A Summary of the Attention and Affect Project

L. Beaudoin, C. Paterson, T. Read, E. Shing, A. Sloman & I. Wright

November 1993

Contents

This document is intended as a brief summary of the attention and affect project. A historical approach is taken - the original objectives of the project are reviewed, followed by an overview of what has been achieved up to now, and then the future direction of the project is extrapolated. The conclusion contains an immanent criticism i.e. the project is judged in terms of the original aims.

1 The Problem

1.1 The original objectives

The project was envisaged as an application of the design-based approach to investigate cognitive processes. The design-based approach is a movement from specifying requirements to developing designs and implementations that satisfy these requirements. In general a mixture of strategies is used - top-down, bottom-up and middle-out - with the aim of providing richer theories of explanation. With this fundamental theoretical stance in mind, the project, as originally conceived, was concerned with developing functional requirements for the architecture of intelligent agents situated in dynamic environments with multiple, asynchronously generated motivators. This is a huge problem area that generalizes to all branches of artificial intelligence. Therefore, to reduce the problem area designs were to be restricted to a 'coarse grained', global level i.e. the project would concentrate on how mechanisms tessellate, not on the inner details of the mechanisms. This constraint was justified by the assertion that 'architecture dominates mechanism' [Sloman 1993]. And any architectures that would be developed were to be 'broad but shallow'[1].

More specific objectives of the project were -

- To concentrate on the functional requirements needed for the processing of motivators and control of attention.
- To explore the design-space of possible architectures that fulfill the requirements; and to analyse different kinds of resource limits and their implications.
- To implement a sequence of designs capable of fulfilling increasing subsets of the requirements; and to test these designs in a domain providing adequate environmental richness.
- To provide an architecturally grounded conceptual framework for describing affective states involving control of attention, such as desires, moods, attitudes, obsessions etc [Sloman 1992].
- To design effective interfaces for demonstrating the key features of the models, and to enable them to be used for teaching.

[1] This phrase first used by Joe Bates at CMU.

[Sloman and Humphreys 1992]

1.2 The domain

To fulfil the project's aim of providing an environment of sufficient complexity for a situated agent the nursemaid domain was proposed [Beaudoin and Sloman 1993]. The domain is a two-dimensional nursery consisting of a number of rooms connected by doors. The nursery is populated by robot babies that wander around and get into various difficulties, such as falling into fatal ditches, running out of battery charge and getting progressively broken. The simulated agent, the nursemaid, can move a camera around the rooms to detect problems and has the use of a claw in which to effect change within the nursery. For example, if a baby is running out of charge she can pick it up and transport it to a battery recharge point. The multiplicity of babies within the nursery ensures that the world is constantly changing and that new motivators are continuously generated. The domain could be extended in several directions to add more requirements and constraints.

2 Current Progress

What follows is a summary of the attention and affect project as it stands today. The summary is *not* intended to be exhaustive and takes a high level view of the extant work.

2.1 General theoretical gains

2.1.1 From requirements to design

The movement from requirements to design is an indeterminate and often arbitrary process. The specific case of designing an autonomous agent from a set of abstract requirements is no different. In order to overcome this, an attempt has been made to formalize the movement. Six types of design decision have been enumerated [Sloman et al. 1992].

- Decisions made according to the design requirements (high-level requirements which are features of the project's objective)
- Decisions made due to constraints which are the results of earlier design decisions regarding other parts of the overall architecture (or from a previous design/prototype)
- Decisions arrived at according to empirical data (e.g psychological); i.e. decisions made to copy the real world (as it actually is)
- Decisions made due to hardware/software limitations (constrained by tools)

- Decisions which are essentially arbitrary (requiring experimentation with the various options)
- Decisions made to test a theory (where no/little empirical evidence exists)

In addition, the design-based approach has been extended (see section 2.4).

2.1.2 Terminological pitfalls

The Attention and Affect project, due to its inter-disciplinary nature, has encountered the problem of terminological confusion. Terms such as 'emotion', 'motivation', 'cognition' and 'attention' are often left undefined or refer to different phenomena [Read and Sloman 1993]. But the problems don't end there. Even when definitions of terms are given they can be simple and arbitrary, and not stated in a form that can be subjected to testing. These problems can significantly slow the research endeavour and lead to unnecessary argument.

A theoretical gain of the Attention and Affect project is that it has highlighted these problems and taken the first steps in relating how the terminological pitfalls can be avoided.

