Modularity and Specialized Learning: Mapping Between Agent Architectures and Brain Organization

Joanna Bryson and Lynn Andrea Stein[1]

Artificial Intelligence Laboratory, MIT
545 Technology Square, Cambridge MA 02139, USA
joanna@ai.mit.edu and las@ai.mit.edu

Abstract. This volume is intended to help advance the field of artificial neural networks along the lines of complexity present in animal brains. In particular, we are interested in examining the biological phenomena of *modularity* and *specialized learning*. These topics are already the subject of research in another area of artificial intelligence. The design of *complete autonomous agents* (CAA), such as mobile robots or virtual reality characters, has been dominated by modular architectures and context-driven action selection and learning. In this chapter, we help bridge the gap from neuroscience to artificial neural networks (ANN) by incorporating CAA. We do this both directly, by using CAA as a metaphor to consider requirements for ANN, and indirectly, by using CAA research to better understand and model neuroscience. We discuss the strengths and the limitations of these forms of modeling, and propose as future work extensions to CAA inspired by neuroscience.

Keywords: Spatial, Structural and Temporal Modularity; Complete Autonomous Agents; Behavior-Based AI; Brain Organization; Action Selection and Synchronization; Perceptual, Episodic, and Semantic Memory

1 Introduction

Although artificial neural networks (ANN) are vast simplifications of real neural systems, they have been a useful technology for helping us think about and model highly distributed systems of representation, control and learning. This work has proven useful both in science, by providing models, paradigms and hypotheses to neuroscientists; and to engineering, by providing adaptive control and classifier systems. In this chapter, we will propose that another area of AI, the agent literature, may also further both science and engineering. We also use agent software architectures to propose an ANN model that addresses issues of modularity, specialized learning, and to some extent synchronization.

ANNs can not so far be used for control systems that attempt to replicate the behavioral complexity of *complete* animals. One reason is that the complexity of such systems

[1] LAS: also Computers and Cognition Group, Franklin W. Olin College of Engineering, 1735 Great Plain Avenue, Needham, MA 02492. las@olin.edu

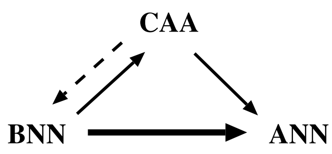


Fig. 1. This chapter addresses the goal of advancing artificial neural networks (ANN) through understanding neuroscience (BNN) via a third field, complete autonomous agents (CAA). This assistance comes both directly, by providing understanding and support for the requirements of such artificial systems, and indirectly, by providing another AI model of biological intelligence .

effectively requires decomposition into modules and hierarchy (see [10] for discussion and references.) This requirement is not theoretical, but practical. In theory, monolithic systems may be Turing complete; but whether attacked by design or by learning, in practice complex control requires decomposition into solvable subproblems.

This principle of modularity is one of several features that characterize the control architectures that *have* been successfully used for complete autonomous agents (CAA). Such agents include autonomous mobile robots [35], virtual reality characters [55], and intelligent environments or monitoring systems [17]. CAA, like animals, must process complex, ambiguous perceptual information. They must also control many forms of action and expression, often with multiple means for achieving any particular behavior. They must also manage a large number of competing and possibly contradictory goals. For example, an animal or animal-like robot may need to balance the need to find food with the need to remain inconspicuous; a character-based tutoring system may need to be clear but not boring; a household robot may need to do the laundry and cook dinner. CAA are situated in the real world in real time. Consequently they must deal with timeliness and synchrony: the order and duration of their expressed behaviors matter.

We begin this chapter with a discussion of modularity as found in nature, then review the literature on CAA in this light. Next, we demonstrate the relevance of CAA to neuroscience by describing a mapping between the common features of CAA architectures and mammalian brain organization. We then examine the implications for learning in the CAA context, which is necessarily specialized due to the representational architecture. We also propose a neural architecture for an agent capable of full, animal-like learning, and future directions, inspired by neuroscience, for CAA itself.

2 Modularity in Nature

There are at least three types of modularity in mammalian brains. First, there is *architectural modularity*. Neuroanatomy shows that the brain is composed of different organs with different architectural structures. The types and connectivity of of the nerve cells and the synapses between them characterize different brain modules with different computational capabilities. Examples of architectural modules include the neocortex, the cerebellum, the thalamus, the hippocampus, periaqueductal gray matter and so forth: the various organs of the fore, mid and hindbrains.

Second, there is *functional modularity*. This modularity is characterized by differences in utility which do not seem to be based on underlying differences in structure or computational process. Rather, the modules seem to have specialized due to some combination of necessary connectivity and individual history. Gross examples include the visual vs. the auditory cortices. Sur et al. [56] have shown at least some level of structural interchangeability between these cortices by using surgery on neonate ferrets. There is also other convincing and less invasive evidence. For example, many functionally defined cortical regions such as V1 are in slightly different locations in different people [45]. Many people recover capacities from temporarily debilitating strokes that permanently disable sections of their brains, while others experience cortical remaps after significant alterations of there body, such as the loss of a limb [50]. This evidence indicates that one of the brain's innate capabilities is to adaptively form functionally modular organizations of neural processing.

