these would have taken us well beyond the scope of this article.

The last point concerns the relationship between short-term and long-term memories. Since neither classification couples nor global mappings could enable

the ordering of appropriate neural formations, which is a necessary requirement for having memories spanning any internal of time, additional brain structures need to have been developed. According to the TNGS these structures include: the cerebellum, the basal ganglia, and the hippocampus. Edelman summarises their memories-related functions in these words:

> "Had the structures and neuroanatomy of the cerebellum, or something like it, not evolved, smoothly coordinated and rapid motion would be compromised. Without the basal ganglia and their specific anatomy, animals would not be able to orchestrate whole symphonies of movements in a plan. Without the functions provided by the hippocampus, whole suites of categorization in a time range between the immediate and those forever stored could not be linked. And without that linkage, no long-term memory could be coherent." (ibid. p. 107).

In summary, new connections triggered by evolutionary development provide the basis for new memory functions beyond those occurring in lower vertebrates.

Concept formation is considered a fundamental ability which appeared prior to the development of language. As such primary linguistic constraints like arbitrariness, or sequentiality do not apply to concept formation. The theory of neuronal group selection

> "proposes that the evolutionary development of specialised brain areas is required before conceptual abilities emerge. The argument supporting this proposal is based on the notion that a simple increase in the number of reentrant maps capable of perceptual categorization is insufficient to account for concepts. Conceptual categorizations are enormously heterogeneous, and general. Concepts involve mixtures of relations concerning the real world, memories, and past behavior. Unlike the brain areas mediating perceptions, those mediating concepts must be able to operate without immediate input. (ibid. p. 108-109).

It seems that brain areas like the temporal, parietal, and especially frontal cortices are quite likely to have been able to support the activities required by the complexities of conceptual categorisation. It is in particular suggested that "the brain constructs maps of its *own* activities" in addition to those (perceptual maps) 'reflecting' aspects of the external world.

One thus envisages what may be called conceptual brain structures responsible for categorising and, even more importantly, recombining neural formations and probably brain activities which had been created or occurred as the result of perceptual categorisations. Such activities can naturally lead to the stage where conceptual brain structures will distinguish classes of global mappings (e.g., an object from a movement class). Edelman finishes his treatment of concept formation with a couple of claims in need of elaboration. He writes:

> "It is also possible to see how events may be categorized as "past" without necessitating their being played out in present brain activities, as they must be for short-term memory and for the hippocampal succession leading to long-term memory. Furthermore, one can see how concept areas, by recursively restimulating portions of global mappings containing previous synaptic changes, give rise to combinations of relationships and categories." (ibid. p 110).

Although plausible these claims: (i) hinge upon some fundamental assumptions which need to be spelt out clearly; and (ii) presuppose the existence and workings of several mechanisms in need of specification and testing. Still, Edelman's account of concept formation solely on the basis of his TNGS tenets, the processes of perceptual categorisation, memory, and learning alterations of reentrant connectivity, open the way for its interesting account of consciousness[11] in terms of further reentry connectivity alterations.

4. AuRA: AN INTEGRATED, BIOLOGICALLY-INSPIRED ROBOT

This section briefly reviews a case study illustrating the impact of cybernetics on the design of intelligent robots. Of the several[12] integrated robot systems we selected AuRA (Autonomous Robot Architecture) for two reasons. First, it attempts to integrate two of the most interesting and promising approaches to planning. Second, it employs concepts from biological systems, thus initiating a much needed conceptual interaction among disciplines.

[11] For a review of the major theories of consciousness, including Edelman's, see Gelepithis [34].

[12] Congdon *et al.* [20] provide an interesting comparison of the CARMEL and FLAKEY robotic architectures (they finished first and second respectively at the 1992 Robot Competition sponsored by the AAAI). Nourbakhsh *et al.* [47] may be consulted for introducing the winning robots for the 1993 event. For a very successful, autonomous, real-world operating, tactical driving system the reader is referred to Pomerleau and Jochem [48], and Jochem and Pomerleau [38]. Commercial robots most of the times look and sound impressive but they usually employ not up to date technology. Fred the robot, for example, can 'see' through walls, detect body heat and motion 150 feet in any direction, navigate over a relatively smooth terrain, and can warn of fire or flood. Fred is already patrolling the World Trade Centre in Boston and is being marketed across the US at a cost of $69,000. It sounds good but its technology is of the early 1980's.