A distinction has had to be made between the conceptual space of mechanisms and behaviours that is to be explored (what we are interested in explaining), and how this space is referred to (the terminology we use to describe it). Good terminology will be grounded in the former space and will arise from a complete or near-complete theory of underlying mechanisms and behaviours. At the present time, ascriptions of terms such as 'happy' or 'sad' to either i. other organisms or ii. simulated agents will neither further our understanding nor unpack colloquial concepts that undoubtedly refer to highly complex phenomena. In addition, concepts approach the concrete only via the interaction of theory and practice. In other words, definitions will tend to be arbitrary and overly abstract until we have a good explanatory theory that is testable. Until this time we must be strict in the use of terminology. For example, a term such as *insistence*[2] has meaning only in terms of a specified virtual machine architecture, but for it to be part of a multi-layer theory it would need to be mapped on to the neural level. Presently ill-defined concepts, such as 'attention', are likely to expand into a network of related phenomena at different virtual machine levels, each level having its own specific (and precise) terminology. This stance can be summed up as: theory first, definitions later; and until that time - watch your language[3].

[2] See section 2.2.

[3] However, it is possible to have a partial specification of a concept which will be refined as the theory develops. E.g. "attention" – or its technical replacement – could be analysed in terms of a collection of types of selections, the precise classes of selections to be determined later, but including: selecting between alternative sets of input data, selecting between different ways of processing the data, selecting between different potential goals (deciding which should become intentions), selecting between different current but suspended plans, selecting between different unsolved problems, etc.

2.2 NML1 - an architecture to simulate motive processing

NML1 is an architecture to simulate motive processing [Beaudoin 1993a]. As it stands, this architecture remains only a framework, with management processes detailed in terms of functional requirements and a coarse specification of their form [Beaudoin 1993b]. The original design and conception of NML1 has been progressively refined through prototype implementation. But code still lags behind theory.

The principal features of NML1's architecture can be split into 3 components.

- Generation and activation of motivators
- Management processes
- Meta-Management

There are two kinds of motivator generation in NML1 [Beaudoin 1993c]. A goal[4] may be created during a high level expansion of a solution to a problem; or by mechanisms running asynchronously to high level processes, responding to activation conditions in the environment, that create new goals or reactivate suppressed goals, a process known as generactivation. The notion of suppression here is important. One of the tenets of the theory driving the project is that a resource limited agent will need to protect high level processes from interruption or distraction from relatively unimportant motivators [Sloman 1987]. In NML1 a filtering mechanism has been designed to implement this requirement. It allows a goal to surface if its *insistence*, a heuristic measure of its urgency and importance, is higher than the filter threshold or if the goal descriptor unifies with a filter structure. In other words, both qualitative and quantitative goal surfacing occurs, providing greater reactivity to the environment.[5] If the goal fails to surface it is suppressed and placed in a short-term store. If a goal surfaces it is considered by higher level processes, which are the goal management processes.

When a goal surfaces it is managed. Management processes determine whether, when, and how motivators are to be executed. In other words, there are three principal functions of m-processes.

- Deciding
  - determine whether a motivator is to be adopted as an intention.
  - determine the extent to which the motivator should be satisfied.
  - needs to take into account the importance, cost and likelihood of success of the motivator.

As the architecture becomes more precisely specified it will be possible to fill out in more detail what is meant by this notion of attention, as the range of options requiring selection become clearer.

[4] A goal is a type of motivator, and a motivator is a type of control state [Beaudoin 1993d].

[5] For example, the knee-jerk reflex in humans could be viewed as qualitative motivator surfacing.

- Scheduling
  - determine temporal ordering of action execution.
  - needs to take into account the urgency of the motivator.
- Expanding
  - determine a possible course of action(s) to satisfy the motivator.
  - needs to take into account beliefs concerning the efficacy and consequences of actions.

It has been identified that the above three aspects of m-processing[6] can only be achieved in conjunction with certain auxiliary functions, such as the ability to predict the outcome of hypothetical decisions i.e. the capacity to 'think ahead', assess motivators and the plans that satisfy the motivators, and assess the overall context or situation. However, prediction and assessment are outstanding problems in NML1 [Beaudoin 1993b].

In addition, the m-processes themselves must be managed to avoid conflicts and over-long deliberation of motivators. This is meta-management.