Thirdly, there is *temporal modularity*. This is when different computational configurations cannot exist contemporaneously. There are at least two sorts of evidence for temporal modularity. First, many regions of the brain appear to have local "winner take all" connection wiring where a dominant "impulse" will inhibit competing impulses [26, 28]. This neurological feature has been used to explain the fact that humans can only perceive one interpretation of visually ambiguous stimuli at a time [48]. Second, many cells in the brain are members of more than one assembly, and can perform substantially different roles in only subtly different contexts (e.g. in the hippocampus [34, 65].) This sort of temporal modularity is not yet well understood, but it could have implications for individual differences in intellectual task performance such as insight and metaphoric reasoning.

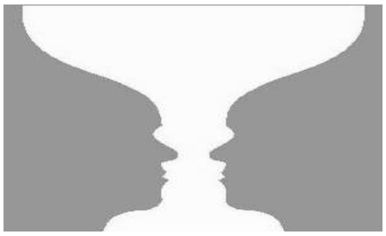


Fig. 2. An ambiguous image. This figure can be seen either as a vase or as two faces, but not both same time. From [25, p. 213]

The presence of these forms of modularity in mammalian brains motivates modular architectures in two ways. First, if we are interested in modeling the brain as a matter of scientific interest, we will need to be able to replicate its modularity. Second, the presence of modularity in the best examples of intelligent control available indicates that modularity is a useful means of organizing behavior. Evolution is not a perfect designer — the mere presence of a solution in nature does not prove it is optimal. However, given the extent and complexity to which the brain has evolved, it is at least worth treating the utility of modularity as a hypothesis. We now consider further evidence for the utility of modularity in another domain of intelligent control.

3 Common Features of Complete Autonomous Agents

Modularity is a key feature of most successful CAA architectures. Other architectural features that have been found useful in CAA architectures are hierarchical and sequential structures for ordering the expression of behaviors, and dedicated, reactive alarm systems capable of switching the focus of attention to different points in a control hierarchy. In this section we review the evidence for the importance of these features. A more extensive review is available in [7].

3.1 Skill Modules

One of the fundamental incentives for AI is the idea of having intelligent technology as a companion, situated with us in our environment. However, this goal was long seen as not technically feasible. Consequently, most early AI systems focussed on only a restricted set of capabilities such as reasoning in a particular domain [6, 60]. Such systems were relatively unconstrained by issues such as timeliness or uncertainty about their perceptions or actions. As a result, early attempts at building CAA, which utilized these techniques, tended to be both slow and enormously costly [e.g. 31, 44, 46].

Around 1986 a new paradigm based on greater modularity became established in mobile robotics [40]. This paradigm, behavior-based AI (BBAI), favored niche-specific solutions for CAA. BBAI agents feature modules composed of sensing and control specialized to a particular task to be executed in a particular situation. Such skill modules are often referred to as "behaviors" in the CAA literature to this day, despite the fact that they actually *generate* behavior rather than represent it. Further, much of expressed behavior is supposed to emerge from the interaction of these 'behaviors' rather than being generated by any one-to-one mapping. The most difficult part of the BBAI discipline falls to the engineer: determining what behavioral modules the agent needs, and guaranteeing that these modules do not interfere with each other in the normal operation of the robot.

3.2 Action Selection

By the late 1990's, modular "behavior" systems were ubiquitous in mobile robotics and virtual reality [30, 35, 54]. Also by this time, the need for structured action selection had become well accepted, despite initial resistance in the BBAI community. Structured action selection is essentially the following of plans, and conventional planning was one of the impracticalities of old-style CAA. However, although constructive planning is too costly for real-time systems [16], action selection based on *established* plans is not. This technique, called *reactive planning*, has been facilitated technologically as well as culturally. Technologically, new plan representations were developed which were more flexible and adaptive to dynamic environmental considerations [e.g. 12, 22, 24, 47].

Some of the systems using these representations derive from traditional planning, with behavior modules relegated to being mere plan primitives, but others maintain the autonomy of the behavior modules, while allowing for an extra system to arbitrate the timing of expressed actions. This latter approach also characterizes the relatively new paradigm of multi-agent systems (MAS) which are now sometimes used to control intelligent systems and environments [61]. Their coordination issues are similar to the arbitration issues of BBAI, and will probably converge on similar solutions.

Structured action selection normally consists of both hierarchical and sequential components. Hierarchies are necessary to reduce the search space for the next action. They focus attention on a particular set of actions likely to be useful in the current context. Sequences are a special case of this process where each action can be followed very quickly and reliably by another, a necessity in certain kinds of fine control [32, 38]. Sequences are not adequately represented by simple chains of productions, partly because their elements may participate in more than one sequence, but also because of timing issues. This problem was discovered and addressed in the Soar architecture, the best established production-based AI architecture, when it was applied to the CAA problem [36].

