

Linking perception and action through motivation and affect

Darryl N. Davis

Computer Science, University of Hull

Contact :-

Dr Darryl N. Davis

Computer Science

Faculty of Science

University of Hull

Kingston-upon-Hull

HU6 7RX, UK

Email: D.N.Davis@hull.ac.uk

Tel: +44 (0) 1482 466469

Fax: +44 (0) 1482 466666

Abstract

Cognition involves the control of behaviour both within external environments and internal to an autonomous system. The control of behaviour is vital to an autonomous system as it responds to its external environment and acts to further its goals. Excepting the most spartan of environments, the potential available information and the associated combinatorics in a perception, cognition and action sequence can tax even the most powerful of symbol processing agents. This research into cognitive architectures is described within a theoretical framework spanning major issues in artificial intelligence and cognitive science. This extends earlier work on motivation with a cognitive model of reasoning that, together with an affective mechanism, enable consistent decision making across a variety of different types of cognitive and reactive processes. We argue that there is no need for emotion in the type of systems we are developing. The affect magnitude concept solves some problems with BDI models, and allows for adaptive decision making over a number of tasks in different domains. The cognitive and affective components are brought together using motivational constructs. We show that the generic cognitive model can adapt to different environments and tasks as it makes use of the motivational models to direct reactive and other situated processes

Keywords

Cognitive architectures, Affect, Motivation

1. Introduction

The control of behaviour is vital to an autonomous system as it responds to its external environment and acts to further its goals. Excepting the most spartan of environments, the potential available information and the associated combinatorics in a perception, cognition and action sequence can tax even the most powerful of symbol processing agents. An agent requires some means of filtering this information. Reactive architectures with behaviors activated by specific sensory signals provide one means of solving this problem. Such architectures are limited to relatively narrow domains and range of tasks. Different reactive architectures are required for differing domains and tasks. Hybrid architectures extend the strengths of the reactive approach with mechanisms that allow a control framework extending beyond the transient states of pure reactive systems. This article considers the control of behaviour in motivated agents. It places recent investigations into affect and affordance within ongoing research into architectures for synthetic intelligence. It makes a case for the linking of perception, cognition and behaviour through affect processes associated with goals. Such an agent does not simply *sense* its environment; it *perceives* it in terms of its own agenda of beliefs, desires, goals and capabilities. The combination of these we term motivations.

The philosophical foundations of cognitive science rest on a number of assumptions (Bickhard & Terveen, 1995). A very important one is that cognition is a natural kind (Fodor 1983, Pylyshyn 1984). It has been suggested that emotion too is natural kind (Charland 1995). There is a growing consensus among theorists and designers of complete intelligent

systems (Minsky 1987, Franklin 2001, Sloman 2001, Minsky 2006) that synthetic minds, to be complete and believable, require a computational equivalent to emotion to complement their behavioural and cognitive capabilities. This need not be a deep model as the thesis behind the work on the OZ project (broad and shallow) demonstrates (Bates et al 1991, Reilly and Bates 1993). This requirement has been highlighted by earlier prominent researchers (Simon 1979, Norman 1980) in their discussions on the nature of cognition in biological systems. In effect to understand how human (and other) minds work, to develop theories about mind and to build computational systems capable of simulating (human and other forms of) mind, both cognitive and affective mechanisms are required. It can be argued that to build autonomous control systems for dynamic environments (whether digital or physical) a synthetic equivalent to cognition and affect is required if these systems are to perform in an adaptable and reliable manner. This claim is particularly apt where the systems are intended for environments associated with noisy and conflicting information. This article makes a case for developing a framework for cognition described in terms of motivation and related control states (for example beliefs, drives, desires, goals, and intentions).

The research described here is part of a long-term project into the study of cognitive architectures. Our research paradigm is that of studying complete agents (Davis 2001). This research is in effect integrative, drawing together a number of threads in cognitive science and artificial intelligence, for example behaviour, decision-making, memory, and learning. Such agents are designed to possess and display a broad range of capabilities. Franklin (1997) describes a complete agent as a system structurally coupled to its environment in