> Meta-management is concerned with dynamically assuring that the m-processes are used judiciously. This may involve selecting which m-processes to assign to which motivators and when to execute particular processes [Beaudoin 1993a].

The distinction between management and meta-management within the architecture is blurred and needs further clarification. However, an aspect of meta-management is detecting problems in the current state of processing, such as perturbances[7], digressions[8], maundering[9], hastiness[10] and oscillating between decisions. Once detected, meta-management should instigate coping strategies. However, detecting global states such as oscillating between decisions remains problematical.

Management and meta-management are the 'brains' of an autonomous agent; but work has not yet progressed to the stage where it can be detailed *how* management of motivators is to work.

[6] There are other management functions waiting to be explored, including, for example, deciding when to abandon a goal, deciding when to re-assess a current plan in the light of new evidence, etc. Also there are kinds of theoretical problem solving and other intellectual activities that an intelligent agent has to interleave with these processes.

[7] A perturbance is a state where a highly insistent motivator keeps resurfacing, disrupting high level processing.

[8] A digression is a state where deliberation of a motivator is pursued for its own sake without returning a decision in a timely manner.

[9] Maundering is a state where a decision to meta-manage a motivator is postponed indefinitely.

[10] Hastiness is the extent to which there are important, urgent, and adopted unsatisfied (but potentially satisfiable) motivators that require management and/or action.

2.3 Computational mechanisms for constraining learning

Learning is essential if an intelligent agent is to cope with multiple, asynchronous motivators arising from a dynamic environment [Shing 1993a]. Without learning an agent will react predictably and rigidly to novel events. However, there is a problem: the more complex the world is the greater the number of input features that can be associated by a learning system. This leads to a combinatorial explosion.

Two hypotheses have been put forward. First, that for useful learning to occur, selection (at various levels in the cognitive architecture) is necessary in order to prune the 'learning space'. For example, processes that monitor the environment (both internal and external), select salient features and evaluate their importance will be able to ignore 'junk' in the event stream. Second, that complex adaptive behaviours will emerge from relatively simple reinforcement learning mechanisms coupled with selection mechanisms.

In order to test these hypotheses reinforcement learning will be simulated within the nursery domain. The nursery domain is particularly suited to this task as it can provide a high number of events that can be associated to an unconditioned response in constrast to traditional machine learning work which has concentrated on small, predictable domains [Shing 1993b].

Examples of types of reinforcement learning are as follows.

- Classical
  - Primary: This is the case of Pavlov's dogs. A bell is rung every time food (the unconditioned stimulus) is presented. After repeated trials the dogs will salivate when the bell is rung without the presence of food. In other words, the ringing bell has become a conditioned stimulus.
  - Seondary: This is the case where sounding a buzzer before the bell is rung (in the case of the Pavlovian dogs) will cause association between the buzzer and food.
- Instrumental
  - This is a kind of trial and error learning. The agent's random actions will cause it to learn. For example, a pidgeon may peck at a lever and receive grain - it will associate this action with receiving food.

Two points need to be added about reinforcement learning. The unconditioned and conditioned stimuli need not occur at the same time or even within a short time interval of each other. A stimulus that occurred in the past can still become associated with an unconditioned stimulus. Also, the degree of learning is proportional to the 'surprisingness' of the stimulus, *cateris paribus.*

An example of reinforcement learning within the nursery domain could be as follows. BabyA falls into the ditch causing it to lose charge and scream. The nursemaid focuses its visual input on BabyA and sees that its colour is red (indicating a low charge). The baby dies causing an unconditioned stimulus—a 'shock' to the nursemaid—to be administered as the nursemaid is innately programmed to dislike the death of one of its charges. The nursemaid must now associate stimuli in the environment, such as the baby screaming or its colour etc., with the death of the baby (the unconditioned stimulus). In a complex environment there will be many such stimuli to be associated. This is where selection mechanisms are needed. Four heuristic selection mechanisms for learning will be implemented.

- Similarity
  - associations formed in the past that are similar to the current situation will be reinforced.
- Unusualness/Novelty
  - unusual stimuli will tend to be associated in preference to others.
- Temporal contiguity
  - stimuli that occur at the same time as the unconditioned stimulus will be associated in preference to others.
- Forward causation
  - the simplifying assumption that stimuli that should be associated with the unconditioned stimulus will occur before the unconditioned stimulus ie. cause precedes effects.[11]

In addition to the above heuristics, other 'selection' mechanisms will reduce the size of 'learning space', such as a visual guidance system and short term memory. For example, a short term memory store will place an upper limit on the past stimuli history the agent will be able to consider for association.