The philosophy underlying AuRA may be best described as an attempt to: (i) integrate aspects of the knowledge-based and reactive approaches to planning; and (ii) employ ideas from biological systems as captured in disciplines like ethology, psychology, neuroscience, and cybernetics. From the reactive approach to planning, it primarily borrows the idea that tasks can be decomposed into a collection of low-level primitive motor behaviours. This idea takes the form of a schema-based methodology we introduce in Section 4.1. The notion of schema can also be seen from the perspective of capturing aspects of the functioning of biological hardware. From this perspective, in the abstracted form of schemas, it is seen as a much more promising approach than the structurally-based approach of artificial neural networks. From the knowledge-based approach to planning, models of world knowledge can be used as long as they: (i) provide expectations for perceptual processes (i.e., *what* to look for); (ii) provide focus of attention mechanisms (i.e., *where* to look for it); (iii) initially configure sensor fusion mechanisms (i.e., *how* to look for it); and (iv) sequence perceptual algorithms correctly (i.e., *when* to look for it). (Arkin [13]).

The first biologically-inspired idea is the notion of action-oriented perception with roots in both Cybernetics (Arbib [8]) and cognitive psychology (Neisser [44]). Its basic principle is that:

> "perception is predicated upon the needs of action: only the information that is germane for a particular task needs to be extracted from the environment" (Arkin [13]).

The second biologically-inspired idea is the notion of homeostasis. The term homeostasis was coined in the early 1930's by the physician Walter Cannon in his book *Wisdom of the Body*. Essentially, it refers to the processes by which a biological system maintains its internal states within a certain range. As such it is vital to the survival of the system since it contributes significantly to the provision of a relatively stable internal state. The basic idea of homeostasis is the principle of negative feedback which has since been recognised as an extremely important principle for almost all physiological processes and provided the basis of Cybernetics. This notion is employed in AuRA by utilising a very simplified model of the mammalian endocrine system (see Arkin [14] for details).

4.1. Design Goals, Methodology, and Architecture

Arkin [12] states six distinct design goals which have motivated the development of AuRA. They can be summarised as follows: (i) applicability over a wide range of navigational domains, for example, the interior of campus buildings, college campus, manufacturing setting, and 3-D navigation in space; (ii) allowing for incremental growth and development; (iii) utilising world knowledge; (iv) encouraging sensor and vehicle independence; (v) implementing aspects of survivability; and (vi) exploiting cognitive, neuroscientific, and ethnological models whenever appropriate for navigational purposes.

The above design objectives and underlying philosophy are coupled with a schema-based methodology to produce AuRA's basic design architecture. Schemas have been variously defined both within and outside robotics.[13] A schema can be characterised by the following two properties: (i) being a primitive unit of behaviour (out of which more complex behaviours can be constructed); and (ii) being a frame like representational and control unit. Arkin [14] distinguishes three principal types: motor (the basic units of motor action); perceptual (the basic units of perceptual activity); and signal (the basic units of homeostatic control). To illustrate the notion of schema as employed by the Georgia's Institute of Technology mobile robotics team consider the following six types of motor schema:

1. move-ahead: the robot moves in a specified direction at a given velocity.
2. move-to-goal ballistically: the robot moves at a constant velocity towards the goal.
3. move-to-goal in a controlled way: the robot moves towards the goal at a decelerating velocity (proportional to its distance from the goal).
4. stay-on-path: (assuming robot on the path) the robot moves at a constant velocity towards the centre of the path; if robot off the path, it moves at a constant[14] velocity towards it.
5. avoid-static-obstacle: robot moves away from detected obstacle.
6. noise: the robot's movement is affected by a random element[15].

[13] According to Arkin [15] the notion of schemas can be traced back to Kant who used it as a philosophical model for the explanation of behaviour. For a recent exposition of schema theory the reader is referred to Arbib [9]. For a presentation of the neuroscientific motivation for schema-based systems, a discussion on the schema-based robot navigation, and a justification in using schemas as computational models of perception and action the reader is referred to Arkin [15].

[14] Usually in the latter case the speed is higher.

[15] Arkin [15] defines this schema only in terms of its vectorial implementation (see section on implementation below).

Section 4.2. gives some examples of schema implementation. The rest of this section briefly presents AuRA's subsystems. The overall design architecture is illustrated in Figure 2.

