Cognitive Architectures for Affect and Motivation

Dr.Darryl. N. Davis

Department of Computer Science,

University of Hull,

HULL, HU6 7RX,

United Kingdom

E-mail: D.N.Davis@hull.ac.uk

fax: +44(0)1482466666

tel: +44(0)1482466469

Cognitive Architectures for Affect and Motivation

Abstract

General frameworks of mind map across tasks and domains. By what means can a general architecture know when it has adapted to a specific task, a particular environment or a specific state of a previously known environment? Our current work on this theme presents an affect and affordance based core for mind. This draws on evidence from neuroscience, philosophy and psychology. However we differentiate between the mechanisms and processes thought to be allied to cognition and intelligent behaviour in biological architectures and the foundational requirements necessary for similarly intelligent behaviour or cognitive-like processes to exist in synthetic architectures. Work on emotion is a morass of definitions and competing theories. We suggest that we should not further this confused framework with unnecessary (and often unneeded) models of emotion for artificial systems. Rather we should look to foundational requirements for intelligent systems, and ask do we require emotions in machines or an alternative equivalent, for example affect, of use in control and self-regulation? This paper addresses this issue with experimentation in a number of simulated and robotic test-beds

Keywords: Cognitive Architectures, Affect, Emotion, Motivation, Robots

1. Introduction

There is an accepted and growing consensus among theorists and designers of complete intelligent systems [1-5] that synthetic minds, to be complete and believable, require a computational equivalent to emotion to complement their behavioural and cognitive capabilities. This requirement for affective states in complete cognitive theories and architectures was highlighted in earlier prominent research [6-7]. This paper confronts the requirement for emotion in intelligent computational systems on a number of grounds. The thesis is that overall the theory of emotion is too disorganised to be of much use in the design of synthetic intelligence and that more pointedly, emotion (*per se*) is not really a requirement for (the majority of) synthetic intelligence theories or systems. It is suggested that a direction given by the less semantically overloaded term affect is more appropriate for synthetic intelligence. The argumentation for affect as a control mechanism makes use of the control state approach to mind, ongoing designs and experimental work with computational agents and cognitive robots.

The paper addresses emotion and affect, before considering motivational states. It presents an architecture design (and theory) that embodies the analytical conclusions of this argumentation. This research into the development of architectures for synthetic intelligence is built around the concept of control states, and in particular motivators. The primary conjecture is that the design and implementation of such architectures can proceed using a systematic control language that obviates the need for ad hoc heuristics to direct the processing within an intelligent system. This control language can be grounded in the theory of affect (and subsume emotion). Experiments using implementations of this architecture (both in synthetic simulations and robotic test-beds) are presented. All these different perspectives come together under the umbrella project of Computational Architectures for Motivation, Affect and Learning (CAMAL).

2. The Red Herring that is Emotion

Norman's pivotal paper [6] suggested emotion-like processes will be necessary for artificially intelligent systems. This section builds an argument that denies the need for emotion in many synthetic systems, while accepting that notable systems have been built based on models of emotion [8, 9]. It would seem plausible for some artificial systems to need an emotion model, for example, a natural language system that addressed conceptual reasoning [10, 11], or a sophisticated human-computer interface system where emotional recognition and correspondence was needed [12], or emotion-based control systems for characters in computer games [13]. However, even in these situations, a highly negative affective affordance towards an event or situation is *computationally* as meaningful as "fear". It is suggested that the question researchers in this area need to address is: what is the requirement for autonomous agents inhabiting synthetic test-beds, computer games or controlling robots to include emotion rather than affect, except for possibly providing working proof of computational models based on (neuro-) psychological theory? The remainder of this section provides a context for emotion in cognitive architectures, referencing pivotal work in psychology and philosophy; section 3 will then address this question.

Theories of emotion can be typified as belonging in one of several types, for example physiological [14-15], evolutionary [16], expression [17], appraisal [18] or goal based [19]. This is partially due to different programmatic objectives within, for example, neurophysiology, psychology and philosophy. Some [8. 20] consider emotions to be a cognition centered set of phenomena, while others [21] consider these terms to be centered on low-level (neuro-physiological) control processes that affect cognition. These arguments are not addressed in depth here – research on emotion is discussed elsewhere [22, 23]. Duffy [24] considers the use of the fuzzy, ambiguous and misleading term "emotion" as fundamentally flawed; and that such terms should be abandoned as confusing, and new or clearly delineated terms used only where such concepts are clearly and unmistakably identified. However, there is such a volume of research in this area (see De Houwer and Hermans [23] for a recent overview) that a significant academic revolution would be required to pursue such a path with any success. While this may be true of disciplines that study human intelligence, the same does not hold for artificial systems.

For psychology, and the philosophy of the mind, emotion is an important concept that is central to any complete theory of mind. Given that, as highlighted immediately above, there is no one accepted theory or definition of emotion, this paper will provide "*just sufficient"* definitions to enable clear differentiation between emotion and affect. From the perspective of philosophy, emotion is a kind of mental phenomenon whose role is to provide an agent an orientation, or an attitude, to the world. To quote Wollheim ([25]:15): *If belief maps the world, and desire targets it, emotion tints or colours it: it enlivens it or darkens it, as the case may be*. These emotion guided perspectives arise as (or from) mental states, and are induced (and induce in turn) mental dispositions. Such mental phenomena can be described using a vocabulary which includes concepts such as *fear*, *hate* etc.

The use of the two concepts *mental state* and *mental disposition* mirror what many theorists believe to be the distinction between emotion and mood - the duration of the experience. Emotions can occur with a rapid onset, with little awareness and with involuntary changes in expression and physiology. Typically, an emotion has a duration of minutes or seconds unless the emotion is invoked again [17]. Mood, however, refers to a longer-term affective state. A mood may arise when an emotion is repeatedly elicited and can last for hours or even days [25, 26]. Nearly all theorists consider emotions to be conscious states, and moods dispositions that can be consciously accessed. As such they can be identified, conceptualised and described using experiential language.

From the very different perspective of neuropsychology, Rolls [27], while accepting that emotions are defined differently by different schools of researchers, ultimately relates his definition of emotion [27:pp40] to rewards and punishments. Most of the psychological theories indexed above fall between the definitions given by Wollheim and Rolls.

Affect can be considered a broader concept and experienced as positive, negative or neutral valences. It is not necessarily accessible to conscious thought, and is not necessarily describable using experiential linguistic concepts and labels such as hate, fear, surprise etc. Affect is a term first introduced by Freud [28] to describe the quantitative aspect of ideas and the responsibility for driving them into consciousness. Clore [29] describes affect as a general term meaning the perceived goodness or badness of something. Frijda [20] agrees with this view that affect is pleasant and unpleasant feelings. In contrast to moods, which refer to more pervasive and sustained states, affect refers to more general and fluctuating changes. The phenomena described here under different definitions have obvious relationships between them. Therefore, we cannot expect that each of these concepts refer to a distinct class of events clearly separated from the others. The following sections amplify on these relatively simplistic definitions.

Given the above proviso, that this research would prefer the use of affect over emotion, two emotion theory types do provide a sound basis for affect in synthetic agents. Emotion in the goal-based theories, for example [19], can be described as "*a state usually caused by an event of importance to the subject*". It involves mental states directed towards an external entity, physiological change, facial gestures and some form of expectation. Scherer defines emotion as "*a sequence of interrelated, synchronised changes in the states of all organismic subsystems (… monitoring subjective feeling) in response to the evaluation of an external or internal stimulus event that is relevant to central concerns of the organism*" [9]. Such emotional processes involve five functionally defined systems involving: information processing over perception; regulation of internal states; decision making over competing motives; the control of external behaviour; and a feedback system across these four. There is considerable overlap between these definitions. In effect, emotions, in socio-biological agents, are affective mental, conative, and/or cognitive states and processes. A coherent hybridization of these theories is possible with the notable exception of the concept of basic emotions. A number of researchers use the concept of the basic emotions. Scherer instead allows for modal forms of emotive processing. Of the many modes that an emotion system can take, some are near identical or relatively similar to the states described as basic emotions. Basic emotions will therefore not be discussed further in this paper. The Scherer concept of modal emotions, with multiple not-accessible-via-consciousness processes, is relevant to our concept of affect as the following two sections detail.

3. Affect not Emotion

Numerous prominent researchers into intelligent systems have suggested that affect-like mechanisms are necessary for intelligence [7, 6, 1, 30, 31]. Sloman [2] has suggested that while emotion is associated with intelligent behaviour, it may not be a prerequisite. If that is the case and that emotion is a side-effect of mechanisms in sophisticated and complex biological architectures, intelligence is now tightly bound to the control of these side-effects through evolution. The development of control mechanisms to harness and cope with the affective associations of the mechanisms necessary for intelligence, over the diachronic intervals associated with evolution, is such that in effect emotion and affect are now central to intelligence in biological systems. However for synthetic (or artificial) architectures, it is possible to generate intelligent systems capable of supporting and utilising affect but not requiring emotion.

Parsimony suggests that at a theoretical, design and implementation level, we need an affective component for intelligence but the need to add emotion *per se* is debatable. Although this seemingly runs contrary to the conclusions of Phelps [32], and the current research of many in cognitive architectures such as Ziemke and Lowe [33] or Gros [34], any artificial system that does not require the use of terms explicitly linked to emotion (for example “hate”, “fear” etc. in a natural language system, e.g. [10]), can arguably be defined using affect and motivation.

The theory of synthetic intelligent systems can therefore progress without the need for emotion but with a requirement for affective control states that can draw on theories of emotion and cognition in biological intelligent systems. This would mean for example that a synthetic system need not model or recognise the emotive state termed “fear” but recognise highly valenced negative internal states and environmental affordances that (potentially) jeopardise its role and tasks in its current environment. Put simply, theories of emotion from the cognate disciplines such as philosophy and psychology can afford functional models of affect for synthetic systems without the need for the theorist or designer of synthetic systems to be concerned with the semantic overloading associated with specific emotions, or indeed the concept of emotion per se.

Furthermore many theories of emotion involve body attitude or facial expression changes that are typically inappropriate for machines. There are few machines (and no generic cognitive systems) that rely on body posture or facial expression for communication other those affective systems that attempt model the emotive state of their user [26, 35]. An immediate benefit is the researcher interested in intelligent synthetic systems can move away from the definitional morass that surrounds much of the work on emotion. The exception of course is where models of emotion are required for sophisticated man-machine interactions. Even there the interactive system only needs to model the emotive or affective state of its user, and not function in terms of emotion.

Previous related research [22, 36] has indeed used emotional models that include basic emotions. The current stance is that emotion, and therefore basic emotions, are unnecessary for many classes of synthetic systems, for example intelligent agents, robots and computer games. We are developing a theory of affect that draws on themes such as control states, affect and motivators [7, 36] and affordances [37, 38]. We define affect in terms of processes and representational structures across a cognitive architecture. Affect is qualitatively defined as negative, neutral or positive and can be mapped numerically over the interval [-1.0, 1.0]. Affect becomes the basis of a control language for agent architectures. It allows external events and objects to take valenced affordances, and allows internal mechanisms to be prioritised via valenced processes.