With the above mechanisms implemented it is hoped that the nursemaid will produce new behaviours when new agents or objects are introduced into the nursery i.e. the nursemaid should be able to recognise the new entity and to associate positive or negative goal-relevant world changes with it. Also, there should be some form of emergent learning—the combination of selection, monitoring and evaluation processes coupled with associative reinforcement learning should produce unexpected behaviours. For example, secondary learning (the nursemaid may learn to associate

[11] Effect may appear to precede cause if the agent fails to notice a cause until after the effect. This may be a common occurrence is a complex environment.

BabyA being *near* a ditch with the subsequent occurrence of a shock) and novel anticipatory action. [Shing 1993c]

2.4 Systemic Design and 'emotional' phenomena

The application of the design-based approach to studying cognitive processes has revealed some of its limitations [Read 1993a]. Systems built using the design-based approach may not generate useful theoretical knowledge as the requirements are not under the same constraints that apply to naturally evolved, biological systems. For example, an evolved, intelligent architecture is likely to consist of compromises between phylogenetically old and young mechanisms. A model built from abstract requirements can be built from scratch without the need for such compromises. Hence, any designs arrived at may not map on to existing animal or human architectures. Also, understanding of the relation between mechanism and behaviour is unclear[12]; therefore, it is difficult to build accurate models of the underlying mechanisms. And finally, due to interdependencies between architectural components modelling individual components as separate wholes can be problematic. Systemic design is a constrained subset of the design-based approach that attempts to overcome these limitations by including the following considerations.

- Evolution
  - Considering the genealogy of the architecture of a species in terms of problems encountered and solutions found during its evolutionary development will
    - reveal architectural redundancy i.e. an aspect of the architecture may have served a purpose in the distant past but is now functionally useless.
    - further constrain the design. E.g. a fundamental evolutionary 'decision' taken many millions of years ago may have circumscribed subsequent phylogenetic development.
    - reveal inter-species architectural congruences E.g. humans may share some aspects of their cognitive architecture with apes, rats and other mammals as i. similar problems will have produced similar solutions and ii. evolutionary change will tend to retain successful architectural components.
  - The design-based approach, by ignoring the above phylogenetic considerations, may have the tendency to produce types of architectures unrelated

[12] Computer science and portions of AI have steadily enriched our understanding of the relation between mechanisms and behaviour - this underlies the design of modern programming languages, and any architectural design.

to the natural world in important ways.[13] This aspect of Systemic design, by forcing the consideration of real organisms over their evolutionary history, attempts to avoid this.

- Resource limitations
  - Considering resource limitations is important for limiting the space of possible designs that satisfy initial requirements. The resource limitations the designer places within a model are often dependent on the goals being pursued i.e. there may be a tendency to 'fix' or 'hard-wire' the model to produce desired results. The Systemic Design approach takes the stance that, wherever possible, all resource limitations should always be taken into account. But this is not always feasible and some selection will be required. An embryonic typology of resource limitations has been elucidated, including environmental, physiological, attentional, learning, and control and representation issues.
- Holism
  - To overcome the problem of component interdependency within a complex architecture, Systemic Design encourages a holistic strategy that is a mixture of bottom up, top down and 'middle out' design. Top down as high level requirements are formulated and then decomposed into a specification of functional mechanisms within the system; bottom up as empirical data can be used to test proposed models and constrain design decisions. The term 'holistic' is meant to convey the idea that a system should be studied from as many points of view as possible, incorporating data from different research fields, with the eventual hope that high level concepts will map onto low level data via intermediate functional levels.

Systemic design has been chosen as a methodology to study 'emotion' in autonomous agents by refining J.A. Gray's neuropsychological model of emotion systems[14] [Read 1993b]. Gray proposes three fundamental systems mediating emotion: a behavioural approach system, a flight—fight system, and a behavioural inhibition system. The functionality of these systems is supported by considerable neuropsychological evidence; however, specification of inputs and outputs is vague and the role of information representation is ignored. In order to remedy this the model will be developed at the computational level and then implemented, incorporating other neuroscience data where appropriate. Research will concentrate on modelling

[13] Of course, the design-based approach would be adequate if the project were solely concerned with designing autonomous agents without wishing to relate such work to the natural world.