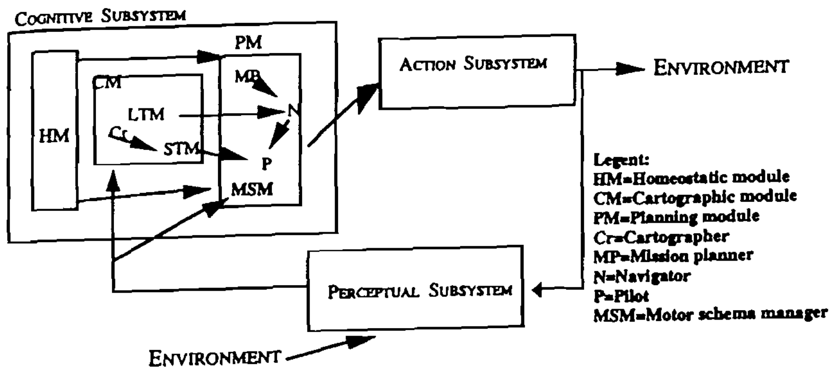

Figure 2. Functional outline of AuRA's subsystems and top-level architecture of its cognitive subsystem.

AuRA consists of three subsystems[16]: perceptual, cognitive, and action (or motor). The perceptual subsystem receives all sensory information and directs it to the cartographic and planning modules; these are two of the three modules comprising the cognitive subsystem, namely, cartographic, planning, and homeostatic. The cartographic module comprises several submodules. One of them, the cartographer, uses the sensory information it receives from the perceptual subsystem, for constructing world models maintained in the module's short memory. A second submodule, the long-term memory (LTM), stores and maintains a priori knowledge. The planning module consist of a hierarchical planner and a motor schema manager (a type of a distributed reactive plan execution system). The former determines a global path through the robot's modelled world. The latter uses the sensory information it receives from the perceptual subsystem, for processing by the perceptual schemas. The homeostatic module addresses some of the survivability issues concerning robots. It monitors the internal conditions of AuRA and feeds the data to both the hierarchical planner,

[16] Termed modules by Arkin; we think it is better to reserve this term to refer to the major components of AuRA's subsystems.

potentially affecting AuRA's high level planning, and the motor schemas themselves, potentially affecting AuRA's reactive behaviour. The action subsystem translates the results of the planning module into specific commands to be carried out by the actual robot R.

4.2. Navigation and Homeostatic Control

In this section we take a closer look at how AuRA's navigational behaviour is carried out. As we saw in the previous section its basic representational and control units are schemas. This mechanism is strongly influenced by the generalised potential fields approach of Krogh [39].

In a nutshell, a schema, as a representational mechanism, needs to be instantiated to provide the potential actions for the control of the robot. An instantiation is represented by a potential field. For example for the move-to-goal motor schema the goal is viewed as a potential well to which the robot is attracted. Schemas of course operate at all levels; as Arkin writes:

> "Motor schemas, when instantiated, must drive the robot to interact with its environment. On the highest level, this will be to satisfy a goal developed within the planning system; on the lowest level, to produce specific translations and rotations of the robot vehicle." (Arkin [10], p. 95).

Motor schemas, like those in Section 4.1., are combined with 'priority rules' to produce more complex (emergent) behaviours; they determine the resolution of conflicts arising from opposing motor schemas. For example, assume that a robot, R, is instructed to move in a particular direction avoiding static obstacles and staying on a path; further assume that at some point in time after the robot started its navigation a static obstacle completely blocks its path. Clearly at this point in time a dead end is reached. Let us see how R would overcome this problem, assuming that the following priority rule holds true: If static obstacle blocks completely the path, then give up the stay-on-path schema. After having reached the dead-end point, R reduces the field allowing itself to wander off the path and, thus, circumnavigating the obstacle. This circumnavigation results in changes in the direction of the force produced by the blocking obstacle, this will inevitably lead to a direction of force indicating that the blocking obstacle has been successfully passed. As soon as this happens the priority rule invoked before ceases to be applicable and the stay-on-path schema produces its original field that causes R to move towards its original path. An alternative way to overcome dead-ends is through the use of the noise schema. In this case the effect of instantiating the noise schema is to modify the robot's velocity, by superimposing

a small random vector, thus enabling R to come away from undesirable equilibrium points[17].

The next couple of paragraphs introduce the two classes of signal schemas (transmitter, and receptor) and the relations between perceptual, signal, and motor schemas. The transmitter class of signal schemas is associated with internal sensors, (e.g., ammeter, fuel tank measuring device, etc.) and provides the feedback required to achieve homeostatic control (for example, they may send information on the rate of consumption). The receptor class of signal schemas is embedded within motor schemas. Receptors schemas modify their vector output by altering their parameters on the basis of the information they receive from the transmitter schema.

The characteristic output vector of each schema is derived from three sources: (i) its goal; (ii) the currently perceived state of the world; and (iii) the robot's internal state (provided by the signal schemas). The sum of all motor schema vectors determines the current direction and velocity of the robot. Figure 3 illustrates the relations among the three types of schemas. The reader should notice that both perceptual and receptor schemas are embedded within motor schemas. Indeed as Arkin remarks, the receptor schemas can be viewed as internal perceptual schemas.