pursuit of its own agenda. Intelligence is therefore an activity related to purposeful behaviour in an environment. This definition of intelligence is not dissimilar to those used by Brooks (1999) and others (Agre & Chapman 1987, Kaelbling 1989) in their criticism of the philosophy underlying much of the abstract intelligence systems studied in AI prior to the behaviour-based revolution. Similar arguments have been made more recently about the potential artificial life offers for the study of cognition (Wheeler 1999). The architectures described in this article accept the grounding issue, for their situated (reactive) aspects, but aim to demonstrate that higher (more *cognitive*) aspects can be generic, with the use of multiple reasoning models, each suited to specific domains and tasks. The project addresses theoretical and design issues associated with cognitive architectures, and attempts to resolve issues through the development of implementations that use a variety of test-beds. The test-beds vary from five-a-side football agents through tileworlds (Hanks et al, 1993) and fungus eaters (Wehrle, 1994) to robotic worlds. The nature of the test-bed reflects the emphasis of the implementation and the theoretical issue(s) under investigation.

2. Emotion, Affect and Theories of Mind

Numerous prominent researchers into intelligent systems have suggested that affect-like mechanisms are necessary for intelligence (Simon 1979; Norman 1980; Minsky 1987) or will arise out of the interaction of the processes necessary for intelligent behaviour (Sloman and Croucher 1987). More recently, Sloman (2001) 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. The question remains how beneficial are emotion or affect systems in artificial systems?

Over the history of psychology, emotion has attracted attention. Hebb (1946) for example could not provide an adequate explanation for observed primate behaviour without the incorporation of emotion. There is no theory of emotion that is consistent across the many competing theory types. Most pointedly with regard to the arguments presented here, it is not clear what level of sophistication is required to experience emotive qualities. So, while the need for emotion in theories of human (primate) mind is not disputed, what emotion actually is and the processes and mechanisms that give rise and support its function is open to debate.

Theories of emotion can be typified as belonging in one of several types, for example physiological (James 1884; Plutchik 1994), evolutionary (Darwin 1892), expression (Ekman 1994), appraisal (Scherer 2001) or goal based (Oatley 1992). This is partially due to different programmatic objectives within, for example, neurophysiology, psychology, philosophy and cognitive science. If a software engineer were to use these theories of emotion as the starting point for a specification of emotion in a synthetic computational system, a number of very obvious comments would be expected. One there is no consistency across these theories. Two, some of the earlier but still prominent theories are

internally inconsistent. Third, most of the theories are so loosely defined that they do not provide for a suitable specification for a computational mind.

Duffy (1962) considers the use of the fuzzy, ambiguous and misleading term “emotion” as fundamentally flawed. Such terms should be abandoned as confusing and new or clearly delineated terms used only where such concepts are clearly and unmistakably identified.

There is such a volume of research in this area 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. However there are many types of artificial system and there are quite legitimate and necessary reasons why a model of emotion (albeit shallow) may be required within these systems (see Sloman et al 2004). The research paradigms of artificial intelligence, cognitive science, computer science and psychology overlap and any purported boundaries are somewhat arbitrary. The question addressed here is not to dispute the importance of emotion for human mind, nor its study in psychology and cognitive science, but to dispute its necessity in the design (and implementation) of intelligent synthetic systems.

Griffiths (2002) suggest that there are different kinds of emotion or emotional process. This is different to the claim that there are basic emotions, for example (Ekman 1994), and more sophisticated emotions that combine the basic emotions with higher level (neocortical) processes. Broadening the scope to include other affective states highlights the diverse nature of these phenomena. There are many potential types (and labels) for the range of affective states. For example my office thesaurus lists twenty-seven synonyms for pleasure (and two antonyms). A trace through the thesaurus following up all antonyms and

synonyms will quickly produce an extensive list of affective terms. It would take the remainder of this paper just to provide linguistic definitions. Highlighting the full extent of the possible relations between them (as in for example a plausible dimension of affect that includes pain, distress, sorrow, torment, grief etc.) is not possible here. These states differ broadly in their situational context, their duration and their possible effects. A complete theory of affect should be able to provide a coherent structure across these issues. It should also provide an account for these in terms of precursors, initiating events, supporting processes, individual and situational differences etc.

There is also the question of what level of control structure sophistication is required for any of these states. It does not make (much or any) sense to discuss how an insect, for example an ant, can grieve over the loss of fellow ants. Why therefore should it make more sense to discuss how a synthetic intelligence, possibly of similar information processing complexity as an ant, can experience affective states qualitatively similar, in type, to grief? It is as yet unclear where it is even sensible to associate the concept of pain with such an organism. The folk psychology of affect is less strict in the application of such terms; for example, a mother may chide her son for “*tormenting*” the ant’s nest. Progress in understanding affect in terms of the information processing complexity of the behavioral control systems of the organism is required if any effort at modeling affective states in synthetic systems is to be something more than silicon folk psychology.