A salient feature of many definitions of emotion is that they are described in terms of goals and roles. Earlier work on agents and control states [36] focused on goal processing. It addressed how goals, and related control states, need to be valenced in a number of different ways, for example intensity, urgency, insistence. At the deliberative level, affective values can be associated with processes and predicates and then relayed as control signals to instantiate and modify aspects of motivators and their associated representations and behaviours. Furthermore, if an agent is to recognize and manage emergent behaviours, and particularly extreme and disruptive control states, this multi-layer model of affect provides the means for reflective (meta-cognitive) processes to do this. This model of affect addresses the need to integrate reflective (meta-cognitive), deliberative, reactive and reflexive level agencies in a synergistic fashion

The valencing of processes and representational structures can be given or the agent can adapt or learn appropriate affordances according to its role and current environment. Affect can provide the basis for perceptual valences that support the external environment affordances appropriate to the agent. As an agent monitors its interactions within itself and relates these to tasks in its external environment, the impetus for change within itself (i.e. a need to learn) is manifested as an affect based state. Such a control state can lead to the generation of internal processes requiring the agent to modify its behaviour, representations or processes in some way. The modification can be described in terms of a valenced mapping between its internal and external environments. This influences the different categories of cognitive and animated behaviour.

4. Mechanisms for Affect and Motivation

This section makes use of computational work on the nature of motivation and affect and looks to how these can be co-joined in perception and cognition. While this theory is in development, and arguably currently incomplete, there follows argumentation, design and simple computational experiments that demonstrate the more cogent aspects.

The control state approach to cognition [7, 31] builds on an assumption that cognition is an epistemic process that can be modeled using information processing architectures. Such information processing architectures can be in any number of non-exclusive control states; for example, learning optimal behaviours for specific task environment configurations while attending to multiple goals. The processing of information can give rise to changes in the current set of extant control states. For instance, perceptual updates leading to belief revision, in a Belief-Desire-Intention (BDI) model, may give rise to goal revision and the activation of meta-level reasoning on what specific BDI model to adopt for the current situation. The nature of the information processing is dependent upon the currently extant control states. The same information may be processed differently in different control states [31, 36, 39].

One of many current taxonomies provides for five broad categories of control state. These broad categories overlap, and indeed the main focus of this research (motivators) are often used as a generic framework that draws all these together.

- Beliefs are internal models, whether assumed or possibly inferred from perceptual acts or from information arising from other control states, such as planning or deductive reasoning; these need not have a rational basis. Beliefs can be prioritised using preference models or valenced by affective states.
- Images are control states using mental images, and alike to Barsalou's [40] perceptual simulations. These images may relate to any perceptual modality in typical or atypical ways.
- Imaginings are control states that embody alternative ways of constructing internal worlds, for example analytical, deductive or analogical reasoning [41]. These can be related to directed problem-solving (e.g. planning) and therefore motivators.
- Behaviours, at the most constrained definition, are actions on an environment. The most primitive type, reflexes, are ballistic mappings from input (i.e. perception) to output (i.e. behavioural response). More sophisticated behaviours on an environment may be large

subsystems built to achieve specific tasks (e.g. reactive sub-architectures). A less constraining definition of behaviour would include mental activity, and blur the definitional border between mental behaviour and imaginings.

- Motivators are problem-solving schemas which link internal and external 'reality', involving perception of events and states, representations and paths to modified states of affairs.

Here the focus is on motivators as a representational form that enables perception, affect and cognition and behaviour to interact. A motivator is a representational schema (or blackboard) that brings together many aspects of perceptual and cognitive processing. The representational framework is outlined in this section with due argumentation. Experimentation with design and computational models is introduced throughout the rest of the paper. Note that this treatment differs from the earlier references given above [22, 36, 38].

Table 1 lists the main components of the motivational construct. The construct provides the means by which an agent can organise data, information and knowledge related to objects, agents, tasks, goals and actions of interest. Experimental work of Bourgne [42] revisited the representational structure and processes associated with motivators [36], but made use of affect and affordances instead of emotion. Theoretical, design and architecture parsimony, has lead to the rejection of emotion as a requirement for intelligence in the agent control systems this research addresses, and as argued above - affect suffices.

To date, main motivator components implemented in CAMAL (see Figure 1) include

- Semantic Content: typically a proposition P denoting a possible state of affairs, which may be, categorically or partially, true or false.
- Motivator Attitude to the semantic content P, for example make true, keep true, make false etc.
- Belief Status: An indication of current belief about the status of semantic content P, represented for example using Boolean or multi-valued logic, or probabilistic affect values.
- Actors and Agents referenced by the motivator
- Objects referenced by this motivator
- Behaviours associated with any intention, plan set and/or past similar motivators
- The current (commitment) status of the motivator, e.g. adopted, rejected, undecided, interrupted, completed.

- Goal Importance. This real value (in the range [0, 1]) is managed by the BDI affective processes in motivator evaluation and goal revision.
- Association Insistence: This real value (in the range [0, 1]) is managed by the BDI affective processes, and is strengthened by goal completion using the associated behaviour.
- Motivator Intensity: In effect the real value (in the range [0, 1]) the affective system and processes associates with the difference between the current and desired state of the motivator

CAMAL uses a cognitive BDI schema to drive a motivational blackboard. Figure 1 shows how belief affordance (described below) enables the architecture to select a focused belief set that mirror its current activities (as highlighted by actions, objects and agents referenced in a current motivator). The motivator enables goal revision and the selection of the next goal based on current goal success and belief importance. The deliberative management of these constructs allows the selection of reactive sub-architectures related to specific objects and actors in the external environment. This in turn drives motivator revision using the *association* construct which enables Belief-Desire-Intention combinations to be ranked based on the likelihood of their success (termed *insistence*). Goal importance, belief status, association insistence and motivator intensity, and motivator success (or failure) are all underpinned by affect values. Together they allow motivators to persist or be updated to reflect the multi-modal focus of attention as the domain model of the architecture is adapted and valenced by the affect model.

A more extensive implementation allows for the affective focus on agent-internal behaviours through, for example, the use of meta-control and meta-cognitive control processes. Associated with motivational structures are attitudes to classes of events and entities relevant to that motivator. These are valenced as internal affordances using affect. The association of perception, behaviour and abstract representations about plans of actions, and the relevance of actions and entities in both the environment and internal worlds, are defined over the same numeric range, [-1, 1]. Affect and affordance become the means by which the agent architecture can weigh its beliefs, processes and control the economics of its processing. It provides a means whereby attention can be directed [43].

These qualities can be defined over processes within a multiple level information processing architecture. Unlike many current architectural theories this need not be defined in terms of three columns and three levels [44, 2, 36]. It is suggested that such architectural stances are limiting and somewhat illusory in that they lead research into a too tightly constrained framework; for further argumentation on this, see [45]. Vijayakumar [46, 47] for example uses a four column, six layer approach in extending CAMAL with reinforcement learning and metacognition.

5. Computational Architectures for Motivation and Affect

This section places recent investigations into motivation, affect and affordance within ongoing research into the development of architectures for synthetic intelligence. In these developing computational systems, activity and behaviour at one level is represented and controlled in other layers. The primary conjecture is that the design and implementation of such architectures can proceed using a systematic control language that obviates the need for ad hoc heuristics to direct the processing within an intelligent system. This control language is grounded in the use of affect with the aim to be consistent across different domains, tasks and levels of processing. The computational work is being developed with no requirement for emotion but rather a reliance on affect (a valencing of and within internal processes) and affordance that together can be used to guide both internal and external activities.

The current work is based on experiments in the theory, design and implementation of affect and emotion based architectures [48, 36, 22, 38]. It builds on the ecological perspectives offered by Gibson [37], and on the work of Simon's control state theory [7]. Preliminary work [48, 36], highly influenced by the Cognition and Affect project [31, 39], centered on motivators and goals, how they come into being and how they are managed. Since that work various closely aligned projects have focused on using affect to drive goal selection and behaviours [49, 42], how to relate Belief-Desire-Intention models to affect [50], the role of meta-cognition in directing the focus of cognitive architectures [46, 47] and the use of the architecture in controlling mobile robots [51]. All these different perspectives come together under the umbrella project of Computational Architectures for Motivation, Affect and Learning (CAMAL).

We use variations on and extensions to the three-column, three layer architecture but are not unequivocally committed to such architectures. For example, some experimentation [50, 38] makes use of an architecture based on cognitive models of reasoning in children (A-CRIBB). The coupling of beliefs, desires and intentions shown in Figure 1 can be seen in columns three and four of Figure 2. The BDI model described below is represented as the arcs from the perception processes across to the motivational blackboard associated with the BDI hierarchies in column 4.

In the four-layer, five-column model of Figure 2, there exist reflexes and reactive behaviours that allow a direct response to sensory events. However, behaviours need not necessarily be environmental activities, but motivational states (intentions), planners or learners that adapt the architecture. Similarly not all perceptual acts relay environmental descriptions. The reflective monitor acts as a “mind’s eye”, activating the meta-level reasoning mechanism in response to the architecture repeatedly failing to achieve tasks or in other out-of-control situations.

There are various different definitions of what constitutes a reactive component. For example a system may have no changeable internal state so that the current input determines the current output. In our broad research, we term these reflexes (the base layer of the architecture and domain dependent); and in this case the output is always the same given the same input. The definition of a reactive system taken here is that the systems output is determined not only by its input, but also by its internal state. For some CAMAL domains reactive behaviours are built around reflexes, and yet further reactive sub-architectures built around those composites as in the robot experiments. The reactive subsystems can provoke processes or be modified at a more abstract level. Other automatic processes *necessitate* the generation of deliberative control states to achieve their goals. The deliberative layer represents those (control state) processes typically studied in thinking, human problem solving etc., plus other processes related to the management of low level actions. Within this paper, deliberative will refer to any system whose output is not only determined by its input and current state, but also by its previous states and/or the current/previous states of other systems. In other words a deliberative system is one whose output is based upon an extended memory beyond that of its own current state. A deliberative system can also be driven by the concept of future states (e.g. goals or desires); and indeed a BDI model is used here. The reflective (or meta-cognitive) processes serve to monitor cognitive behaviour or control it in some other way. Extreme affective states (symptomatically categorised as a loss of control or perturbance) are effectively clamped by means of self-regulatory (affective) processes within the architecture. This model is quite general, and the effect of altering the relative size and importance of the layers is an open issue.