3.3 Environment Monitoring

The difficulty with structured action selection in a dynamic environment is that it can leave the agent unprepared to cope with sudden, unexpected events. Consequently, the third necessary feature of CAA architectures is an independent, parallel environment monitoring or 'alarm' system for switching attention. Such a system must operate in parallel with the agent's main attention, and must be able to recognize critical situations very rapidly with minimal 'cognitive' overhead. That is, it cannot require the reactive planning system itself, but must rely on simple perception. All modular CAA systems have this feature. In the early BBAI system, *every* action was continuously monitoring the environment with its specialized perception, prepared to take control of the agent whenever necessary [5]. In systems closer to conventional planning, such as PRS [24], control cycles between the conventional action selection and the monitoring system. In our own system, context is quickly re-checked at the top, motivational level of the hierarchy between every arbitration step, and the entire hierarchy is revisited on the termination of any subtask [8].

4 Mapping CAA Features to Mammal Brain Structure

To summarize the last section, complete agent architectures have converged on three sorts of architectural modules in order to support complex, reactive behavior. First, skill modules — a functional decomposition of intelligent ability into skilled actions and their supporting specialized perception and memory. Second, hierarchical structures that support action selection, often known as reactive plans. These structures are used to focus attention on behaviors likely to be useful in a particular circumstance and provide temporal ordering for behavior. And third, an environment-monitoring or alarm system for switching the focus of action-selection attention in response to highly salient environmental events.

If this sort of organization is necessary or at least very useful for intelligent control, then it is also likely to be reflected in the organization of animal intelligence. As we explained in the introduction, animals have evolved to face similar problems of information management. In this section, we look at how these CAA principles relate to what is known of mammal brain architecture.

4.1 Skill Modules

In Section 2 we discussed modularity in mammalian brains. Using that terminology, we consider the skill modules of CAA to correspond roughly to functional modularity, particularly in the neocortex, and perhaps to some extent to temporal modularity. However, there is no direct correlation between brain modularity and CAA skill modules. For example, a skill module for grasping a visual target must incorporate the retinas, visual cortex, associative cortices, motor pre-planning and motor coordination. It would also need to exploit somatic and proprioceptive feedback from the grasping limb, though some of this complexity might be masked by interfacing to other specialist modules.

The reason there is not a direct correlation between CAA skill modules to mammalian functional modules is because CAA modules tend to be end-to-end. That is, they encapsulate both perception and action. Much of functional cortical modularity in mammals tends to be more general, for example the visual, auditory, somatic and motor cortices. Some of the temporal modularity in the parietal cortex and the hippocampal formation may correspond more directly to CAA modularity, but this lacks the parallelism typically recommended in CAA, and also the localized representation of perception and motor skills. Taking these sorts of context-specific parietal and hippocampal representations along with some supporting perception and action representations from other cortical areas is probably the best approximation of the typical CAA skill module.

Since a concern of this volume is modeling neural modularity in ANN, we should point out that even though CAA skill modules aren't typically analogous to cortical regions, this is for reasons of practicality. Within the agent discipline, modularity in CAA is primarily to support an orderly decomposition of intelligence into manageable, constructible units. But if one is more interested in modeling the brain directly, one could well use known or theorized cortical modularity as a blueprint for skill decomposition in a CAA agent [e.g. 63]. We would also like to note that more primitive neural systems such as those found in insects to some extent lack the generality of mammalian brains, and more closely match common CAA modularity. For example, spiders have multiple pairs of eyes, some of which seem to be dedicated to single skill modules, such as mating [37].

4.2 Action Selection

The basal ganglia has been proposed as the organ responsible for at least some aspects of action selection [27, 43, 49, 51]. In a distributed parallel model of intelligence, one of the main functions of action selection is to arbitrate between different competing behaviors. This process must take into account both the activation level of the various 'input' cortical channels and previous experience in the current or related action-selection contexts.

The basal ganglia is a group of functionally related structures in the forebrain, diencephalon and midbrain. Its main 'output' centers — parts of the substantia nigra, ventral tegmental area, and pallidum — send inhibitory signals to neural centers throughout the brain which either directly or indirectly control voluntary movement, as well as other cognitive and sensory systems [42]. Its 'input' comes through the striatum from relevant subsystems in both the brainstem and the forebrain. Prescott et al. [49] have proposed a model of this system whereby it performs action selection similar to that proven useful in CAA architectures.

Arbitrating between subsystems is only part of the problem of action selection. Action patterns must also be sequenced with appropriate durations to each step. The duration of many actions is too quick and intricate to be monitored via feedback, or left to the vagaries of spreading activation from competing but unrelated systems [32, 38]. Further, animals that have had their forebrains surgically removed have been shown capable of conducting complex species-typical behaviors — they are simply unable to apply these behaviors in appropriate contexts. In particular, the periaqueductal grey matter has been implicated in complex species-typical behaviors such as mating rituals and predatory, defensive and maternal maneuvers [39]. However, there appears to be little literature as to exactly how such skills are coordinated. There is also little evidence that learned skills would be stored in such areas. We do know that several cortical areas are involved in recognizing the appropriate context for stored motor skills [e.g. 1, 57]. Such cortical involvement could be part of the interface between skill modules and action selection (see further [8].)