There are many questions that research into the emotions and affect needs to address. Are all of the possible affective states appropriate to computational modeling? If not, which are plausible and why? For example how can a machine experience joy? Wright and colleagues

(1996) used the CogAff architecture as the basis for an account of grief, but they do not imply that their computational designs would be capable of suffering so. Are there categories of affect that are needed if the theory of affect (and hence emotion) is to progress? For example, is joy akin to pleasure, in the same way that grief is akin to pain? Cognitive systems that attempt to model human functioning and cognate theories need to explain how these are alike and the different levels of abstraction over the affective states. Such mind models are qualitatively different to the (insect or at best perhaps pigeon level) systems currently being developed by AI practitioners. Do the decision and arbitration functions and processes required in these latter systems really require the conflict resolution processes to validate their choices in terms of a shallow and sometimes arbitrary use of emotion?

Earlier research into computational emotion and affect (Davis & Lewis 2003, 2004) concluded that for the design of artificial minds, we can distinguish between affect and emotion in terms of magnitude and type. Fishbach et al. (2004) demonstrated in a psychological study that there is a transfer effect between achievement of goals and the means by which the goal was achieved. The transfer mechanism was shown to impact the *magnitude* of affect experienced in regard to the means in question, as well as its *kind* (involving, e.g., promotion-type affect or prevention-type affect). The kind of affect relates to emotive qualities such as anger, joy, etc. Such decision-making based on emotion types has been used in a number of artificial systems, for example (Frijda & Swagerman, 1987; Ortony et al, 1988; Scherer, 1993; Velásquez, 1998). Given that we are not focussed on linguistic capabilities of mind, and for the reasons listed above and in (Davis & Lewis,

2004), this dimension of affect is now omitted. Its inclusion offers no bonus and its exclusion reduces theoretical confusion and complexity. However mechanisms defined in terms of the *magnitude* of affect offer benefit to the design of the perceptual, cognitive and intentional aspects of our architectures.

3. A Multi-Layer Model of Affect

Rolls (1999) describes four broad categories of brain complexity: mechanisms that function in terms of taxes (for example reward and punishment); mechanisms capable of stimulus response learning via taxes; mechanisms capable of stimulus reinforcement association learning and two-factor learning; and finally explicit systems that guide behaviour through syntactic operations on semantically grounded symbols. One approach to cognitive systems is to consider control systems that work in terms of the taxonomy of *Taxes* that Rolls (1999:table10.1) provides. Taxes are primary reinforcers, positive or negative for example reward and punishment, and provide a currency for decision making. The taxonomy is differentiated primarily in terms of the five sensory modalities, reproduction and a collection of diverse reinforcers related to social and environmental interactions. The reinforcers, either positive or negative, can be mapped onto drives and affective states through control mechanisms. In the somatosensory modality for example pain is a negative reinforcer, while touch positive. Control over action is a positive reinforcer. An important feature of Rolls theory of Taxes is an increasing sophistication of affect models from implicit drives to implicit goals to explicit motivations defined in terms of the less sophisticated. There is no wholesale replacement of the simpler mechanisms, but an

expansion in terms of number, type and complexity. We suggest the magnitude of affect provides a viable taxes mechanism by which such mechanisms can function, and demonstrate how this can be utilised in our computational model of affordances.

We are developing a theory of affect that draws on themes such as control states and motivators (Simon 1967; Sloman 1987, 1993; Davis 2001) and affordances (Gibson 1979). We define affect in terms of reinforcers over processes and representational structures. These are qualitatively defined over negative, neutral or positive values, as in the work of Rolls (1999), or numerically over the interval (-1.0,1.0). The reinforcers are used to modify the affordance model associated with goals and the means to achieve goals. Some initial work (Davis 2003) looked to develop a fuzzy (and/or neural) valued processing model that maps across these measures at different levels of the architecture through the affect column shown in Figure 3. This builds on and relates to the valences developed for motivational constructs (Davis 1996, 2003). Hence, affect forms the basis for a control language in an agent architecture. It allows external events and objects to take valenced affordances, and allows the results of internal mechanisms to be prioritised and compared via a consistent model of valenced processes. At the deliberative level, affective values can be associated with processes and control signals to instantiate and modify aspects of motivators and their associated representations. Furthermore, if an agent is to recognize and manage emergent behaviours, whether positive or disruptive, this multi-layer model of affect provides the means for reflective processes to do this. This model of affect addresses the need to integrate reflective, deliberative, reactive and reflexive level agencies in a synergistic manner.