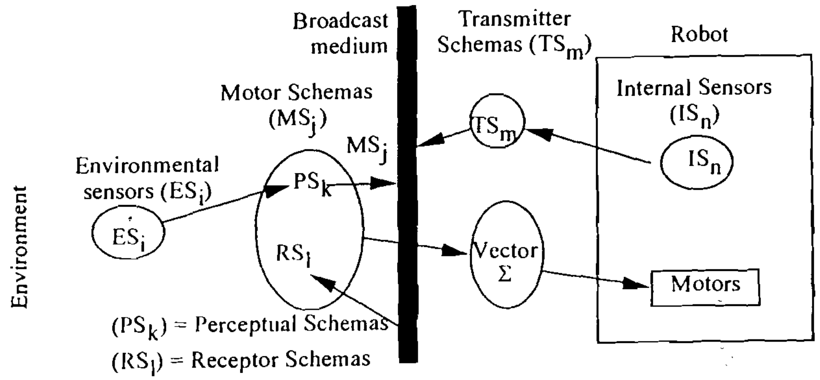

Figure 3. Interschema relationships (adapted from Arkin [14], p. 204).

[17] Undesirable equilibrium points arise whenever at least two instantiated motor schemas counter balance each other

AuRA's planning module is hierarchical and consists of: a mission planner; a navigator; a pilot; and the motor schema manager. The mission planner receives and interprets commands from a human, determines the nature of the mission, sets parameters for the navigator and pilot, etc. In particular, the mission planner defines the notion of 'optimality'.

The navigator produces a plan in terms of a set of goals and 'asks' the pilot to satisfy one such goal at a time. The overall navigation plan is, essentially, a global path consisting of a series of piecewise linear path segments. This global path is constrained by the robot's overall mission and is based upon information received from a static map (called the meadow map- see below) representing the terrain to be navigated and stored in R's long term memory. Each segment is described in terms of a set of perceptual and motor schemas[18]. The concurrent instantiation of these schemas navigate the robot accordingly. AuRA's navigator accepts a start and a goal point from the mission planner and determines a 'reasonable' path to achieve the set objective. It is thus in contrast to the usual approach of attempting to construct the 'best' or 'optimal' path for a given initial situation and goal state. A 'reasonable' path is defined as a path that can be successfully executed by a robot given its available resources (e.g., time or/and energy). An 'optimal' path may be defined as a path that maximises a particular parameter (for example, fastest-least time; shortest-least distance; etc.). This is a reasonable strategy and as Arkin ([11], p. 55) remarks:

> "even if an optimal path (by whatever definition) was attainable by the navigator, it could only be based on partial information (i.e., the modeled world). Since the robot's environment is subject to unmodeled and moving obstacles, there is no a priori guarantee that any path produced by any navigator is the best path, given only incomplete world knowledge (although the path can be optimized relative to the current world model)."

Search is carried out through the application of a heuristics-enhanced A* algorithm.

The meadow map is stored in the robot's long term memory (LTM). Aerial maps (for outdoor terrains), and usual architectural diagrams (for indoor terrains) are used as basis for the building of the meadow maps. It should be noticed that the meadow map is only a representation of a particular terrain and as such it does not usually represent all the objects, or indeed obstacles, existing or created

---

[18] An example of a typical leg of a path may be: go on for 30 meters on this path then turn left 90 degrees at the traffic lights.

within the terrain. Consequently the navigator takes no account of any unrepresented objects in specifying the path to be executed. As a result untenable paths may well be presented to the pilot for execution and it is the pilot's job to ask the navigator for an alternate route on the basis of his current perception of the world stored in short term memory (STM). STM representations are LTM descriptions instantiated by sensor data (visual and ultrasonic) provided by the robot's perceptual system.

The pilot is concerned with: (i) satisfying one subgoal from the navigator at a time; and (ii) avoiding unmodelled obstacles. Goal satisfaction is subject to constraints received from the navigator such as criteria for failure to attain a goal. In such a case the navigator is informed and replanning is initiated. It should be noted that the pilot can make alterations to the route as long as such alterations fall within acceptable limits. Finally, the pilot transforms the received goals into a collection of motor behaviours and perceptual strategies. The motor schema manager is the execution arm of the pilot carrying out the collection of motor behaviours and perceptual strategies produced by the latter.