4.3 Environment Monitoring

Our proposal for the the mammalian equivalent to the environment monitoring and alarm systems is more straight-forward. It is well established that the limbic system, particularly the amygdala and associated nuclei, is responsible for triggering emotional responses to salient (particularly dangerous, but also reproductively significant) environmental stimuli. Emotional responses are ways of creating large-scale context shifts in the entire brain, including particularly shifts in attention and likely behavior [15, 18]. This can be in response either to basic perceptual stimuli, such as loud noises or rapidly looming objects in the visual field, or to complex cortical perceptions, such as recognizing particular people or situations [15]. Again, there can be no claim that this system is fully understood, but it does, appropriately, send information to both the striatum and the periaqueductal grey. Thus the amygdalic system meets our criteria for an alarm system being interconnected with action selection, as well as biasing cortical / skill-module activation.

4.4 Conclusion

In conclusion, it is difficult to produce an completely convincing mapping of CAA attributes to neural subsystems, primarily because the workings of neural subsystems are only beginning to be understood, but also because the levels of generalization are not entirely compatible. The primary function of a CAA architecture is to facilitate a programmer in developing an agent. Consequently, complexity is kept to a minimum, and encapsulation is maximized. Evolution, on the other hand, will eagerly overload an architectural module that has particular computational strengths with a large number of different functions. Nevertheless, we have identified several theories from neuroscience that are analogous to the features of CAA.

5 Adaptivity in Modular Systems

The modularity of an intelligence has obvious ramifications for learning. In particular, the generic CAA architecture described in Section 3, having three sorts of elements, has three sorts of adaptability.

First and foremost, there is adaptivity within the skill modules or behaviors. A behavior module must have the ability to represent perceptual experience in a way that provides for mapping motivation to appropriate actions. In other words, adaptive perceptual state *defines* what a behavior is and does. Specialized learning is ubiquitous in nature, to the point that leading researchers consider it by far the dominant form of natural adaptivity [23, 52]. Such attributes are naturally modeled in CAA under behavior oriented design [9].

Given that action selection requires structure, a natural extension of the CAA systems described above would allow the agent to learn new reactive plans. There are at least three means by which this could be done. The most commonly attempted in AI is by *constructive planning*. This is the process whereby plans are created by searching for sets of primitives which, when applied in a particular order to the current situation, would result in a particular goal situation [e.g. 21, 62]. Another kind of search that has been proposed but not seriously demonstrated is using a genetic algorithm (GA) or GA-like approach to combine or mutate existing plans [e.g. 14]. Another means of learning plans is to acquire them socially, from other, more knowledgeable agents.

Constructive planning is the most intuitively obvious source of a plan, at least in our culture. However, this intuition probably tells us more about what our consciousness spends time doing than about how we actually acquire most of our behavior patterns. The capacity for constructive planning is an essential feature of Soar and of most three-layer-architectures; however it is one that is still underutilized in practice. We suspect this will always be the case, as it will be for GA type models of "thinking", because of the combinatoric difficulties of planning and search [16]. Winston [66] states that learning can only take place when one nearly knows the answer already: this is certainly true of learning plans. Search-like algorithms for planning in real-time agents can only work in highly constrained situations, among a set of likely solutions.

Social or mimetic learning addresses this problem of constraining possible solutions. Observing the actions of another intelligent agent provides the necessary bias. This may be as simple as a mother animal leading her children to a location where they are likely to find food, or as complex as the imitation of complex, hierarchical behavioral patterns (in our terminology, plans) [13, 64]. This may not seem a particularly promising way to increase intelligence, since the agent can only learn what is present in its society, but this is not the case. First, since an agent uses its own intelligence to find the solution within some particular confines, it may enhance the solution it is being presented with. This is famously the case when young language learners regularize constructed languages [3, 33]. Secondly, a communicating culture may well contain more intelligence than any individual member of it, leading to the notion of cultural evolution and mimetics [20]. Thus although the use of social learning in AI is only beginning to be explored [e.g. 53], we believe it will be an important capacity of future artificial agents.

Finally, we have the problem of learning *new* functional and/or skill modules. Although there are many PhD theses on this topic [a good recent example is 19], in the taxonomy presented in this paper, most such efforts would fall under the parameter learning for a single skill module or behavior. Learning full new representations and algorithms for actions is beyond the current state of the art for machine learning. Such a system would almost certainly have to be built on top of a fine-grain distributed representation — essentially it should be an ANN. However, again, the state of the art in ANN does not allow for the learning and representation of such complex and diverse modules.