The architecture, BDI and motivational models are configured at run-time using domain models that allow the architecture to instantiate itself in a previously known best or specifically designed configuration. In some operational modalities the architecture can modify, conflate or expand according to its own requirements once running. High level and low level processes coexist and interact in a holistic manner through the use of affect. In effect, goal processing, planning, decision making, behaviour selection, attitude shift and other cognitive processes are not purely abstract but exist in relation to other autonomous (and affective) processes. They are, in effect, embodied within the context of their interactions with their underlying processes and the agent's relationship(s) with its environment.

6. BDI Reasoning: CRIBB and a-CRIBB

CRIBB (Children's Reasoning about Intentions, Beliefs and Behaviour) is a computer model based upon a general theory for belief-desire reasoning in children [52]. It simulates the knowledge and inference processes of a competent child solving false-belief tasks [53]. A simulation run in CRIBB starts by giving propositions containing facts and perceptions about some scenario in sequential steps according to the time interval in which the propositions arise. On the basis of the given propositions and the inferences drawn, CRIBB answers test questions about the cover story. The questions can be about its own beliefs or about the intentions, beliefs and behaviour of another person in the scenario.

CRIBB represents propositions about physical states of a given situation and the intentions, beliefs, perceptions and behaviour of others. Its knowledge base consists of four types of practical syllogisms and three other inference schemata, which represent the relations between these propositions. Practical Syllogisms denote knowledge about the relations between intentions, behaviour and beliefs of another person. The three other classes of inference schemata relate perception-belief, belief-time and fact-time. These are split into primary and secondary representations. Primary representations are the system's own beliefs about the situation and the behaviour of other people. Fact-time inferences, propositions about facts along a time scale, are classed as primary representations. Belief-time and perception-belief inference schemata are both types of secondary representation as they contain beliefs about the system's own and others' beliefs. A further element of CRIBB is a consistency mechanism that detects and resolves contradictions in belief sets. This is invoked each time a new proposition is added, in order to ensure the consistency of its knowledge base. Related work [50] suggests that adding affect to cognitive desire and intention models of CRIBB result in more effective processing and task management.

The major difference between CRIBB and a-CRIBB is the latter uses an affect value from its goals to order perceptual propositions. This affects belief order, in particular truth maintenance, and the simple goals the a-CRIBB agents act out. Rather than attempt to completely and accurately model the agent's world, affect can be used to guide attention so an agent is drawn to aspects of the environment deemed to be of importance. Assigning an affective affordance enables a process by which perceptions can be filtered according to their importance. Hence given a set of Perceptions (P), an affect preference (A) and a set of Beliefs (B), the updating of a set of Belief propositions (r, s, q, p, ¬p) occurs as follows:

$P := \{r, s, q, p\}$

$A := \{importance(high, p), importance(low, r)\}$

$B := \{ \neg p\}$

$A \otimes P \rightarrow AP$

$AP := \{p, s, q, r\}$

$AP \otimes B \rightarrow B'$ (Resolution: $\{X_{t+1}, \neg X_t\} \rightarrow \{X_{t+1}\}$)

$B' := \{p, s, q, r\}$

The perception set, P, is modified by the Affect preferences (A) to produce AP. This contains the same perceptions as before, but, the order in which the perceptions are processed is changed according to the affective affordance attached to each one. The new belief set, B′, contains the perceptions which have been processed in the order that accords with their significance to that individual. The affect ordered belief set then drives the Desire and Intention reasoning schema, leading to the selection of action that reflects the current focus of the agent at the Belief level. The latest versions of CAMAL [54] extend this a-CRIBB BDI model further with the affect model being used to drive the entire BDI reasoning cycle. This allows Beliefs to be given a likelihood of belief measure (based on affordance and affect models), Desires (or goals) an importance and Intentions (or associations) an insistence value.

7. Linking Perception, Affect and Cognition

This research into the development of architectures for synthetic intelligence is built around the concept of control states, and in particular motivators. The primary conjecture is that the design and implementation of such architectures can proceed using a systematic control language based on affect. This control language should be domain-less to the extent that it can be used across multiple non-contingent domains (and applications) and yet be capable of capturing the control metaphors most suitable for any specific domain. This control language is grounded in the use of affect and consistent across different domains, tasks and levels of processing. The computational work is being developed with no requirement for emotion but rather a reliance on affect (a valencing of and within internal processes) and affordance (a motivational based perception of events) that together can be used to guide both internal and external acts.

One problem addressed in the design and constructed architectures to date has been relating action, perception and feedback from action via perception to the constructs that initiated action. It is suggested that the type of connectionist perceptual system suggested by Barsalou [40] may (ultimately) offer the means to provide such links. Although connectionist approaches to modelling some or all of the sub-architectures supporting motivational constructs have not been excluded, and may offer deeper conceptual and implementation scope, the approach so far makes use of behaviour-based and symbol architectures.

Consider the a-CRIBB model described above, but with the addition of motivational constructs as the core representational schema [38, 54]. The motivational construct links perception, belief, desire, intention and actual action as is evident from the components associated with the construct given in Table 1, modulated by the affective (BDI) reasoning developed from the a-CRIBB research. This allows us to define the information and knowledge constructs shown in Figure 1. Activity and behaviour at one level is represented and controlled at other layers, and captured in the motivational structure. We have extended the a-CRIBB model to produce associations. Associations are a construct that consist of a belief, a desire, an intention, and an association value (*insistence*). The associations provide an indication of the past success of a specific set of plans given the agent's current beliefs and desires. This allows the agent to consistently determine the most appropriate set of plans based on its beliefs and desires; and to modify them where they fail. Associations take the general form:

association(BeliefSet, Goal, BehaviourSpecification, Insistence)

Associations can be pre-defined (typically a small number related to high priority tasks for specific environment configurations), or formed when the architecture is initialised, or dynamically created. The a-CRIBB based BDI model drives the association model, so that motivators are created to reflect those associations whose belief basis is currently true, or if running in probabilistic mode, those beliefs with a high degree of belief. Motivator selection proceeds through the balancing of goal importance against association insistence; in effect, the likelihood of goal success given a belief basis using the association specified behavior. Alternative metrics can be used for motivator selection and current experimentation involves combining degree of belief (in a Belief), goal importance and association insistence. A further CAMAL project also incorporates degree of trust in perceptual models in motivator generation and selection. A motivational blackboard enables belief revision, goal review and the updating of associations to be coordinated. Behavioral feedback is filtered back to the relevant section of the blackboard. Affect provides the means whereby all these processes are valenced. The feedback induced revision of the affective values ensures adaptation.

8. Experiments

The CAMAL architecture (Figure 2) and variants has been extensively used in simulation, for example: multiple robot simulation test-beds [36, 48]; Five A-Side football [49, 42], Fungus Eater environments [50, 46]; and with physical robots [52]. Here experimental results from three series of experiments addressing different aspects of the architecture are presented. The first addresses the effect of adding affect to the BDI schema used in CRIBB. The second set of experimentation considers what happens as different parts of the architecture are used individually or disabled within the overall architecture. Both these series of experiments use a variation of the Fungus Eater test-bed. The last series of experiments uses the architecture to control a mobile robot.

8.1 Experiments with CRIBB and a-CRIBB

For this experiment CRIBB and a-CRIBB are compared using a standard agent test-bed (the Fungus Eater [55, 56]). Test-bed parameters can be systematically varied to control the type of agents present (CRIBB or a-CRIBB), the number of each agent, the number of Fungus, bad-fungus (poison), medicine (to negate the poison) and the agent's metabolism [50].

Table 2 (a,b,c) show results from an experiment where the number of agents exceeds what is environmentally viable. The statistics collected are: Energy left at the end of the time interval; amount of ore collected; and the survival time of the agent. The environment contains 10 fungus (10% bad fungus), 10 ore, 5 small fungus, 1 golden ore and 1 medicine object. Only one type of agent was used at a time. All the results are an average value from multiple runs and the time survived is shown as a percentage of the time interval that the agents survived on average.

The results from these experiments show that as the number of agents is increased, the amount of energy left decreases for both agent types. The a-CRIBB agents maintain their energy level between 80 and 60 which corresponds to high and low energy drive thresholds respectively. The amount of ore collected by any one agent decreases as the number of agents increases. The a-CRIBB agents out-perform the CRIBB agents in the amount of ore that is collected. a-CRIBB agents can survive for 100% of the time intervals, but as the number of agents increase the survival rate decreases. The a-CRIBB agents' survival rate reaches 92.64% at the lowest. The CRIBB agents, however, do not survive 100% of any time interval, and the survival rate substantially decreases as the number of agents increase. The results from these experiments led to the a-CRIBB BDI schema being incorporated into the design and implementation of CAMAL.

8.2 Experiments with a Society of Mind CAMAL Architecture

SMCA (Society of Mind CAMAL Architecture) is a redesign of CAMAL using the Society of Mind metaphor [1, 46, 47]. SMCA comprises of six reflexive, eight reactive, fifteen deliberative, nineteen perceptual, fifteen learning and fifteen meta-control behaviours. These behaviours are encapsulated as agents and activated using K-lines within SMCA. SMCA extended the CAMAL implementation with meta-cognition processes that enable the learning of goal directed behaviour; previous research had only referenced these capabilities at the theoretical and design stages, with only relatively shallow implementations.

SMCA was also tested using the Fungus Eater scenario. Test-bed parameters can be systematically varied to control the type of agents present (full CAMAL or partial CAMAL subtypes), the number of each agent, the number of Fungus, bad-fungus (poison), medicine (to negate the poison) and the agent's metabolism. Test-bed agents were given an artificial physiology, which causes the agent to defer from its primary goal of collecting Ore if energy was determined to be low, and then favour Fungus collection. The Society of Mind approach to the redesign and implementation of CAMAL enabled various components in the architecture to be deactivated, and so allow comparison of CAMAL with its components in stand-alone mode. Figure 3 and 4 show the learning agent compared to the whole architecture. As the figures illustrate, in the initial stages the reinforcement learning agent was found to collect more fungus than the BDI agent. However over both goals and time, the more extensive SMCA (CAMAL) architecture provided far greater capability (as expected).

In the second set of SMCA results presented here, the architecture without any reflective layer (cognition1 with learning, meta-control and meta-cognition components missing) is compared with the full implementation. This addition of the meta-layers allows for CAMAL to differentiate between the various cognitive models, each favouring specific qualities such as physiological and/or goal-oriented behaviour. Again multiple runs of various scenarios allowed averaged results to be collected. The meta-cognition agents not only outlived the non-meta-cognitive agents (Figures 5 and 6), but the meta-cognition agent was able to collect 82% of resources, compared to 50% of resources collected by the cognitive agent

In the third set of SMCA results presented here, the architecture including meta-control with full deliberative and learning capabilities but without the meta-cognitive layer processes (Norm selection) is compared with the full implementation. The full implementation allows for the meta-cognitive layer to select high level attitudes (Norms) and so switch the deliberative context for the architecture to reflect current needs, activities and environment. Again multiple runs of various scenarios allowed averaged results to be collected. Again the architecture with the meta-cognition agents not only outlived the meta-control implementation (Figures 7 and 8), but the meta-cognition architecture collected 82% of resources, while the meta-control configuration collected 68% of resources.