For a robot to be really autonomous it must both be able to cope with (adverse) environmental events and sustain its internal state within an acceptable range of parametric variation. Work in robotics had so far ignored the latter requirement. One of AuRA's basic working hypotheses is that "many aspects of self-preservation can be implemented by taking advantage of a control system analog of the endocrine system in mammals." (Arkin [14], p. 198). Some of the major aspects of the mammalian endocrine system which have been incorporated in Arkin's approximate implementation of it are summarised[19] below:

1. It is concerned with the regulation of the robot's internal state.
2. Transmitter schemas employ a non hierarchical broadcast mechanism.
3. Receptor schemas implement robot's targetability.
4. Specification of parameters within the R-schemas influence motor schemas.
5. Maintenance levels are maintained without high-level planning assistance.
6. Internal sensing data can affect high-level planning.

A simulated demonstration has shown that the robot's navigational behaviour changes considerably as a function of its fuel reserves. When ample fuel is available it chooses the fastest albeit longer path. As fuel reserves decrease it starts producing a series of shorter paths with slower associated speeds. Such simulations may well open up the way for a robust implementation of a homeostatic control system.

---

[19] For more information the reader is referred to Arkin [12, 14].

5. COMPARISON AND CONCLUSIONS

Since so far no effort has been made for any theories of mind, either embodied or not, to address the same tasks and cognitive phenomena, no comparison along such lines is possible. In addition, since the cardinality of the set of human response functions, tasks, and psychological phenomena is immense, a fruitful comparison along such lines seems to be doomed for quite some time, if not forever.

The alternative is to follow Newell [45] and specify a set of constraints, or design requirements, which any theory of mind should satisfy and compare a proposed theory on the basis of such a class. Table 1 provides such a list along with indications of whether a research programme has addressed each specified constraint or design requirement, and of a rough, qualitative measure of success. An early version of most of these requirements, in terms of the notion of an *ideal robot*, may be seen in Gelepithis [27, 29]. For a different set of constraints, heavily influenced by Soar, see Newell [45]. The rest of this section briefly comments on some of the proposed requirements.

Constraint 2 implies capability to operate in real time; though not, necessarily, in a sense understood by humans.

Language could have been considered under design requirement 7, but its significance and complexity put it in a class of its own.

Requirement 5 refers to a fundamental process which presupposes both *understanding* and, through the latter, *meaning*. It is very important to be realised that the assignment of meanings, and making sense of something are processes fundamentally different from those involved in the use of symbols and abstractions. Although, work on cognitive functions constitutes the bulk of research in Artificial Intelligence both the PSSH and the situated action approaches have ignored the process of *communication*. This I consider to be a significant omission. For individual, isolated voices on the significance of communication see Gelepithis ([27], for its centrality too; [29, 30], for its invariance nature too; [33]), Bobrow [16], Grosz [36].

Finally, I would like to note that although the ability to communicate is closely related to the constraint of using human language, it is considerably different. In a nutshell, the use of human language does not necessitate communication, nor it forces the restructuring of the mind as the processes of communication and understanding do (Gelepithis and Goodfellow [35]; Gelepithis [31]). Constraints 8 and 9 are of outmost importance if machine intelligence independent of human intelligence is/were to be developed. Gelepithis [32] discusses and briefly justifies constraint 8. The next paragraphs provide brief support for

Table 1. Comparison of embodied theories of mind in terms of constraints and design requirements.

Constraints or design requirements	**Soar**	**Act**	**TNGS**	**AuRA**
1. Use of: (i) domain knowledge; and (ii) commonsense knowledge for problem solving.	Vast amounts of (i) very successfully.	Very considerable success in using (i).	Not really.	Limited use of (i).
2. Be able to operate in environments of, at least, Earth-level complexity.	Interfaces only.	No.	Primarily simulations.	Limited.
3. Operate autonomously, but within a social community.	No.	No.	No.	Aspects of Survivability, i.e. homeostatic control.
4. Acquisition and use of language to, at least, human-level complexity.	Rudimentary.	Moderate (in simulation terms).	Epigenetic development of syntax.	No.
5. Able to communicate.	Not addressed.	Not addressed.	Not addressed.	Not Addressed.
6. Be conscious.	No.	No.	Not implemented theory.	Incorporates an ego-model of itself.
7. Be able to develop: (i) skills, e.g., through learning; and (ii) judgement, e.g., through maturation.	(i) Yes. (ii) No.	(i) Yes. (ii) No.	Blocks-world level learning of skills.	(i) Navigational skills at most. (ii) No.
8. Develop *own* representational system	Not addressed.	Not addressed.	Not addressed, but approach promising.	Not Addressed, but Approach promising.