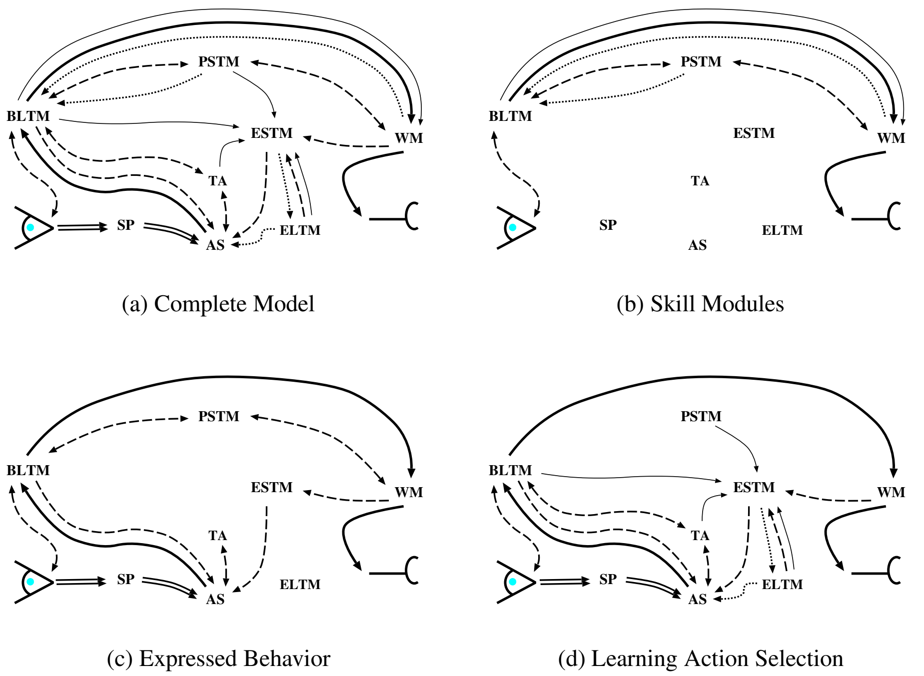
6 Requirements for a Behavior Learning System

If the current state of the art were not an obstacle, what would a system capable of *all three* forms of adaptivity described in the previous section look like? We think it would require at minimum the elements shown in Figure 3(a). In this section, we will explain this model.

Consider first the behavior or skill module system (Figure 3(b)). The representation of the CAA skill modules has been split into two functional modules: the Behavior Long Term Memory (BLTM) and the Perceptual Short Term Memory (PSTM). The persistent representation of the skill modules' representations and algorithms belong in the former, the current perceptual memory in the latter. There is further a Working Memory (WM) where the representation of the behaviors from the BLTM can be modified to current conditions, for example compensating for tiredness or high wind. In a neurological model, some of these representations might overlap each other in the same organs, for example in different networks within the neocortex or the cerebellum. As in BBAI, the skill modules contain both perception and action, though notice the bidirectional arrows indicating expectation setting for perception.

The full path for expressed action is shown in Figure 3(c). This takes into account both standard action selection and environment monitoring. Here, with the learning arcs removed, we can see recommendations flowing from the behavior system to action selection (AS). Action selection also takes into account timing provided by a time accumulator (TA, see below) and recent action selections (decisions) stored in episodic short term memory (ESTM). Expressed action takes into account current perceptual information in PSTM as well as the current modulated version of the behaviors in WM.

We have also provided a separate path for basic perceptual reflexes such as alarm at loud noises or sudden visual looming. The module for recognizing these effects is labeled SP for Special Perception. In nature this system also has connections to the cortical system, so that reflexive fear responses can be developed for complex stimuli,



(a) Complete Model

(b) Skill Modules

(c) Expressed Behavior

(d) Learning Action Selection

Fig. 3. An architecture for allowing adaptation within skill modules, of new plans, and of new skill modules. Icons for sensing and action are on the lower left and right respectively. Dashed lines show the flow of information during an active, attending system. Dotted lines are pathways for consolidation and learning. The heavy solid line is the path of expressed behavior; the double line represents the constant perceptual pathway for environmental alerts. The fine lines indicate references: the system pointed to references representations in the system pointed from.

but this capacity is not necessary for a minimal fully-learning CAA configuration. It is, however, necessary to isolate the fundamental system from possible modification by the skill module learning system.

To make action selection adaptive (Figure 3(d)) we provide first a time accumulator (TA) as proposed by Pöppel [48] and Henson [29] and episodic short term memory (ESTM) as postulated by a large number of researchers (see [41] for experiments and review.) Episodic long term memory (ELTM) is included for good measure — as consolidated experience, it might also represent other forms of semantic memory, or it might actually be homologous with BLTM.

Finally, in keeping with [41, 58], this model assumes that many of the modules make reference to the state of other modules rather than maintaining complete descriptions themselves. This is considered an important attribute of any system which needs to hold a large number of things which are learned very quickly, because it allows for a relatively small amount of state. Such reference is considered important in computer science as a means to reduce the probability of conflicting data sets, and is also a likely feature of evolved systems, where existing organization is often exploited by a variety of means.