Overall, these experiments demonstrate that the meta-cognition concept provides a powerful tool towards developing efficient computational models. Meta-cognition is used to specify Norms that allow the use of alternative BDI models and so change the operational mode of the CAMAL architecture with respect to planning, reasoning, decision making, problem solving, and learning.

8.3 Experiments with a mobile robot

The final experiment presented here gives summary results from the use of CAMAL to control mobile robots. robo-CAMAL [51] was developed to demonstrate how CAMAL could be used to resolve many of the issues associated with the anchoring problem [57], and how to learn to act and adapt in a physical environment. robo-CAMAL is an autonomous mobile robot that inhabits a bounded maze which can include specific known objects, e.g. further mobile (colour coded) robots and a ball, plus any other object, which is usually recognised as *unknown.*

The symbol grounding problem concerns the difficulties of generating symbols using perceptual systems, and the meaning of those symbols [58]. The anchoring problem is a subset of the grounding problem. It investigates how links are generated and maintained between symbols used within an agent's cognitive architecture, and the data obtained via the agent's perceptual system.

In terms of robo-CAMAL's learning ability, it uses an anchoring mechanism to identify the objectives of its goal, and the association model (the BDI couplings) to understand the consequences of its actions. As mentioned above, an association is a coupling between a belief, a desire, an intention, and a magnitude value (i.e. the association value). At a high level view robo-CAMAL goes through a number of phases:

- observe the environment using the perceptual anchoring mechanism;
- initiate a BDI association using the affect model;
- perform the action described by the BDI association, using the domain model details of the link between the intention symbol and the reactive control system;
- observe the environment once the action has been performed using the perceptual anchoring mechanism;
- feed the consequences of the action on the environment into the BDI association model, and modify BDI association value using reinforcement learning.

If the observed state of the environment after the action conforms to the required desire state, then the association mechanism increases the association value. However, if the environmental state after the action is not the required goal state, the association value is decreased. This process ensures that successful associations will develop higher association values, whilst unsuccessful associations will develop lower values.

Adaptation experiments were designed to determine if robo-CAMAL has the ability to modify its goals to reflect changes in its environment. The robo-CAMAL architecture was instantiated with the three goals hit(blueball), hit(redrobot), and hit(blackrobot). The correct associations were given to the architecture at start-up. robo- CAMAL was then allowed to run for three minutes in a variable environment. The environment contained any combination of the three possible objects blueball, redrobot, and blackrobot. The object combination was changed at intervals of one minute. At each deliberative cycle, the agent's internal deliberative state was recorded. This included the agent's current belief set, the association values, and the goal importance values. In addition, the number of actual collisions with the objects present was recorded. Some of the results of this experiment can be seen in figures 9 to 12.

Figure 9 shows the various found(X) beliefs present during one of the experimental runs. Each line represents one of the possible found beliefs. If a belief is present, it was given a value as indicated by the vertical axis in Figure 9. If the found(X) belief is not present, the value is zero. Figure 9 shows that initially robo-CAMAL holds the beliefs found(redrobot) and found(blackrobot) at various times. Then at point a, the belief found(blueball) is present. It is around this point that the found(redrobot) belief is no longer present. At point b the belief found(blueball) is no longer held. This coincides with the way the objects were varied within the environment. At first the redrobot and the blackball were present in the environment. After one minute the redrobot was replaced with the blueball. After the second minute the blueball was removed. This mirrors figure 9 in that the redrobot was found at the start of the run, the blueball was found during the middle of the run, and the blackrobot was present thought the whole of the run.

One point to note is the spacing of the found beliefs. The total number of deliberative processing cycles for this three minute experiment is around 250. As the environment was changed after each minute, it would be expected that the beliefs would alter to reflect that change at around 80 and 160 deliberative processing cycles. This however is not the case. The found(X) belief alters its profile at around 50 and 150 deliberative processing cycles. This is due to the way in which the deliberative and reactive levels of the architecture interact. The deliberative component sets the reactive component to run for a number of reactive cycles. The reactivate level returns control when an event occurs, or when it completes the required number of reactive cycles. This means that when there are few events occurring in the agent's local vicinity, the deliberative component is not as active. Therefore at the deliberative level, the number of processing cycles per minute is dependent on the number of events that occur within that minute.

Figure 10 shows the importance value for each goal over time. Initially robo-CAMAL achieves the hit(blackrobot) and hit(redrobot) goals. This can be seen at points a and b. At these points the importance value for each goal jumps to a value of 0.52. Once a goal has been achieved, its importance value is set to 0.5. The step increase for the goal importance value is 0.02. The goal importance value does not get recorded until after it has been incremented. This means that when the importance value jumps to, or remains at 0.52, the relevant goal has been achieved. In addition, if a goal fails, its importance value is reduced to 0.1. Therefore, in figure 10 if the goal importance value drops to 0.12 that goal has failed. This failure can be seen at points c, d, and e where each of the three goals fail. Point f shows the success of the goal hit(blueball). This corresponds well with when the redrobot was removed and the blueball added.

An important point to note can be seen at point g. Here the importance value of the goal hit(blueball) increases beyond 0.7. One question is why has robo-CAMAL not attempted to achieve this goal when its importance value is so high in comparison to the others? The answer can be found by examining the association values. These can be seen in figures 11 and 12. A motivator is chosen based on not only the goal importance, but the likelihood of successfully achieving the goal (i.e. the association with the highest value and the corresponding goal and current belief). Figures 11 and 12 clearly show that the association value related to the goal hit(blackrobot) is significantly higher than the one related to the goal hit(blueball). This means that when combining the two values, the motivation to hit the blackrobot is greater than the one to hit the blueball. Therefore robo-CAMAL continues to attempt the goal hit(blackrobot). This result is as expected. Even though robo-CAMAL believes, correctly, that the blueball is present in the environment, it also believes that it is more likely to achieve the goal hit(blackrobot). This was the reason for including the association value in the mechanism to decide robo-CAMAL's current BDI motivation.

The same reasoning can be applied to robo-CAMAL's behaviour after 150 processing cycles. It is clear from Figure 10 that at points h, i, and j, robo-CAMAL attempts the other goals. At these points the goal importance value becomes so great that it outbalances the low association values. This result is again expected and was the reason for incrementing the goal importance value. This ensures that robo- CAMAL periodically attempts a previously failed goal to see if it has become achievable.

These result show that robo-CAMAL has the ability to adapt to a variable environment, and attempt the goals it believes achievable at the right time. These and other experiments [51] demonstrated that robo-CAMAL is capable of learning general behaviours. That is, with no understanding about the nature of its actions, robo-CAMAL can successfully learn which action to perform in order to achieve a specific goal. This is accomplished by performing and observing the effect of its actions on the environment. robo-CAMAL successfully adapted to its environment when it held the correct beliefs. However, when the agent formed the incorrect beliefs it failed to correctly modify its goals to reflect the actual environment. Instead it modified its goals to coincide with its incorrect beliefs. In general the results show that robo-CAMAL performed as expected, and, yet again, substantiates the adoption of the affect model to valence the BDI schema.

9. Discussion

This work has now taken a number of new directions as coworkers pursue their own agendas. These new directions take the design associated with CAMAL and the concept of an underlying affect mechanism that can be used to compare process priority, or rank goals and intentions but reframe the research according to specific interests. The issues confronted and results from implementation and experimentation inform the central project.

In applying the Society of Mind approach to cognition [1] to the CAMAL architecture [47], the SMCA experiments show that metacognition agents have the organizational intelligence and (optimal) behaviours to constitute a "Society of Mind". The metacognition agents demonstrate a complete control mechanism in managing an (affective) metabolism, and balancing resource and task motivations. The agents exhibit optimal decision making capabilities near a decision variable boundary, using mechanisms based on norms (*attitudes*) and affect. Through the meta-level control of their different Belief-Desire-Intention models, at any given point, some agents in the Society of Mind are active, while others are static. This arrangement of activities within their planning, reasoning, decision making, self reflection, problem solving and learning capabilities, from different combination of agents proves the viability of the concept of metacognition as a powerful catalyst for control and self-reflection. SMCA, with its metacognition selected alternative BDI models, demonstrated task effectiveness, goal achievement, and the ability to perform well in novel situations.

The mapping of the architecture from simulation, using a variety of test-beds, to robotics has raised a number of interesting observations [51, 54]. While predominantly successful in the way it performed, robo-CAMAL failed in two ways. The first was due to the opportunistic learning mechanism implemented. This meant that robo-CAMAL in some circumstances was unable to optimise its behaviour. The second failure was due to the limitations of the anchoring mechanism in robo-CAMAL (specifically the shallow vision system). In some instances, robo-CAMAL mistook the blueball for the blackrobot. When this occurred the agent failed to modify its goals to reflect the actual environment. This work is now being developed further using a Bayesian approach to formalize the affect model, and with a new camera and vision system to be used in conjunction with a deeper perceptual (anchoring) model. This new perceptual anchoring model will combine the improved sensors with neural learning mechanisms and address some of the issues raised by Barsalou [40] and van der Velde and de Kamps [59]. While the connectionist research domain has moved on from the critical analysis of Fodor and Pylyshyn [60], many difficult questions still need to be addressed, or revisited in light of the change in architecture design. van der Velde addresses the symbol grounding problem for the domain of natural language understanding, but for CAMAL in dynamic real-world environments involving motivated interactions, the issues will be different again. Learnt neural perceptual systems will typically store exemplar representations of specific environmental actors and objects, and the linking to specific symbols in the CAMAL BDI and motivational model may require adaptive real-time parameterization as the architecture shifts its focus of attention from say the black-robot in the northwest corner of the environment to the black-robot near the east wall. Similar issues are identified by Franklin [61]. Two particular thorny issues are:

- What dimensions to this type of perceptual symbol system can exist and why? Note the requirements for a psychological theory may well differ from a theory of use to cognitive science and differ again from the requirements of a theory of cognitive architectures.
- What is the minimal set of dimensions or qualities to this type of perceptual system required to build cognitive architectures? What sort of computational experiment will demonstrate the usefulness (or otherwise) of this approach? Most standard test-beds are limited in their usefulness as identified by Hawes et al. [62], and Hanks et al. [63] earlier, in that while single task comparisons allow direct comparison, the nature of research and implementations at different institutes will necessarily invoke differences that undermine meaningful comparison of experimental results.

10.Conclusion

There are many views on the place of emotion and affect and indeed what constitutes a mind or an architecture for cognition. CAMAL pursues a perspective informed by motivational and affective control states mitigated by cognitive models of reasoning and learning. This research pursues the following three major objectives:

- A synthesis of concepts based on an analysis and investigation of how different perspectives on autonomy, affect and motivation map onto computational frameworks.
- Insights into the nature of heterogeneous mechanisms across different processing, representational and knowledge levels, and their explanatory role in describing mental phenomena.
- A developing framework enabling reactive and emergent behaviours to be combined with the more abstract decision making processes associated with cognitive agents.