9. Combine perceptual and motor information with *own* belief systems.	Not addressed.	Not addressed.	Not addressed.	Not Addressed.
10. Be creative	Not addressed.	Not addressed.	Not addressed.	Not Addressed.
11. Explain human neonate's capabilities for development.	Not addressed.	Not addressed.	Addressed.	Not Addressed.
12. Emotion.	Not addressed.	Not addressed.	Not addressed.	Not Addressed.

design requirement 9 in terms of the nature of vision and manipulation. With respect to scene understanding, I restrict myself to the problem of understanding the *nature* of a scene, in other words, I restrict myself to the 'computational' level of scene understanding (Marr [41]). Artificial vision has made considerable progress in identifying objects in a scene and describing their geometric relations but is currently far behind exhibiting most of the hypothetico-deductive reasoning involved in human scene-interpretation.

Humans are able to *make sense* out of an extremely varied and virtually infinite number of visual *stimuli*. Light signals from objects and scenes of the world around us impinge upon the retina of the human eye and a most intricate series of processes take place which result in what we call vision. A scene rarely consists of a single object. Scene understanding depends on the kind of entities present in the scene and the types of interaction among them. Use of such constraining knowledge is of fundamental importance in successful scene understanding and presupposes considerable classificatory, and deductive capabilities on behalf of the system. Furthermore, in most cases, additional world knowledge is required which is not deducible from the scene entities themselves but from our common-sense knowledge of say, gravity, causality, the survival instinct, human intentions, etc. In some cases further knowledge of the specific scene context like atmospheric conditions, ambient light or the angle of lighting are required.

We know that the visual image depends on: (i) the properties and coincidental features of an object (for example, its shape, size, material composition, surface properties, position and orientation); (ii) the nature of illumination and related effects (e.g., shadowing, reflectance); and (iii) the nature of the visual system itself[20]. Furthermore, at least as far as humans are concerned, the assignment of *meaning* to the retinal image depends, also, on one's belief systems; a

[20] A visual system's own representational capacity and fidelity depends upon their functional range in the electromagnetic spectrum, their design, and their objectives and overall knowledge.

term I use to refer to one's expectations, general and domain knowledge, and so on. In humans therefore, information derivable from two very different sources needs to be combined in order to fully understand a scene: the retinal image of the scene itself and the particular human's belief system. This interaction between aspects of the subjective and the objective is one of the primary characteristics of human vision, and therefore, seems to me imperative that it should be a constraint of any theory of mind.

With respect to manipulation it is assumed that three sensing operations are adequate for determining a complete tactile sensing system Staugaard [54]. These three fundamental operations are: (i) *Joint force:* Sensing the force applied to the robot gripper, as well as to the robot wrist, elbow and shoulder joints; (ii) *Touch:* Sensing the pressures applied to various points on the gripper surface; and (iii) *Slip:* Sensing any movement of the object while it is being grasped. On this basis, I believe that there are two constraints which follow directly from Staugaard's data type requirements, and which have not received any attention despite their significance in the quest for artificial intelligence. They may be summarized as follow. The first concerns the assignment of *meaning* to the data supplied by the three sensing operations. The second concerns the augmentation and (re)organization of memory to accommodate the semantically loaded data and *integrate* them with the data supplied by the other perceptual modalities as well as with the intelligent robot's belief systems.

I use creativity as a generic term referring to the class of characteristically-human processes whose members include intuition, tacit knowledge, and insight.

Finally, constraints 11 and 12 need to be addressed by any theory of human mind, but they do not *necessarily* constitute constraints for an artificial mind.

Whether any of the proposed theories can be extended to account for all 12 constraints is a significant open question. What is important is that unified theories of mind should, at this stage of the development of Psychology and Artificial Intelligence, take precedence over the discovery of yet another regularity. The road to the unraveling of the mysteries of mind is fascinating but full of thorny paths, blind alleys, and loopholes. Excellent theoretical work is urgently required to guide further development of all the subfields of cognitive science.

REFERENCES

[1] Anderson, J.R., *Language, Memory and Thought*. Lawrence Erlbaum Associates, Hillsdale, NJ 1976.

[2] Anderson, J.R., *The Architecture of Cognition*, Harvard University Press 1983.

[3] Anderson, J.R., *The Adaptive Character of Thought*, Lawrence Erlbaum Associates, Hillsdale, NJ 1990.

[4] Anderson, J.R., The place of cognitive architectures in a rational analysis, In: *Architectures for Intelligence* (ed. by K. VanLehn), The 22nd Carnegie Mellon Symposium on Cognition, Lawrence Erlbaum Associates 1991.

[5] Anderson, J.R., *Rules of the mind*, Laurence Erlbaum Associates, Inc., Hillsdale, NJ 1993.

[6] Anderson, J.R. and Bower, G.H., *Human Associative Memory*, Winston, Washington DC 1973.