7 Future Directions: From Neuroscience to CAA

Fortunately, we do not expect that implementing such a complex system is necessary for most CAA applications. In general, the adaptive needs of the agent can be anticipated in advance by the designer, or discovered and implemented during the process of developing the agent. We do, however, suspect that some of the systems being discovered and explored in neuroscience may soon become standard functional modules in CAA, in the same way that action selection and alarm systems are now. We finish our chapter with some lessons from neuroscience which might further CAA, and hopefully therefore indirectly benefit the ANN community.

We expect that one of the capacities often ascribed to the hindbrain, that of smoothing behavior, should be broken into a separate module. This allows modules that create motor plans to operate at a relatively coarse granularity. It also allows for the combination of influences from multiple modules and the current situation of the agent without complicating those skill modules. The only CAA architecture we know that explicitly has such a unit is Ymir [59], where the Action Scheduler selects the most efficacious way to express messages given the agent's current occupations. For example, if the agent has decided "cognitively" to agree with something, but it is currently engaged in listening rather than speaking, it may nod its head rather than say "yes" and interrupt the other speaker. This sort of capacity is also present in a number of new AI graphics packages which allow for the generation of smooth images from a script of discrete events [2, 4]. The fact that such work is not yet seen in robotics may be partially due to the fact that a physical agent can take advantage of physics and mechanics to do much of its smoothing [11], but as robots attempt more complex feats such as balancing on two legs, providing for smoothed or balanced motions may well deserve dedicated modules or models similar to those cited above.

We also expect that having comprehensive but sparsely represented records of episodic events will become a standard mechanism. Episodic records are useful for complimenting and simplifying reactive plans by recording state about previous attempts and actions, thus reducing the chance that an agent may show inappropriate perseveration or redundancy when trying to solve a problem. Further, as mentioned previously, episodic memory can be a good source for consolidating semantic information, such as noticing regularities in the environment or the agent's own performance. These records can in turn be used by specialized learning systems for particular problems, even if a full-blown skill learning system has not been implemented.

Many researchers are currently working on emotion modules for CAA. We remain skeptical of the need for an independent emotion module for two reasons. First, there is a great deal of evidence that the basic emotions evolved independently at different times in our history. This suggests that a single emotion module might not be appropriate. Second, emotions are intimately involved in action selection. In fact, Damasio [18] implies that any species-typical behavior pattern is effectively an emotional response.

This suggests that it is impossible to separate emotions from motivation in action selection.

8 Conclusions

In this chapter we have attempted to further the advance of ANN by unifying it and neuroscience with another branch of artificial intelligence: complete autonomous agents. We have described both the utility and costs of modularity, and explained the sorts of mechanisms the CAA field has found necessary to work with these systems. We have discussed how these mechanisms relate to neuroscience, and in turn have suggested how current advances in neuroscience might further the field of CAA. We have also proposed a complex model for a modular agent intelligence capable of true mammal-like learning, which would rely heavily on ANN representations. ANN, with its distributed representations for machine learning, is the most promising field for developing a system capable of developing its own representations and algorithms. Unfortunately, to date such learning is, as far as we know, beyond the state of the art.

Although our future work section focussed on our own area, CAA, we hope that the advances in this area of AI can also contribute to the advance of neural networks research. We expect that a modular ANN would require both similar types of modules and similar interfaces between them as we have described here. Further, we hope that as the field of CAA advances, it can become as useful a tool for helping brain scientists think about and model the sorts of representations and interactions they are attempting to understand.

Acknowledgements

Thanks to Will Lowe for his comments and suggestions.

References

[1] W. F. Asaad, G. Rainer, and E. K. Miller. Task-specific neural activity in the primate prefrontal cortex. *Journal of Neurophysiology*, 84:451–459, 2000.

[2] A. Baumberg and D. C. Hogg. Generating spatio-temporal models from examples. *Image and Vision Computing*, 14(8):525–532, 1996.

[3] Derek Bickerton. *Language & Species*. The University of Chicago Press, Chicago, Illinois, 1987.

[4] Matthew Brand, Nuria Oliver, and Alex Pentland. Coupled hidden markov models for complex action recognition. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR97)*, 1997.

[5] Rodney A. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, RA-2:14–23, April 1986.

[6] Rodney A. Brooks. Intelligence without reason. In *Proceedings of the 1991 International Joint Conference on Artificial Intelligence*, pages 569–595, Sydney, August 1991.

[7] Joanna Bryson. Cross-paradigm analysis of autonomous agent architecture. *Journal of Experimental and Theoretical Artificial Intelligence*, 12(2):165–190, 2000.

[8] Joanna Bryson. Hierarchy and sequence vs. full parallelism in reactive action selection architectures. In *From Animals to Animats 6 (SAB00)*, pages 147–156, Cambridge, MA, 2000. MIT Press.

[9] Joanna Bryson. Making modularity work: Combining memory systems and intelligent processes in a dialog agent. In Aaron Sloman, editor, *AISB'00 Symposium on Designing a Functioning Mind*, pages 21–30, 2000.