Affect in SMCA provides the basis of an agent-internal micro-economy which is used as a decision metric, with affective values used as a currency. This rationalization of affect across the architecture, and as also used in the other experiments, has enabled decision-making and behaviour selection, across all levels and columns of the architecture to be grounded in a consistent and adaptable mechanism. This has allowed artificial life and emergent behaviours to be combined with reactive and the more abstract decision making processes associated with cognitive architectures. As we develop the architecture and experiment with further test-beds, the limitations and benefits of this approach provide pointers for further work.

The metacognition concept provides a powerful tool towards developing efficient and quality computational models. The SMCA research investigates the concept of metacognition as a powerful catalyst for control, unification and self-reflection. Metacognition is used on BDI models (via the use of norms) with respect to planning, reasoning, decision making, self reflection, problem solving, learning and the general process of cognition to improve performance. The reactive class of agent, in turn, provides a computational platform for the deliberative agents. The design of deliberation mechanisms for the fungus testbed includes five different types of BDI agents. The BDI determines which of the reactive or reflexive control mechanisms are active according to the goals to satisfy. These goals are either task related or agent-internal resource related, and determine the number of different types of reflexive and reactive agent required for this specific testbed. The benefits are shown in the results comparing cognition and metacognition.

If we address the type of emotional processes highlighted by Scherer [9], and consider the affective processing designed and implemented in the various CAMAL architectures, we can find a good correspondence. Taking the five functionally defined systems individually, CAMAL provides for information processing over perception. Whether in simulated or robotic worlds, sensory information is mapped onto belief structures. We differentiate between the source of beliefs (e.g. assumptive, deductive or perceptual), and (if required) allow beliefs related to currently important objects (and actors) to be treated differently. We have experimented with a-life and (simulated) physiological states and allow the regulation of internal affective states. The affective importance measure allows CAMAL to rank goals and allow decision making over competing motives. The affective insistence measure allows the control of external behaviour through the building of associations that link beliefs, goals and intentions. Based on the results of the a-CRIBB experiments, CAMAL modulates the rationality of its BDI schemas with affective mechanisms, allowing goal and behaviour adaption to current world and internal states. Finally the results of actions, the success or failure of a motivator linking belief, goal and association enables the "cognitive" feedback system to inform these four processes over a reasoning cycle. And this can be done without using shallow interpretations of "fear", "surprise", "anger" etc., or a any need for an emotion system – affect suffices.

The work on the architecture with robots has highlighted the issues associated with symbol meaning, and the nature of problems when the architecture has to link symbol grounding and learning. The type of perceptual system suggested by Barsalou and van der Welde may offer the means to provide such links in a non-superficial manner. Although connectionist approaches to modelling some or all of architectures supporting motivational constructs have not been excluded, the approach so far makes use of behaviour-based and symbol architectures. This raises questions to be addressed in current and future research, but as Franklin [61] suggests the combination of neural, reactive and deliberative mechanisms offers great promise for tackling the issues raised by the papers referenced in the first section of this paper.

Finally, there are controversies about the terminology used in this area of research. Motivation, drive, goal and affect are used to refer to and mean a number of different things. Konidaris and Barto [64] make the point that a motivational system is central to agent autonomy. Like Stoytchev and Arkin [65], they use a low level (quantitative) drive system which is used to provide metrics for a reinforcement learning system There is no universal definition of these terms across (or even within) the fields of philosophy, psychology, cognitive science and artificial intelligence. And this is particularly so for emotion [33, 34, 66].

As our work moves forward, we may need to address whether the necessity for symbol meaning (arising from perceptual grounding) requires that the semantic depth of the architecture is sufficient for emotion tags such as “surprise” to be used. It may well prove that until the architecture is sufficiently advanced with linguistic or communication capabilities, such as detailed in [67], we will have no need to use emotion, and rely purely on the affect model to drive cognition.

References

1. Minsky, M.L. The Society of Mind, William Heinemann Ltd., 1987.
2. Sloman, A. Beyond shallow models of emotion, *Cognitive Processing, 2(*1):177-198, 2001.
3. Franklin, S. A consciousness based architecture for a functioning mind, *Proceedings of the AISB'00 Symposium on How to Design a Functioning Mind*, April 2000.
4. Robinson, P. and El Kaliouby, R. Computation of emotions in man and machines, *Phil. Trans. R. Soc. B December 12, 2009 364:3441-3447*.
5. Vallverdú, J. and Casacuberta, D. Handbook of Research on Synthetic Emotions and Sociable Robotics, Idea Group Inc (IGI), 2009.
6. Norman, D.A. Twelve issues for cognitive science, *Cognitive Science, 4:1-33*, 1980.
7. Simon, H.A. Motivational and emotional controls of cognition, Reprinted in Models of Thought, Yale University Press, 29-38, 1979. (originally Psychological Review, 74(1):29-39, 1967)
8. Ortony, A., Clore, G.L. & Collins, A. *The Cognitive Structure of Emotions*. Cambridge University Press, 1988
9. Scherer, K. Toward a Concept of 'Modal Emotions'. In P. Ekman and R. Davidson (Eds): Nature of Emotion, Oxford University Press. New York, 1994.
10. Barrett, L.F., Lindquist, K.A. and Gendron, M. Language as context for the perception of emotion, Trends in Cognitive Sciences, 11(8): 327-332, August 2007.
11. Neviarouskaya, A. AffectIM: An Avatar-based Instant Messaging System Employing Rule-based Affect Sensing from Text, Dissertation, University of Tokyo, 2008.
12. Chanel, G, Kierkels, J.J.M., Soleymai, M. and Pu, T. Short-term emotion assessment in a recall paradigm, International Journal of Human-Computer Studies, 67(8): 607-627, 2009.
13. Arrabales, R., Ledezma, A. and Sanchis, A. Towards Conscious-like Behaviour in Computer Game Characters, IEEE Symposium on Computational Intelligence and Games, 2009.
14. James, W. What is an Emotion? *Mind.* 9, 188-205, 1884.
15. Plutchik, R. *The Psychology and Biology of Emotion.* Harper Collins. New York, 1994.

16. Darwin, C. The Expression of Emotions in Man and Animals. London. Murray, 1872.

17. Ekman, P. The Nature of Emotion: Fundamental Questions. Oxford University Press. New York, 1994.

18. Scherer, K. Appraisal Considered as a Process of Multilevel Sequential Checking. In K. Scherer, A. Schorr, T. Johnstone (Eds). *Appraisal Processes in Emotion.* Oxford University Press. New York, 2001.

19. Oatley, K. *Best Laid Schemes*. Cambridge. Cambridge University Press, 1992.

20. Frijda, N. *The Emotions*, Cambridge University Press, 1986.

21. Frankel, C.B. and Ray, R.D. Emotion, intention and the control architecture of adaptively competent information processing. Symposium on How to Design a Functioning Mind, AISB'00 Convention, April 2000

22. Davis, D.N. Multiple Level Representations of Emotion in Computational Agents, Emotion, Cognition and Affective Computing, AISB2001:Agents and Cognition, University of York, 2001. isbn: 1 902956197

23. De Houwer, J. and Hermans. D. Cognition & Emotion: Reviews of Current Research and Theories, Psychology Press, 2010.

24. Duffy, E. *Activation and Behaviour*, Wiley & Sons, 1962.

25. Wollheim, R. On The Emotions, Yale University Press, 1999.

26. Picard, R. *Affective Computing*, MIT Press, 1997.

27. Rolls, E.T. The Brain and Emotion, Oxford University Press, 1999.

28. Freud, S. (originally 1894) The Defence Neuro-psychoses. In E. Jones (Ed), Sigmund Freud: Collected Papers. BasicBooks. New York. 1959.

29. Clore, G.L. Why Emotions are Never Unconscious. In P. Ekman and R. Davidson (Eds). Nature of Emotion. Oxford University Press. New York, 1994.

30. Sloman, A. and Croucher, M. Why robots will have emotions. *Proceedings of IJCAI87, 197-202, 1987.*

31. Sloman, A. The mind as a control system. In Hookway, C. and Peterson, D. editors, Philosophy and Cognitive Science. Cambridge University Press, 1993.

32. Phelps E. Emotion and cognition: Insights from studies of the human amygdala. Annual Review of Psychology, 2006, 24(57):27–53.

33. Ziemke, T. and Lowe, R. On the Role of Emotion in Embodied Cognitive Architectures: From Organisms to Robots, Cognitive Computation, 2009, 1:104–117

34. Gros, C. Cognition and Emotion: Perspectives of a Closing Gap Cognitive Computation, 2010, 2(2): 78 – 85.

35. Cohn, J.F. Foundations of human computing: facial expression and emotion**.** Proceedings of the 8th international conference on Multimodal interfaces, Banff, Alberta, Canada, pp.233 – 238, 2006

36. Davis, D.N. Control States and Complete Agent Architectures, *Computational Intelligence, 17(4):621-650*, 2001

37. Gibson, J.J. The Ecological Approach to Visual Perception, Boston: Houghton Mifflin, 1979.

38. Davis, D.N. and Lewis, S.C. Computational Models of Emotion for Autonomy and Reasoning. *Informatica (Special Edition on Perception and Emotion Based Reasoning)*, 27(2):159-165, 2003.

39. Beaudoin, L. *Goal Processing in Autonomous Agents*. Ph.D. Thesis, Cognition and Affect Group, Computer Science, University of Birmingham, 1994.

40. Barsalou, L.W. Perceptual symbol systems. Behavioural and Brain Sciences, 22, 577-609, 1999.

41. Sowa, J.F. and Majumdar, A.K. Analogical Reasoning. In: Conceptual Structures for Knowledge Creation and Communication, Springer Berlin / Heidelberg, 2003.

42. Bourgne, G. Affect-based Multi-Agent Architecture (for a 5-aside football simulation), Thesis, Computer Science, University of Hull, 2003.

43. Norman, D. A. & Shallice, T. Attention to action: Willed and automatic control of behaviour. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.) Consciousness and self-regulation, Vol. 4 (pp. 1-18). New York: Plenum Press. 1986.

44. Gat, E. Three-layer architectures. In Kortenkamp, D., Bonasso, R.P. and Murphy, R. editors, *Artificial Intelligence and Mobile Robots: Case Studies of Successful Robot Systems,* pages 195-210. MIT Press, Cambridge, MA, 1998.

45. Singh, P. and Minsky, M. An Architecture for Combining Ways to Think, Proceedings of the International Conference on Knowledge Intensive Multi-Agent Systems, 2003.

46. Vijayakumar. M.V. A Society of Mind Approach to Cognition and Metacognition in a Cognitive Architecture, PhD Thesis, Computer Science, University of Hull, August 2008.