[7] Anderson, J.R. and Thompson, R., Use of analogy in a production system architecture, In: *Similarity and analogical reasoning* (ed. by S. Vosniadou and A. Ortony), Cambridge University Press (1989), 267-297.

[8] Arbib, M.A., *The Metaphorical Brain*, John Wiley 1972.

[9] Arbib, M.A., Schema theory, In: *The Encyclopedia of Artificial Intelligence, 2nd ed.*. (ed. by S.C. Shapiro), Wiley Interscience, NY (1992).

[10] Arkin, R.C., Motor schema-based mobile robot navigation, In: *The International Journal of Robotics* **8**:4 (1989), 92-112.

[11] Arkin, R.C., Navigational path planning for a vision-based mobile robot, *Robotica* 7 (1989), 49-63.

[12] Arkin, R.C., The impact of cybernetics on the design of a mobile robot system: A case study, In: *IEEE Transactions on Systems, Man and Cybernetics* **20**:6 (1990), 1245-1257.

[13] Arkin, R.C., The multiple dimensions of action-oriented robotic perception: Fission, fusion, and fashion, *Workshop on Sensory Aspects of Robotic Intelligence, AAAI Fall Symposium Series*, Monterey, CA (1991).

[14] Arkin, R.C., Homeostatic control for a mobile robot: Dynamic replanning in hazardous environments, In: *Journal of Robotic Systems* 9:2 (1992), 197-214.

[15] Arkin, R.C., Modeling neural function at the schema level: Implications and results for robotic control, In: *Biological Neural Networks in Invertebrate Neuroethology and Robotics*, Academic Press (1993).

[16] Bobrow, D.G., Dimensions of interaction, *AI Magazine* **12**:3 (1991), 64-80.

[17] Changeux, J-P. and Dehaene, S., Neuronal models of cognitive functions, In: *Cognition* **33** (1989), 63-109.

[18] Chomsky, N., *Rules and Representations*, Basil Blackwell 1980.

[19] Chomsky, N., *Language and Problems of Knowledge*, The MIT Press 1988.

[20] Congdon, C., Huber, M., Kortenkamp, D., Konolige, K., Myers, K., Saffiotti, A., and Ruspini, E. H., CARMEL versus FLAKEY: A comparison of two winners, *AI Magazine* **14**:1 (1992), 49-57.

[21] Craik, F.I.M. and Lockhart, R.S., Levels of processing: A framework for memory research, *J. Verbl. Lea and Verbl. Behav.* **11** (1972), 671-684.

[22] Edelman, G.M., *Neural Darwinism: The Theory of Neuronal, Group Selection*, Basic Books, NY 1987.

[23] Edelman, G.M., *Bright Air, Brilliant Fire: On the Matter of the Mind*, Basic Books, Harper Collins 1992.

[24] Feldman, J.A. and Ballard, D.H., Connectionist models and their properties, *Cognitive Science* 6 (1982), 205-254.

[25] Forgy, C.L., *On the Efficient Implementation of Production Systems*, Ph. D. dissertation, Computer Science Dept., Carnegie-Mellon University 1979.

[26] Gathercole, S.E. and Baddeley, A.D., *Working Memory and Language*, Lawrence Erlbaum Associates 1993.

[27] Gelepithis, P.A.M., *On the Foundations of Artificial Intelligence and Human Cognition*, Ph.D. Thesis, Department of Cybernetics, Brunel University, England 1984.

[28] Gelepithis, P.A.M. Knowledge, truth, time, and topological spaces, *Proc. of the 12th International Congress on Cybernetics*, Namur, Belgium (1989), 247-256.

[29] Gelepithis, P.A.M., The possibility of Machine Intelligence and the impossibility of Human-Machine communication, *Cybernetica XXXIV*:4 (1991), 255-268.

[30] Gelepithis, P.A.M., *Understanding and Communication: Invariance and Centrality*, Working paper '93-2, School of Information Systems, Kingston University Library 1993.

[31] Gelepithis, P.A.M., Aspects of the theory of the human and the human-made minds, Paper presented at the *13th International W'shop of the European Society for the Study of Cognitive Systems*, Wadham College, University of Oxford, 5-8 September, 1995.

[32] Gelepithis, P.A.M., Revising Newell's conception of representation, *Cognitive Systems* 4:2 (1995), 131-139 (Special issue on Representation).

[33] Gelepithis, P.A.M., A rudimentary theory of information: Consequences for information science and information systems, *World Futures* 49 (1997), 263-274.

[34] Gelepithis, P.A.M., Comparison of theories of consciousness, *J. of Consciousness Studies*, submitted.