[10] Joanna Bryson. The study of sequential and hierarchical organisation of behaviour via artificial mechanisms of action selection, 2000. MPhil Thesis, University of Edinburgh.

[11] Joanna Bryson and Brendan McGonigle. Agent architecture as object oriented design. In Munindar P. Singh, Anand S. Rao, and Michael J. Wooldridge, editors, *The Fourth International Workshop on Agent Theories, Architectures, and Languages (ATAL97)*, pages 15–30, Providence, RI, 1998. Springer.

[12] Joanna Bryson and Lynn Andrea Stein. Architectures and idioms: Making progress in agent design. In C. Castelfranchi and Y. Lespérance, editors, *The Seventh International Workshop on Agent Theories, Architectures, and Languages (ATAL2000)*. Springer, 2000. *in press*.

[13] Richard W. Byrne and Anne E. Russon. Learning by imitation: a hierarchical approach. *Brain and Behavioral Sciences*, 21(5):667–721, 1998.

[14] William H. Calvin. *The Cerebral Code*. MIT Press, 1996.

[15] Niel R. Carlson. *Physiology of Behavior*. Allyn and Bacon, Boston, 2000.

[16] David Chapman. Planning for conjunctive goals. *Artificial Intelligence*, 32:333–378, 1987.

[17] Michael H. Coen. Building brains for rooms: Designing distributed software agents. In *Proceedings of the Ninth Conference on Innovative Applications of Artificial Intelligence (IAAI97)*, Providence, RI, 1997.

[18] Antonio R. Damasio. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Harcourt, 1999.

[19] John Demiris. *Movement Imitation Mechanisms in Robots and Humans*. PhD thesis, University of Edinburgh, May 1999. Department of Artificial Intelligence.

[20] Daniel Dennett. *Darwin's Dangerous Idea*. Penguin, 1995.

[21] Richard E. Fikes, Peter E. Hart, and Nils J. Nilsson. Learning and executing generalized robot plans. *Artificial Intelligence*, 3:251–288, 1972.

[22] James Firby. An investigation into reactive planning in complex domains. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 202–207, 1987.

[23] C.R. Gallistel, Ann L. Brown, Susan Carey, Rochel Gelman, and Frank C. Keil. Lessons from animal learning for the study of cognitive development. In Susan Carey and Rochel Gelman, editors, *The Epigenesis of Mind*, pages 3–36. Lawrence Erlbaum, Hillsdale, NJ, 1991.

[24] M. P. Georgeff and A. L. Lansky. Reactive reasoning and planning. In *Proceedings of the Sixth National Conference on Artificial Intelligence (AAAI-87)*, pages 677–682, Seattle, WA, 1987.

[25] Henry Gleitman. *Psychology*. Norton, 4 edition, 1995.

[26] Stephen Grossberg. How does the cerebral cortex work? learning, attention and grouping by the laminar circuits of visual cortex. *Spatial Vision*, 12:163–186, 1999.

[27] Kevin Gurney, Tony J. Prescott, and Peter Redgrave. The basal ganglia viewed as an action selection device. In *The Proceedings of the International Conference on Artificial Neural Networks*, Skovde, Sweden, September 1998.

[28] D.O. Hebb. *The Organization of Behavior*. John Wiley and Sons, New York, New York, 1949.

[29] Richard N. A. Henson. *Short-term Memory for Serial Order*. PhD thesis, University of Cambridge, November 1996. St. John's College.

[30] Henry Hexmoor, Ian Horswill, and David Kortenkamp. Special issue: Software architectures for hardware agents. *Journal of Experimental & Theoretical Artificial Intelligence*, 9(2/3):147–156, 1997.

[31] Berthald K. P. Horn and Patrick H. Winston. A laboratory environment for applications oriented vision and manipulation. Technical Report 365, MIT AI Laboratory, 1976.

[32] George Houghton and Tom Hartley. Parallel models of serial behavior: Lashley revisited. *PSYCHE*, 2(25), February 1995.

[33] Simon Kirby. *Function, Selection and Innateness: the Emergence of Language Universals*. Oxford University Press, 1999.

[34] T. Kobayashi, H. Nishijo, M. Fukuda, J. Bures, and T. Ono. Task-dependent representations in rat hippocampal place neurons. *JOURNAL OF NEUROPHYSIOLOGY*, 78(2):597–613, 1997.

[35] David Kortenkamp, R. Peter Bonasso, and Robin Murphy, editors. *Artificial Intelligence and Mobile Robots: Case Studies of Successful Robot Systems*. MIT Press, Cambridge, MA, 1998.

[36] John E. Laird and Paul S. Rosenbloom. The evolution of the Soar cognitive architecture. Technical Report CSE-TR-219-94, Department of EE & CS, University of Michigan, Ann Arbor, September 1994. also in *Mind Matters*, Steier and Mitchell, eds.

[37] M.F. Land. Mechanisms of orientation and pattern recognition by jumping spiders (salticidae). In R. Wehner, editor, *Information Processing in the Visual Systems of Arthropods*, pages 231–247. Springer-Verlag, 1972.