4. Control States and Motivations

The control state approach to cognition (Simon 1967, Sloman 1993) builds on an assumption that cognition is an epistemic process that can be modelled using information processing architectures. Such information processing architectures can be in any number of non-exclusive control states. The processing of information can give rise to changes in the current set of extant control states. 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. Some of these changes are predictable, some deliberative, others reactive. Other changes are emergent and may be unpredictable.

Previous research has attempted to organize motivational control states in an initial taxonomy as a starting point for future research (Davis 2001). Figure 1 provides one taxonomy of five major types of motivational control states with subcategories. The processing requirements and representational qualities associated with these categories vary but become more generic as the taxonomy is ascended. The relationships between these and further motivational states are described below. The term motivation is used with caution and is meant to be inclusive of all internal (and external) control states that focus a cognitive architecture to some specific type of problem solving or behavioural activity. Savage (2003), building on Epstein (1982), suggests a differentiation between instinctive and motivational behavior. Three features (individuation, expectancies and affect) can be used to discriminate between the two, or at least provide indications to the extent that any system is instinctive or motivational. "Individuation", is an observable phenomenon, and relates to a capacity for goals to be realized by one of a variety of actions. This flexibility in

achieving goals is not seen with instinctive behaviour. Instinctive behaviour sees the same limited or identical repertoire of actions across a species in similar circumstances. For example, mayflies only breed in very specific circumstances and there is no variation across the species. “Expectancies” are similar to “affordances” (Gibson, 1979). They reflect different aspects of the cognitive processing associated with realizing goals under varying circumstances. Expectancies resemble high level perceptual qualities or environmental situations interpreted according to the context of current goals. “Affective” qualities are internal qualities associated with motivational behaviour and defined in terms of somatic, neural or psychological activity.

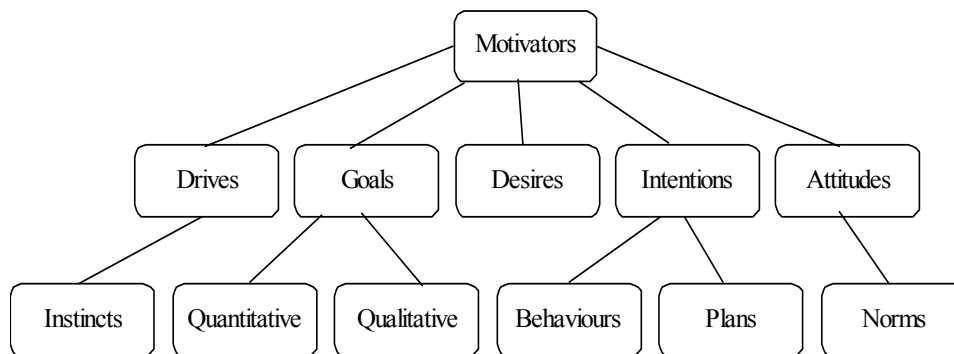


Figure 1. Taxonomy of Motivational States

Drives are low-level, ecological, physiological, and provide the basis for many biological agent’s behaviour in the world. Drives are periodic, short-lived and are defined in terms of resources essential for an agent. Much of the work on motivations in biological agents (McFarland, 1993; Toates, 1998) and artificial life animats (Pfeifer, 1996) is described in terms of drives. In biological agents such drives include thirst, hunger, and reproduction. Instincts, as discussed above, are a very constrained form of drives. The mechanisms

underlying such motivational states are typically pre- or non-conscious. Thresholds for the onset and satiation of such drives are variable and dependent upon processes internal to an agent and external factors arising through the agent's interaction with its environment.

Drives can be modeled relatively easily in computational agents using intrinsic control structures (for example quantitative goals – see (Konidaris & Barto 2006)), or in

architectures capable of supporting adaptive state automata or fuzzy logic mechanisms.