[14] Jeffrey A. Gray: Brain Systems that Mediate both Emotion and Cognition, Cognition and Emotion, 1990, 4(3), 269-288, Lawrence Erlbaum Associates Ltd.

the information processing that occurs in the septo-hippocampal system (part of the limbic system) of a rat. The simulated rat, incorporating Gray's three emotion systems, will be placed within various experimental paradigms, such as a shuttle-box with administered shocks and rewards. Results from the simulation can then be compared with empirical data to see if the model replicates rat behaviour with the aim of exposing any shortcomings of Gray's model. Another objective of this research is to provide a coarse-grained functional mapping of brain systems at three levels: the neural, computational and behavioural. And, as a corollary to this, it is hoped that the terminology arising from work in other parts of the project[15] will be linked to the functions of specific brain systems. Finally, the temporal implications of brain system functionality will be explored, with the aim of resolving the 'role of cognition in emotion' debate (Lazarus v. Zajonc) [Read 1993c]. In this way, it is hoped that the efficacy of Systemic Design will be demonstrated and underline the importance of cross-fertilization between differing research fields.

2.5 AIMAE - action, reaction and learning

AIMAE (Agent Information Management and Execution) [Paterson 1993a] is an architecture in development that is intended to support both reactivity to environmental events and goal deliberation in a balanced and successful way. Previous AI planning systems tend to fall into one of two categories: those that are predominately 'reactive' i.e. agent behaviour is event driven, or those that are predominately 'classical' i.e. agent behaviour is goal driven.[16] Or, colloquially, those that 'jump before they think' and those that 'think before they jump'. But for an autonomous agent in a complex, changing environment a balance must be struck between these two extremes [Paterson 1993d].

A novel approach has been outlined to strike this balance. In AIMAE a plan is represented as strategy, which is a goal along with the subgoals directly linked to that goal. In addition, cross-indexing tables are used to link strategies with goals enabling efficient plan construction. At run-time the agent will attempt to satisfy a top level goal by choosing from a number of strategies. The choice of strategy is determined by its *rating*, which is a profile of the strategy's general efficacy over

[15] Eg. concepts such as *insistence* or *goal filtering*.

[16] The distinction between event driven and goal driven behaviour is not straightforward. Even highly deliberated plans formed from goals generated by internal events with relative autonomy from environmental occurrences can still be viewed, in the final analysis, as event driven. For example, while washing the dishes Jon decides he will dedicate his life to helping the needy. This generated goal has no obvious correlation to contemporaneous events. Rather, it is the outcome of a complex interaction between Jon's personal history and some of his fundamental dispositions. So in this sense, even Jon's decision to help the needy is event driven. What makes the distinction between event and goal driven behaviour somewhat artificial is that agent behaviour is an interaction of the two.

a number of dimensions.[17] In this way, plan formation is determined not only by the top level goal but also by current events. For example, goal A may be satisfied by strategy 1 or 2. However, the urgency of goal A is high and strategy 2 has a superior termination time rating. Hence strategy 2 is chosen. In other words, plan construction in AIMAE is both 'classical' in the sense that it is goal generated, and 'reactive' as the event stream has a direct impact on plan formation through run-time selection of strategies. In addition, the system will attempt to interleave planning with execution [Paterson 1993b].

The ratings for a particular strategy are not fixed. Each time a strategy is used its ratings are re-computed. A strategy that nearly always satisfies a goal in an efficient manner will tend to have a 'good' rating, which will be reinforced each time it is successfully applied. Consequently, it will be chosen more frequently. This is a kind of learning, where a strategy is a hypothesis that is continually tested in the environment, the results of which are incorporated into its rating. It is hoped that this approach - that of continuous learning - will partially overcome the problem of when and how learning should take place.

Strategies can traverse the movement from hypothesis to belief via repeated, successful application. For example, strategy 1 and strategy 2 are both candidates for satisfying goal A. Deciding which one to use is indeterminate as their ratings are similar. After many applications of strategy 1 and 2 their ratings diverge until strategy 2 becomes the best strategy for satisfying goal A in most situations. AIMAE will approach the stage where she will *always* choose strategy 2 when attempting to satisfy goal A. It is no longer one hypothesis amongst many - but a firm belief.[18] AIMAE, at this stage, is concerned only with one aspect of learning - that of improved strategy selection; other aspects of learning, such as strategy analysis and strategy creation, are yet to be tackled.