[38] K. S. Lashley. The problem of serial order in behavior. In L. A. Jeffress, editor, *Cerebral mechanisms in behavior*. John Wiley & Sons, New York, 1951.

[39] Joseph S. Lonstein and Judith M. Stern. Role of the midbrain periaqueductal gray in maternal nurturance and aggression: *c-fos* and electrolytic lesion studies in lactating rats. *Journal of Neuroscience*, 17(9):3364–78, May 1 1997.

[40] Chris Malcolm, Tim Smithers, and John Hallam. An emerging paradigm in robot architecture. In *Proceedings of IAS*, volume 2, Amsterdam, Netherlands, 1989.

[41] James L. McClelland, Bruce L. McNaughton, and Randall C. O'Reilly. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3):419–457, 1995.

[42] Frank A. Middleton and Peter L. Strick. Basal ganglia output and cognition: evidence from anatomical, behavioral, and clinical studies. *Brain and Cognition*, 42 (2):183–200, 2000.

[43] Jonathon W. Mink. The basal ganglia: focused selection and inhibition of competing motor programs. *Progress In Neurobiology*, 50(4):381–425, 1996.

[44] Hans P. Moravec. The stanford cart and the CMU rover. In I. J. Cox and G. T. Wilfong, editors, *Autonomous Robot Vehicles*, pages 407–419. Springer, 1990.

[45] O. Nestares and David J. Heeger. Robust multiresolution alignment of MRI brain volumes. *Magnetic Resonance in Medicine*, 43:705–715, 2000.

[46] Nils Nilsson. Shakey the robot. Technical note 323, SRI International, Menlo Park, California, April 1984.

[47] Nils Nilsson. Teleo-reactive programs for agent control. *Journal of Artificial Intelligence Research*, 1:139–158, 1994.

[48] E. Pöppel. Temporal mechanisms in perception. *International Review of Neurobiology*, 37:185–202, 1994.

[49] Tony J. Prescott, Kevin Gurney, F. Montes Gonzalez, and Peter Redgrave. The evolution of action selection. In David McFarland and O. Holland, editors, *Towards the Whole Iguana*. MIT Press, Cambridge, MA, to appear.

[50] V. S. Ramachandran and S. Blakeslee. *Phantoms in the brain: Human nature and the architecture of the mind*. Fourth Estate, London, 1998.

[51] Peter Redgrave, Tony J. Prescott, and Kevin Gurney. The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience*, 89:1009–1023, 1999.

[52] T. J. Roper. Learning as a biological phenomena. In T. R. Halliday and P. J. B. Slater, editors, *Genes, Development and Learning*, volume 3 of *Animal Behaviour*, chapter 6, pages 178–212. Blackwell Scientific Publications, Oxford, 1983.

[53] Stefan Schaal. Is imitation learning the route to humanoid robots? *Trends in Cognitive Sciences*, 3(6):233–242, 1999.

[54] Phoebe Sengers. Do the thing right: An architecture for action expression. In Katia P Sycara and Michael Wooldridge, editors, *Proceedings of the Second International Conference on Autonomous Agents*, pages 24–31. ACM Press, 1998.

[55] Phoebe Sengers. *Anti-Boxology: Agent Design in Cultural Context*. PhD thesis, School of Computer Science, Carnegie Mellon University, 1999.

[56] M. Sur, A. Angelucci, and J. Sharma. Rewiring cortex: The role of patterned activity in development and plasticity of neocortical circuits. *Journal of Neurobiology*, 41:33–43, 1999.

[57] J. Tanji and K. Shima. Role for supplementary motor area cells in planning several movements ahead. *Nature*, 371:413–416, 1994.

[58] T. J. Teyler and P. Discenna. The hippocampal memory indexing theory. *Behavioral Neuroscience*, 100:147–154, 1986.

[59] Kristinn R. Thórisson. A mind model for multimodal communicative creatures & humanoids. *International Journal of Applied Artificial Intelligence*, 13(4/5): 519–538, 1999.

[60] Alan M. Turing. Computing machinery and intelligence. *Mind*, 59:433–460, October 1950.

[61] Gerhard Weiss, editor. *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*. MIT Press, Cambridge, MA, 1999.

[62] Daniel S. Weld. Recent advances in AI planning. *AI Magazine*, 20(2):93–123, 1999.

[63] Michael Wessler. A modular visual tracking system. Master's thesis, MIT, 1995. Artificial Intelligence Laboratory.

[64] Andrew Whiten. Primate culture and social learning. *Cognitive Science*, 24, 2000. in press.

[65] S. I. Wiener. Spatial, behavioral and sensory correlates of hippocampal CA1 complex spike cell activity: Implications for information processing functions. *Progress in Neurobiology*, 49(4):335, 1996.

[66] Patrick Winston. Learning structural descriptions from examples. In Patrick Winston, editor, *The Psychology of Computer Vision*. McGraw-Hill Book Company, New York, 1975.