Drives do not necessarily require deliberative mechanisms, but can instantiate further motivational states (such as *Desires* or *Goals*) that do. For example a Drive to maintain an agent's energy level may give rise to multiple Desires: to maintain its energy level; to seek energy sources; and to collect energy sources. The addition of other Drives (for tasks and targets) requires some form of arbitration mechanism to decide on a best course of action.

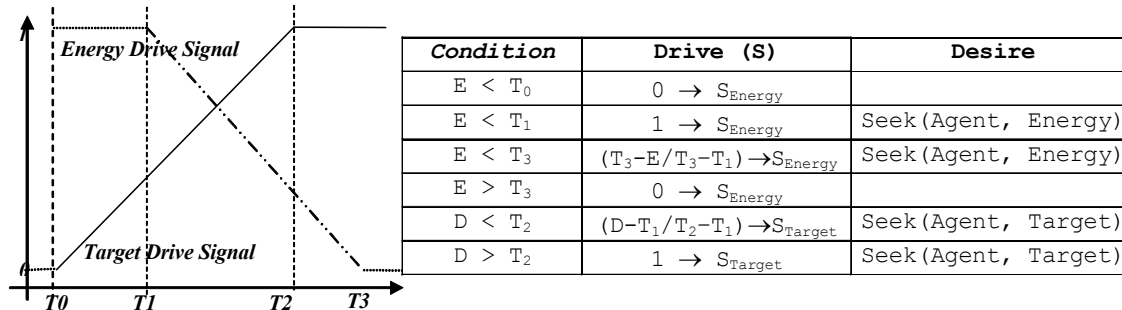


Figure 2. Model for Energy and Target Drives (with Desire Equivalents)

Quantitative goals are the types of goals discussed in engineering control theory (Sontag, 1998) and reinforcement learning, for example (Maes, 1989; Toates, 1998). They can encompass drives (for example the Target Drive Signal in figure 2). The differentiation from drives allows for more flexible representations and process models. Iglesias and Luengo (2005), uses mechanisms based on artificial life principles for goal generation in

intelligent virtual characters. Mathematical models are used for a number of Drives that underlie a heterogeneous set of internal states, for example energy, sociability, boredom and anxiety. Anxiety is defined in terms of a measure of the frustration caused by trying to perform a goal and failing in the attempt, as given by the formula $A(t,k) = 100t/a_k$ (for goal k at time t with time a_k to achieve the goal). These internal states are used (as a feasibility metric) together with an affective quality (wish rate), again defined mathematically to generate a goal priority, which is used for goal selection. Once selected goal metrics are subsequently modified to reflect their success. While not attempting to fit with any theoretical model of goal selection, it demonstrates that applications can benefit from the use of affect models to prioritise activity. Konidaris and Barto (2006) define a number of critical properties for (robot) motivational models. These include metrics for goal priority, action selection, reward and feedback. They build upon (Spier 1997), and demonstrate that drive-based metrics used in choosing between reactive behaviours, should be comparable to those used in providing adaptive behaviour. Here we extend this perspective to be inclusive of other categories of control state.

Qualitative goals are seen in many artificial intelligence models of problem-solving. For example, Newell and Simon (1972) describe problem solving as a search through problem space. Problem space can be described in terms of states and paths between states, where the solution state is defined as the goal state. Qualitative goals are the types of motivators discussed in most planning literature (Nilsson 1998), where an agent or robot must act to achieve or maintain some state of the environment. In many respects these are analogous to the end states associated with Desires. While processes involving the manipulation of

knowledge are typically referred to as symbolic, they need not necessarily be so - see (Kosslyn 1975, Freeman & Skarda 1990, Brooks 1991). In the model being developed here problem-solving is a component of goal-directed action or control. Where there is an obvious path for the transformation of an initial state into the goal state, these state space operators can be modeled as reflexive or reactive behaviours. The conditions for the applicability of these components can be packaged as qualitative goals whose plan is to use specific behaviours. Where there is no obvious path or method of achieving a solution, a series of partial (sub-) goals may be tried and reasoned about in terms of qualitative goals over sub-symbolic behaviour, or more orthodox planning.