47. Davis, D.N. and Vijayakumar M. V. A. "Society of Mind" Cognitive Architecture based on the Principles of Artificial Economics, International Journal of Artificial Life Research. Volume 1(1): 51-71, January-March 2010.

48. Davis, D.N. Reactive and motivational agents. In: *Intelligent Agents III*, J.P. Muller, M.J. Wooldridge & N.R. Jennings (Eds.), Springer-Verlag, 1996.

49. Nunes, H.A. Investigation of Motivation in Agents Using Five-Aside Football, M.Sc. Thesis, Department of Computer Science, University of Hull, 2001.

50. Lewis, S.C. Computational Models of Affect, Ph.D. Thesis, Department of Computer Science, University of Hull, 2004.

51. Gwatkin, J. Robo-CAMAL: Anchoring in a Cognitive Robot, PhD Thesis, Department of Computer Science, University of Hull, July 2009

52. Bartsch, K. and Wellman, H. Young children's attribution of action to beliefs and desires. *Child Development*, *60,* 946-964, 1989.

53. Wahl, S. and Spada, H. Children's Reasoning about intentions, beliefs and behaviour. *Cognitive Science Quarterly, 1*, 5-34, 2000.

54. Davis, D.N. Linking perception and action through motivation and affect. Journal of Experimental & Theoretical Artificial Intelligence, Vol. 20, 1, March 2008, pages 37 - 60.

55. Toda, M. The Design of a Fungus-Eater: A Model of Human Behaviour in an Unsophisticated Environment. Behavioural Science. 7:164-183, 1962.

56. Wehrle, T. New Fungus Eater Experiments. In P. Gaussier and J. Nicoud (Eds). From Perception to Action. IEEE Computer Society Press, Los Alamitos, 400-403, 1994.

57. Coradeschi, S. and Saffiotti, A. An introduction to the anchoring problem. Robotics and Autonomous Systems, 43:85-96, 2003.

58. Harnad, S. The symbol grounding problem. Physica D, 42:335346, 1990.

59. van der Velde, F. and de Kamps, M. Learning of control in a neural architecture of grounded language processing, Cognitive Systems Research, 11(1):93-107, 2010.

60. Fodor, J.A. and Pylyshyn, Z.W. Connectionism and Cognitive Architecture: A Critical Analysis, Cognition. 28:3-71, 1988.

61. Franklin, S. Perceptual Memory and Learning: Recognizing, Categorizing, and Relating. Symposium on Developmental Robotics: American Association for Artificial Intelligence (AAAI), Stanford University, Palo Alto CA, USA, 2005.

62. Hawes, N. Wyatt, J. and Sloman, A. Exploring Design Space for An Integrated Intelligent System. Knowledge-Based Systems, 22(7), pages 509--515, Elsevier. (Published as a best paper from Artificial Intelligence 2008 (AI-2008), The 28th SGAI International Conference on Artificial Intelligence). September 2009.

63. Hanks S. Pollack M. and Cohen P.R. Benchmarks, Test-beds, Controlled Experimentation, and the Design of Agent Architectures, AI Magazine, 14(4):17-42, 1993.

64. Konidaris, G. and Barto, A. An Adaptive Robot Motivational System, From Animals to Animats 9, Lecture Notes in Computer Science, Springer Berlin / Heidelberg 1611-3349 Volume 4095, 2006

65. Stoytchev, A. and Arkin, R. Incorporating Motivation in a Hybrid Robot Architecture, Journal of Advanced Computational Intelligence and Intelligent Informatics, 8(3): 269-274, May 2004.

66. Charland, L.C. Emotion as a natural kind: Towards a computational foundation for emotion theory, Philosophical Psychology, Vol. 8, No. 1, 59-85, 1995.

67. Loula, A., Gudwin, R., El-Hani, C.N. and Queiroz, J. Emergence of self-organized symbol-based communication in artificial creatures, Cognitive Systems Research, 11(2):131-147, 2010.

Figures

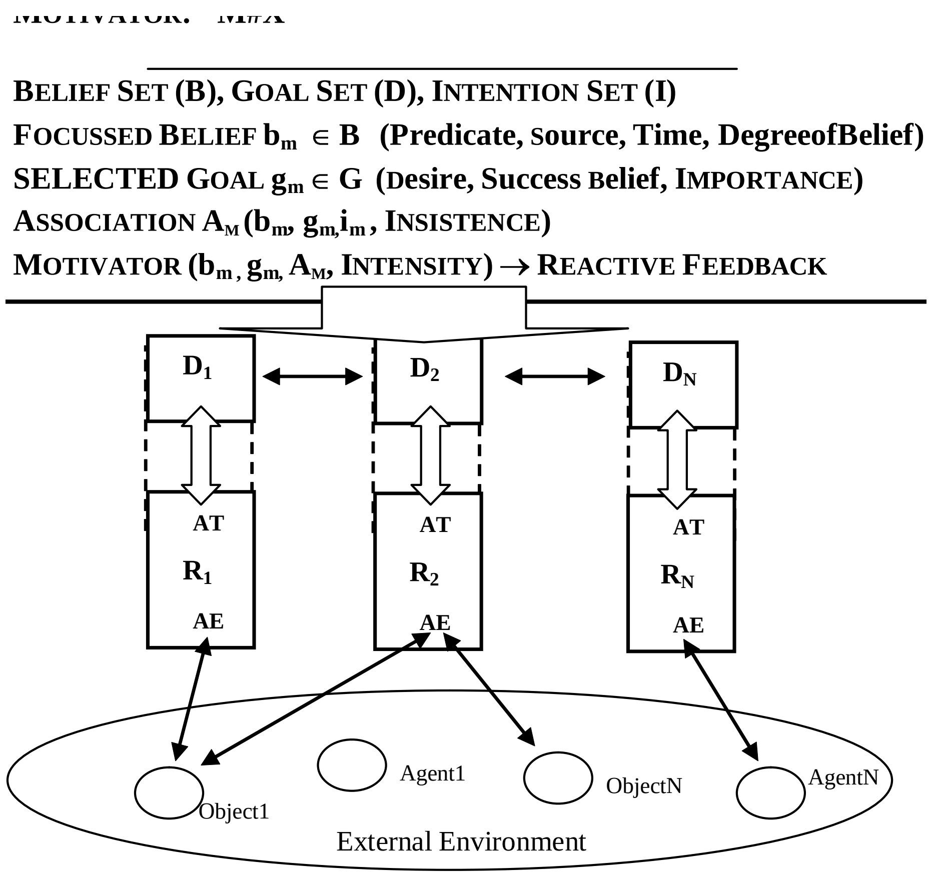

Figure 1. Motivator coupling of BDI and affect in CAMAL enabling the linking of Reactive (R_1, R_2, R_N) and Deliberative (D_1, D_2, D_N) processes through Affect (AT), and the mapping of Affordances (AE) onto Actors and Objects in an external environment

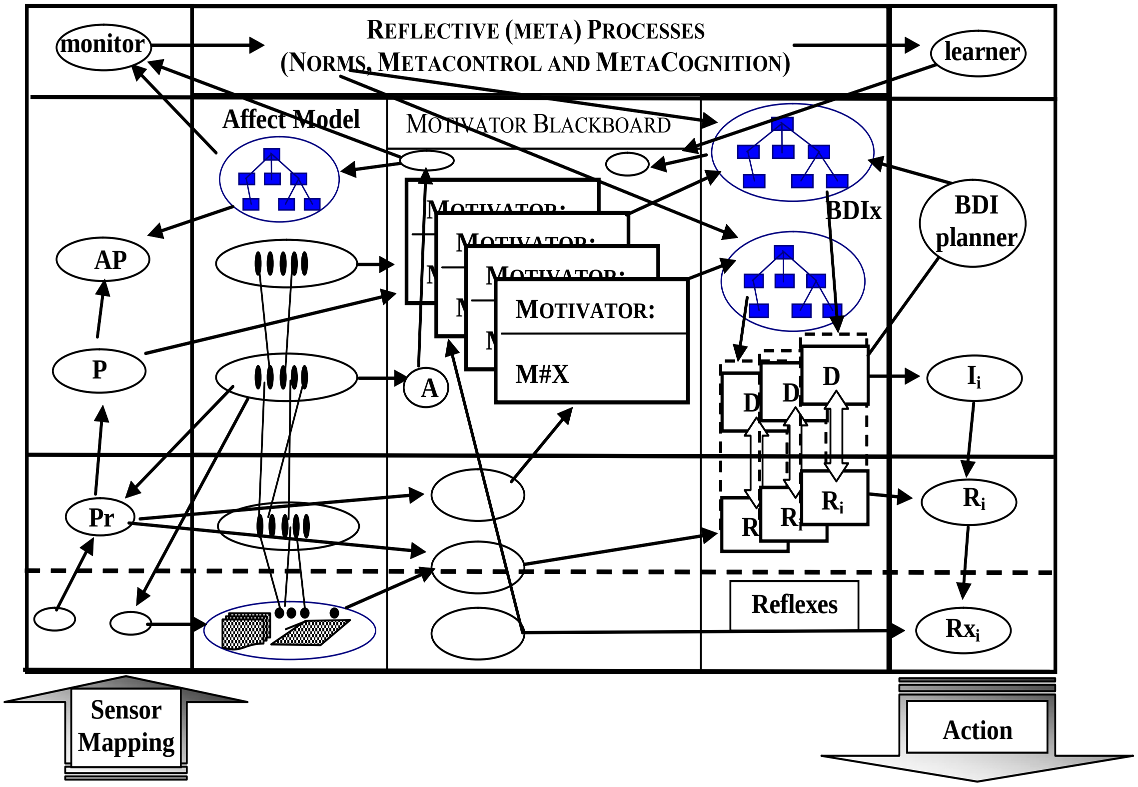

Figure 2. The four-layer, five-column CAMAL (Cognitive Architecture for Motivation, Affect and Learning) model.

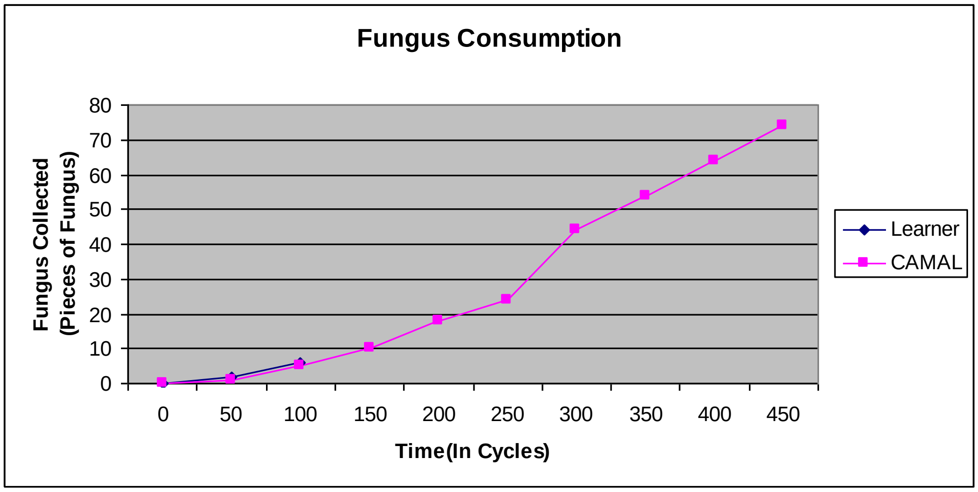

Figure 3: Fungus Consumption for Reinforcement Learner and SMCA (CAMAL) agents.