When implemented, it is expected that AIMAE will exhibit the following behaviours [Paterson 1993c].

- Automatic behaviour
  - AIMAE has a management module, with functionality similar to that of management in NML1, and an interpreter. The interpreter requests help from the manager when problems in goal expansion arise. However, it is possible for the interpreter to automatically expand a whole strategy frame without recourse to management functions. I.e. in this case no high level processing is necessary for goal satisfaction. This is automatic

[17] Dimensions include i. degree of *satisfaction* ii. *efficiency* of strategy execution iii. *time* needed for strategy execution and iv. the *frequency* of strategy use.

[18] It is recognised that the use of the term 'belief' in this context is not without its dangers. The justification for anthropomorphizing is that the use is illustrative rather than technical.

behaviour - no ‘attention’ is required.[19]

- Action slips
  - The manager is needed to check that the invocational pre-conditions of a strategy are met. Therefore, it is possible that automatic behaviour can lead to inappropriate strategies being selected that do not lead to goal satisfaction. This is an action slip.
- Learning
  - Over time AIMAE should develop winning strategies tailored to achieving specific goals. Its efficacy within the environment should increase.

AIMAE is still under development. The proposed architecture will be tested within the nursery domain. It is hoped that the use of strategies and ratings to integrate planning, execution and learning will enable AIMAE to operate successfully in a complex, changing environment.

3 Future Work

Here we examine where the project is heading.

3.1 New research

New research will initially concentrate on the relationship between urgency, importance and insistence within the nursery domain, striving to obtain a clearer understanding of the content of these concepts. Consideration of these problems is viewed as a good ‘entry-point’ into the more general problem of motive processing in a resource limited agent. For example, it is expected that investigation of filtering mechanisms for goal surfacing will lead on to problems of high level goal management.[20]

An aim of this research is to relate the heuristic computation of insistence to the urgency and importance of a goal, with a view to demonstrating the fallibility of goal surfacing. It may then be possible to provide architecturally grounded explanations for ‘errors’ in human motive processing [Wright 1993].

[19] Posture control is an example of goal satisfaction that requires no high level attention.

[20] High level management processes can control the goal filter. Therefore, any investigation of goal surfacing will need to consider the effect of these processes.

3.2 Possible directions

Work in the Attention and Affect project can be divided into four areas.

- **Motive processing.** Concerned with describing a framework for an architecture that can cope in a complex, dynamic environment. Involved with specifying management and meta-managment processes.
- **Learning and attention.** Concerned with computational mechanisms that implement a subset of animal learning.
- **Low level brain systems mediating 'emotion'.** Concerned with refining a neuropsychological theory and plugging the gap between the neuronal and behavioural levels of description.
- **Learning with planning.** Concerned with specifying how learning can interact with run-time plan generation.

Future work could extend all four areas. Fruitful areas might be

- Investigating how management and meta-managment processes work.
- Further investigation of types of learning. Comparing and perhaps integrating reinforcement learning with strategy selection on ratings. How do they differ? Could a combination of the two work? Is one reducible to the other? etc.
- Implementations of other neural level theories. Further comparisons of models with empirical data. More sophisticated models. Terminological mapping through neural—computational—behavioural levels.
- Adding a planning system to Gray's model.
- Extending the typology of design decisions.
- How is urgency and importance computed? How does this relate to real world situations? How do people solve the problems of the nursemaid (build a game nursery for people to play).
- Clarifying the distinction between management and meta-management processes.
- Integrating the learning of cognitive reflexes that bypass management and planning functions with earlier designs.
- etc.

The top level goal of this kind of work is to produce a coarse grained theory of the architectures implemented in the brains of intelligent agents.

4 Conclusion

Here the project is judged with respect to the original aims. It must be borne in mind that work is still in progress.

- The research has not concentrated solely on the functional requirements needed for the processing of motivators and control of attention. It has branched out into areas such as learning and lower level mechanisms mediating emotion.[21]
- Work has progressed in exploring the design-space of possible architectures [Beaudoin 1993e, Beaudoin 1993f]. Therefore, this aim has been partially achieved but the problem is such that it has no definite end point.
- The aim of implementing a sequence of designs capable of fulfilling increasing subsets of the requirements has not even begun. NML1 remains an incomplete prototype. Other nursemaids are at the design stage.
- An embryonic, architecturally grounded, conceptual framework for describing affective states involving control of attention has been outlined. For example, in [Beaudoin and Sloman 1993][22] the process of motivator surfacing is explicated in terms of the underlying architecture. However, an architecturally grounded understanding of 'affective' states such as desires, moods and attitudes is still a long way off.
- The educational tool has not been built. Neither has an effective interface been implemented demonstrating the key features of the models.