[35] Gelepithis, P.A.M. and Goodfellow, R., An alternative architecture for intelligent tutoring systems: Theoretical and implementational aspects, *Interactive Learning International* **8**:3 (1992), 171-175.

[36] Grosz, J.B., Collaborative systems, *AI Magazine* **17**:2 (1996), 67-85.

[37] Holland, J.H., Holyoak, K.J., Nisbett, R.E. and Thagard, P.R., *Induction: Processes of inference, Learning, and Discovery*, The MIT Press 1986.

[38] Jochem, T.M. and Pomerleau, D.A., Life in the fast lane: The evolution of an adaptive vehicle control system, *AI Magazine*17:2 (1996), 11-50.

[39] Krogh, B., A generalised potential field approach to obstacle avoidance control, Technical Paper **MS84-484**, SME-RI; Society of Manufacturing Engineers, Dearborn, Michigan, USA (1984).

[40] Laird, J., Hucka, N. and Huffman, S., An analysis of Soar as an integrated architecture, *SIGART Bulletin* 2:4 (Aug. 1991).

[41] Marr, D., *Vision*, W. H. Freeman 1982.

[42] McClelland, J.L. and Rumelhart, D.E., An interactive model of context effects in letter perception: Part-I An account of basic findings, *Psychological Review* **88** (1981), 287-330.

[43] Minsky, M., *The Society of Mind*, Heinemann, London 1986*1987.

[44] Neisser, U., *Cognition and Reality: Principles and Implications of Cognitive Psychology*, W. H. Freeman 1976.

[45] Newell, A., *Unified Theories of Cognition*, Harvard University Press 1990.

[46] Newell, A., Rosenbloom, P.S. and Laird, P.J., Symbolic architectures for cognition, *Foundations of Cognitive Science* (ed. by M.I. Posner), MIT Press (1989).
[47] Nourbakhsh, K., Morse, S., Becker, C., Balabanovic, M., Gat, E., Simmons, R., Goodridge, S., Potlapalli, H., Hinkle, D., Jung, K. and Vactor D.V., The winning robots from the 1993 robot competition, *AI Magazine* **14**:4 (1993), 51-62.
[48] Pomerleau, D.A. and Jochem, T.M., A rapidly adapting machine vision system for automated vehicle steering, *IEEE Expert* **11**:2 (1996), 19-27.
[49] Pylyshyn, Z.W., The role off cognitive architecture in theories of cognition, In: *Architectures for Intelligence* (ed. by K. VanLehn), The 22nd Carnegie Mellon Symposium on Cognition, Lawrence Erlbaum Associates, Hillsdale, NJ (1991).
[50] Reeke Jr., G.N., Sporns, O. and Edelman, G.M., Synthetic neural modeling: The "Darwin" series of recognition automata, *Proc. of The IEEE* **78**:9 (1990), 1498-1530.
[51] Rosenbloom, P.S., Newell, A. and Laird, J.E., Toward the knowledge level in Soar: The role of the architecture in the use of knowledge, In: *Architectures for Intelligence* (ed. by K. VanLehn), Lawrence Erlbaum Associates, Hillsdale, NJ (1991).
[52] Rumelhart, D.E. and McClelland, J.L., An interactive model of context effects in letter perception: Part-II The contextual enhancement and some tests and extensions of the model, *Psychological Review* **89** (1982), 60-94.
[53] Schank, R.C., *Dynamic Memory: A Theory of Reminding and Learning in Computers and People*, Cambridge University Press 1982.
[54] Simon, H.A., Cognitive architectures and rational analysis: Comment, In: *Architectures for Intelligence* (ed. by K. VanLehn), The 22nd Carnegie Mellon Symposium on Cognition, Lawrence Erlbaum Associates, Hillsdale, NJ (1991).
[55] Staugaard Jr., A.C., *Robotics and Artificial Intelligence*, Prentice-Hall 1987.
[56] Tamble, M., Newell, A. and Rosenbloom, P.S., The problem of expensive chunks and its solution by restricting expressiveness, *Machine Learning* **5**:3 (1990), 299-348.
[57] Tulving, E., *Elements of Episodic Memory*, OUP 1983.
[58] VanLehn, K., *Architectures for Intelligence*, Lawrence Erlbaum Associates, Hillsdale, NJ 1991.
[59] Wallach, H. and Averbach, E., On memory modalities, *American Journal of Psychology* **68** (1955), 249-257.

[1]Kingston University
Kingston upon Thames
KT1 2EE, England
UK
~~E-mail~~: petros@kingston.ac.uk

Current:
petros.gelepithis@philosophy.ox.ac.uk
petros2004@phs.uoa.gr

Part Four: ANNs - Stability and Application Domains