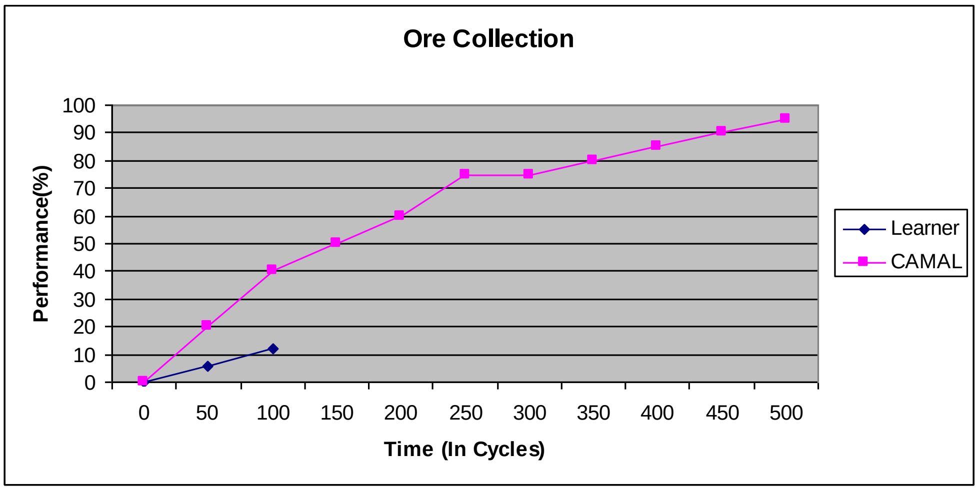

Figure 4: Ore Collection for Reinforcement Learner and SMCA (CAMAL) agents.

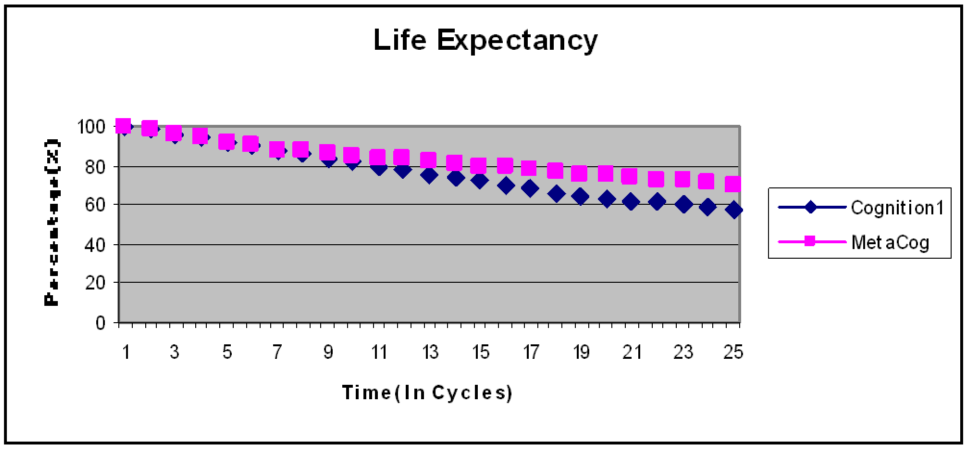

Figure 5: Life expectancy for Cognitive and Meta-cognitive agents.

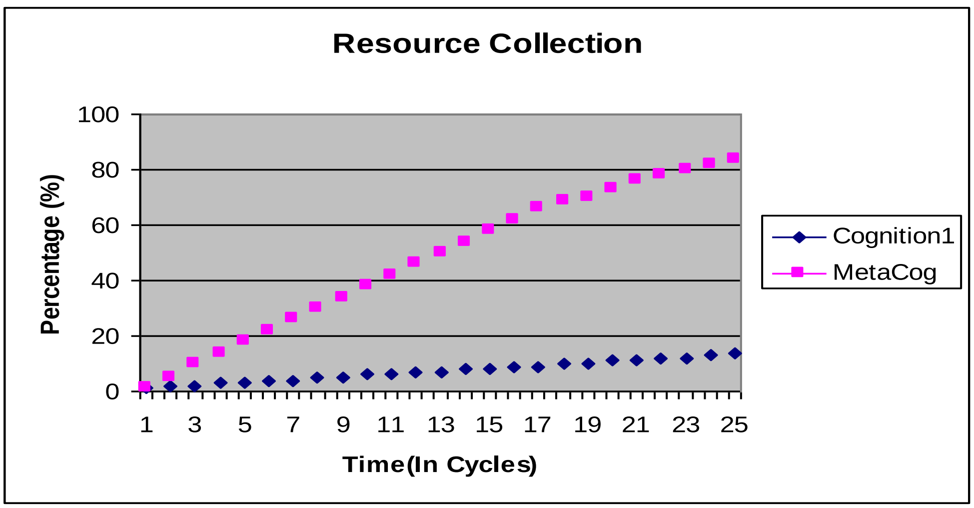

Figure 6: Fungus and Ore Collection for Cognitive and Meta-cognitive agents.

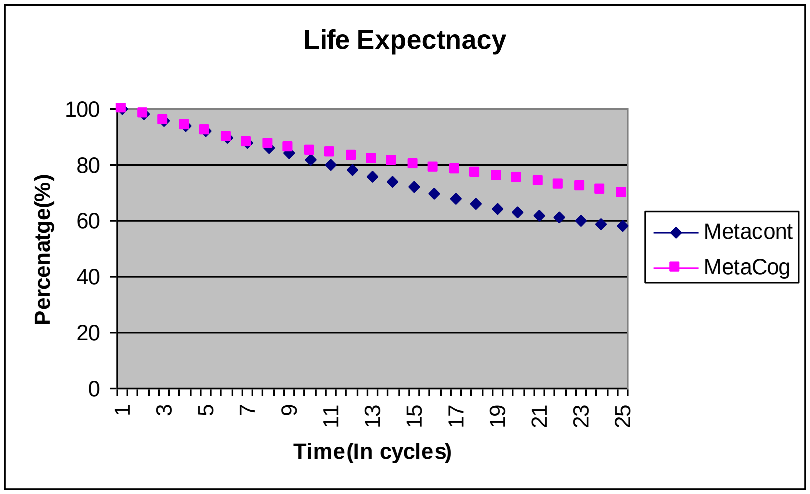

Figure 7: Life expectancy for Meta-control and Meta-cognitive architectures.

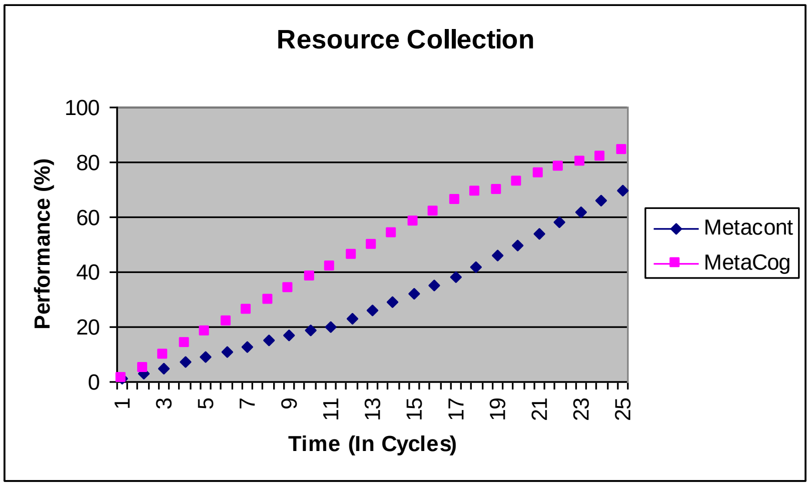

Figure 8: Fungus and Ore Collection for Meta-control and Meta-cognitive architectures.

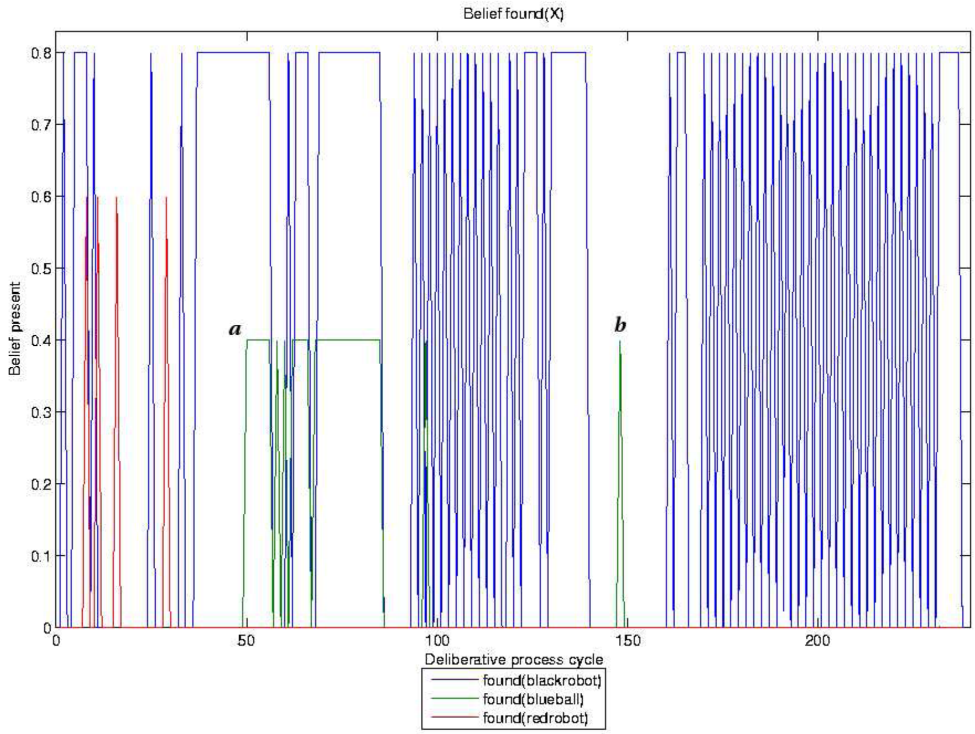

Figure 9: Periods when the belief found(X) *was present in robo-CAMAL adaptation experiment.*