There is a long way to go!

[21] This may be deemed a good thing but we are concerned here with relating the project to its original aims.
[22] Section 9: A Scenario in the Nursemaid Domain Revisited.

References

[Beaudoin and Sloman 1991] A Proposal for a Study of Motive Processing. Unpublished manuscript. School of Computer Science, University of Birmingham.

[Beaudoin and Sloman 1993] A Study of Motive Processing and Attention. In Proceedings of AISB93, A. Sloman, D. Hogg, G. Humphreys, A. Ramsay and D. Partridge (Eds), pp 229–238, Oxford: IOS Press.

[Beaudoin 1993a] Thesis - work in progress. Chapter 4 - Process Specification. School of Computer Science, University of Birmingham.

[Beaudoin 1993b] Personal communication.

[Beaudoin 1993c] Thesis - work in progress. Chapter 5 - Architecture. School of Computer Science, University of Birmingham.

[Beaudoin 1993d] Thesis - work in progress. Chapter 3 - The Concept of a Goal. School of Computer Science, University of Birmingham.

[Beaudoin 1993e] Thesis - work in progress. Chapter 6 - Analysis of the Design Space. School of Computer Science, University of Birmingham.

[Beaudoin 1993f] Analysis of the design space surrounding a model of motivator processing. Notes for a talk at the Workshop on Architectures Underlying Motivation and Emotion - WAUME 93, University of Birmingham.

[Paterson 1993a] Planning and Autonomous, Affective Behaviour. Thesis proposal. School of Computer Science, University of Birmingham.

[Paterson 1993b] Planning for Behaviour. Internal document. School of Computer Science, University of Birmingham.

[Paterson 1993c] Planning, Learning and Autonomous Behaviour. Notes for Emotions Meeting, School of Computer Science, University of Birmingham.

[Paterson 1993d] Planning and Autonomous Behaviour. Notes for Workshop on Architectures Underlying Motivation and Emotion - WAUME 93, University of Birmingham.

[Read 1993a] Systemic Design: A Methodology for Investigating Emotional Phenomena. (In the process of being updated). Unpublished document. Cognitive Science Research Centre, School of Computer Science, University of Birmingham.

[Read 1993b] Phd Year Two Progress Report, Using Systemic Design to Study Emotion. School of Computer Science, University of Birmingham.

[Read 1993c] Seminar notes. School of Computer Science, University of Birmingham.

[Read and Sloman 1993] The Terminological Pitfalls of Studying Emotion. Paper presented at the Workshop on Architectures Underlying Motivation and Emotion - WAUME 93, University of Birmingham.

[Shing 1993a] Computational Mechanisms for Constraining Learning. Paper presented at the Workshop on Architectures Underlying Motivation and Emotion - WAUME 93, University of Birmingham.

[Shing 1993b] Research Progress Report 5: Computational mechanisms for constraining learning. Cognitive Science Research Centre, School of Computer Science, University of Birmingham.

[Shing 1993c] Notes for seminar. School of Computer Science, University of Birmingham.

[Sloman 1987] Motives, Mechanisms and Emotions. Emotion and Cognition, 1, 217-234. Reprinted in M.A. Boden (ed) The Philosophy of Artificial Intelligence "Oxford Readings in Philosophy Series" Oxford University Press, 1990.

[Sloman 1992] Towards an Information Processing Theory of Emotions. Notes for Cognitive Science seminar, October 1992.

[Sloman and Humphreys 1992] The Attention and Affect Project. Appendix to JCI proposal.

[Sloman, Shing, Read, Beaudoin 1992] Notes from an emotions meeting. School of Computer Science, University of Birmingham.

[Sloman 1993] Why We Need To Study Architectures. Notes for workshop on architectures underlying motivation and emotion - WAUME93, Birmingham.

[Wright 1993] Research Progress Report 1: Urgency, Importance and Insistence - goal surfacing in a resource limited agent. School of Computer Science, University of Birmingham.