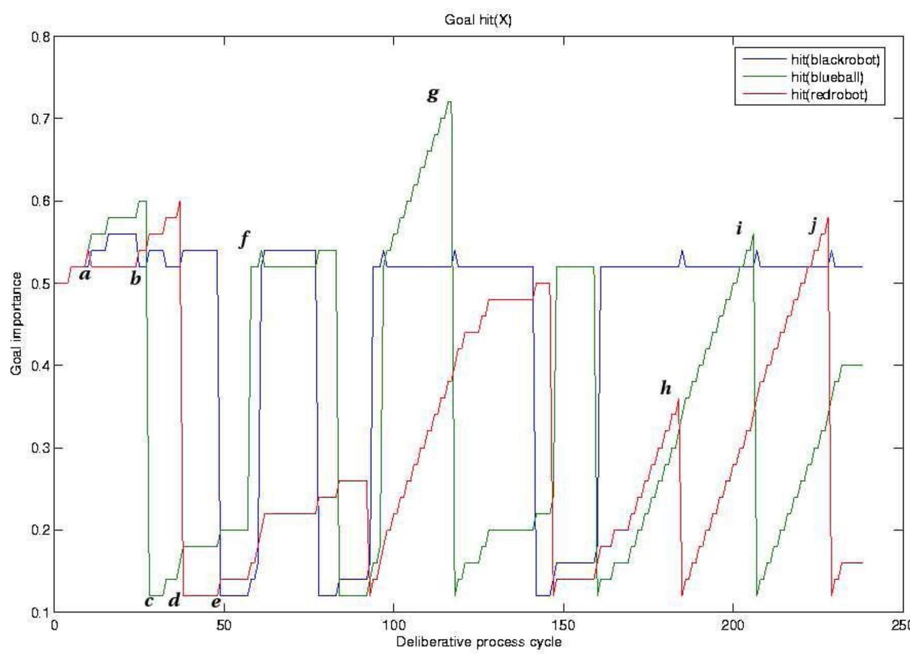

Figure 10: Goal importance value in robo-CAMAL adaptation experiment.

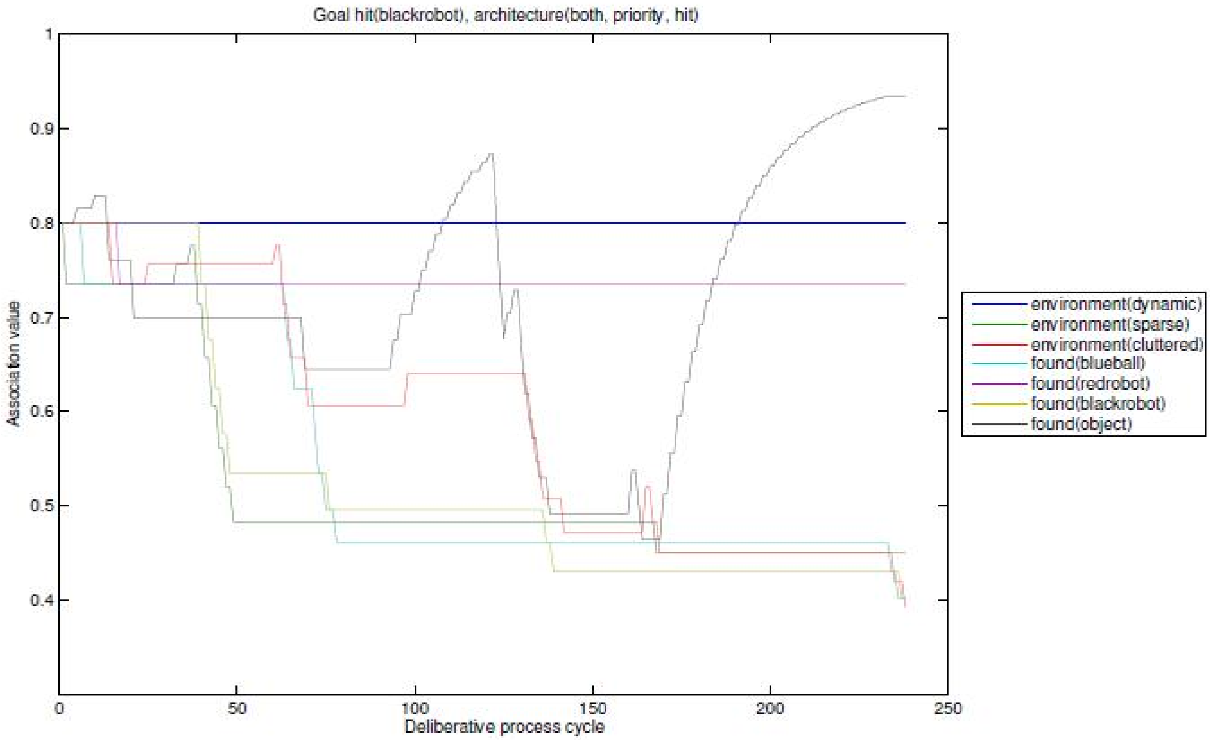

Figure 11: hit(blackrobot) association value in robo-CAMAL adaptation experiment.

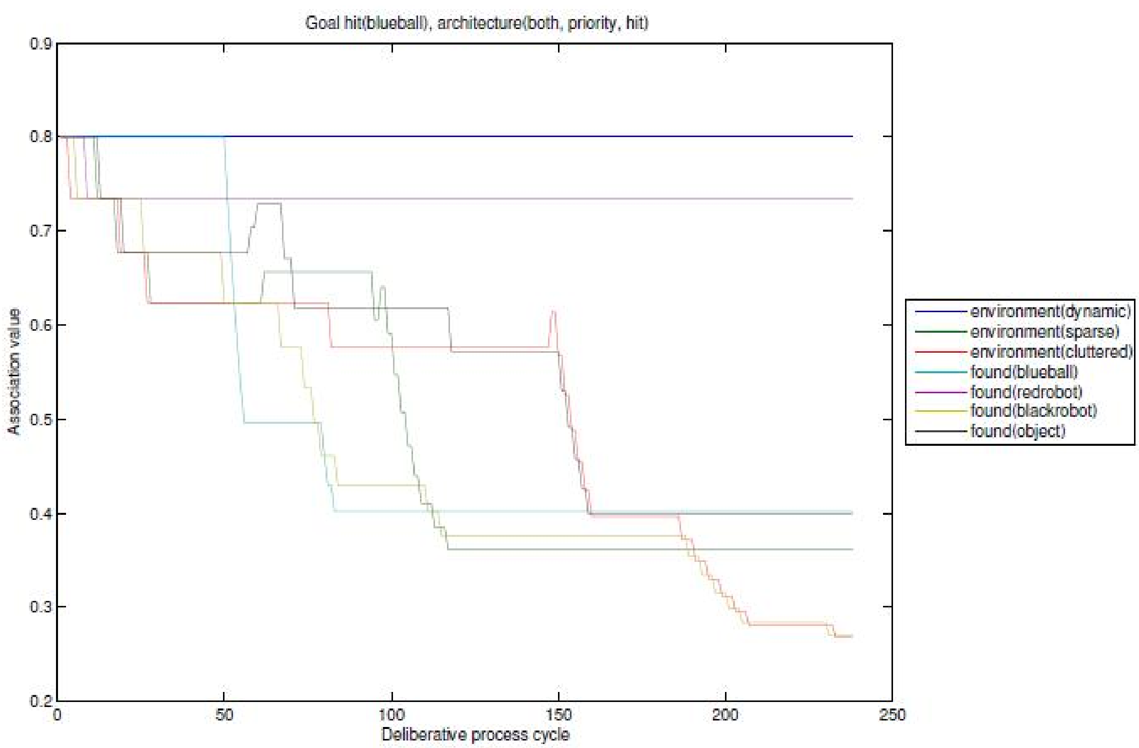

Figure 12: hit(blueball) association value in robo-CAMAL adaptation experiment

Tables

Component	**Meaning**
Actors and Entities	Other agents (actors) and objects referenced by this motivator
Belief Indicator	Indication of current belief about the status of semantic content P: e.g. *true*, *partially true*, *false* (later CAMAL also uses Real valued *DegreeofBelief*)
Commitment Status	The current status of the motivator, e.g. *adopted*, *rejected*, *undecided*, *interrupted*, *stalled*, *unconsidered*, *completed*.
Decay Function	Defines how insistence decreases while motivator is not *adopted*.
Dynamic State	The process state of the motivator e.g. *being considered*, *nearing completion*, etc.
Affect Key	Processing keys to the affective model and their situational triggers for the motivator.
Importance Value	Importance (e.g. *neutral, low, medium, high, unknown*). This may be intrinsic or based on an assessment of the consequences of doing or not doing
Insistence Value	Heuristic value determining interrupt capabilities. This should correspond to a combination of the motivator's importance and urgency.
Intensity	This influences the likelihood of (continuing) to being acted on.
Management Information	The state of relevant management and meta-management processes.
Motivational Attitude	The motivator's attitude to semantic content P : *make true*, *keep true*, *make false* etc.
Plan Set	Possible plan or set of plans for achieving the motivator.
Rationale	If the motivator arose from explicit reasoning – motivators need not.
Semantic Content	A proposition P denoting a possible state of affairs, which may be *true* or *false* (as given by *Belief Indicator*)
Urgency Descriptor	How urgent is this descriptor – this may be qualitative (e.g. *high*, *low*) or quantitative (for example a time-cost function).

Table 1. The Full Motivational Construct

No of Agents	*10 Time Intervals*		*20 Time Intervals*		*50 Time Intervals*	
	CRIBB	a-CRIBB	CRIBB	a-CRIBB	CRIBB	a-CRIBB
2	32.5	79.5	20	80	4	77
4	29.3	76.5	9.5	72.5	3.5	71.75
8	33.63	76.5	8	64.57	21.25	65
10	29.3	65.6	5.5	65.6	0.6	62.1

Table 2a. a-CRIBB Fungus Eater Experiments (Energy)

No of Agents	*10 Time Intervals*		*20 Time Intervals*		*50 Time Intervals*	
	CRIBB	a-CRIBB	CRIBB	a-CRIBB	CRIBB	a-CRIBB
2	3.85	6.25	5.65	9.75	6.65	13.4
4	3.44	4.6	4.93	7.95	5.98	14.85
8	3.05	3.54	4.66	6.44	5.63	10.54
10	3.12	2.97	4.57	6.11	5.22	10.96

Table 2b. a-CRIBB Fungus Eater Experiments (Ore Collected)

No of Agents	*10 Time Intervals*		*20 Time Intervals*		*50 Time Intervals*	
	CRIBB	a-CRIBB	CRIBB	a-CRIBB	CRIBB	a-CRIBB
2	92	100	70.5	100	41.2	100
4	91.4	100	66	100	33.06	100
8	88.6	99.4	62.15	96	33.6	94.36
10	89.5	98.2	58.7	95.65	28.64	92.64

Table 2c. a-CRIBB Fungus Eater Experiments (Life expectancy as % of original population)