Desires can be viewed as (deliberative symbolic) statements defining preferred agent and environment states. Task-directed behaviour in embodied reactive agents define intentions to transform current agent and environmental states in an implicit way. In either case, excepting the representational issue, this is analogous to their use in BDI architectures (Georgeff & Lansky 1987), for example *desire(agent, stateY)*. *Desires* can be underpinned by Drives and in turn underpin Goals and other purposeful behavior, for example *Intentions*. In BDI architectures (Bratman, 1987) intentions are adopted strategies for achieving desires. As such they can involve the adoption of specific plans and/or behaviours to achieve the intended desire. Thangarajah et al. (2002) detail a number of formal models (and rules) that enable a logically consistent set of Goals to be maintained given a non-consistent set of Desires. In an effort to overcome known problems with the formal models used in specifying goal management in BDI agents, they invent preference, priority and goal ordering operators, but fail to specify the basis for the metrics used in

these operators. CRIBB (Wahl & Spada 2000) provides a different perspective on BDI, representing the reasoning capabilities of a five year child (Bartsch & Wellman 1989). The CRIBB model has been extended with an affect and affordance model (Lewis 2004) that maps onto attributes of a motivational structure in the schemas described below.

The remaining category identified here, attitudes, are pre-dispositions to respond to specific sensory or cognitive cues in specific ways. In experiments with Five-a-side football (Bourgne 2003), agents adopt different attitudes (for example, attack, defend) in response to team orders or game situations; the change in the attitude requiring a change to the motivational model then in force. The motivational model affects perception and the criteria by which goals are selected. The work on norms in computational agents (Epstein 2001, Staller & Petta 2001) is relevant to this category. Norms provide for preference operators, typically based on social and interactive models, that enable goals to be prioritised. A change in the currently invoked Norms can invoke changes in the active preference operators within domain and tasks; or require different task-models to be adopted. The former affects how reinforcers are used in goal, association and motivator feedback. The latter requires a redefinition of the current goal (and association) model being used; i.e. different goals and potential means of achieving goals. Experimentation across different task domains and task definitions within domains has led the development of an architecture with mechanisms to allow dynamic and adaptive decision models. Again the affect model provides the framework for these (meta-)cognitive adaptations.

Anything other than the simplest of architectures should be able to support multiple instantiations of a combination of the motivational states given above. Different approaches

to building synthetic agents will favour alternative solutions. Reactive approaches may favour implicit drive and goals through the integration of task related behaviours, for example (Agre & Chapman, 1987). The problem remains how to prioritise between behaviours given the possibility of multiple actions, i.e. arbitration mechanisms. Behavior weighting, inhibition and subsumption (Brookes, 1998) are well used. There however remains the problem of allowing comparisons between transient and more long-term states. The use of pure deliberative mechanisms will coerce decisions to be made between reactive drives and more abstract goals in terms of conflict resolution mechanisms perhaps inappropriate for drives.

The use of a multiple level control system allows for the management of this otherwise disparate set of motivational states. It allows Drives best managed at the reactive level, to be dealt with at that level, while the management of goals and desires occurs at the deliberative level. Attitudes and norms are managed at the deliberative and meta-management (reflective) levels. This raises the problem of possible conflicts between the actions required of these layers, requiring some means of activity prioritisation that is consistent across a hybrid architecture. Even in non-complex environments, situations where several behaviours are in conflict will occur frequently. Heuristic models can provide an evaluation function to resolve this. The experimental and theoretical research presented here allows activity prioritisation to be made in terms of affect magnitude, and its interpretation in different contexts. The affect model can be used to compare drives, desires, goals and the means to achieve them (i.e. intentions). It allows decision making across different types of processes to remain consistent, and allows feedback and adaptation

where required. The adopted affect models provide dynamic evaluation functions that demonstrate a processing consistency across mechanism such as perception, goal selection and the different processing paradigms inherent in hybrid architectures. The schema by which this is coordinated we term motivations.

5. Representing and Controlling Motivations

Motivations are more encompassing than drives, goals and indeed plans and behaviours for accomplishing goals or satisfying drives. Motivations include not only the descriptions of goals and their actor and environmental contexts (for example basis for actions, belief states etc.), the means to achieve specific goals (whether through plans or behaviours) but define an affective context. In the computational systems described in the following section, activity and behaviour at one level is represented and directed within other layers. For example a specific reactive sub-architecture with a particular sensor set, can be associated with specific belief, desire combinations with a given affordance. This motivator, once adopted, causes the reactive sub-architecture to run with no further deliberative processing. The reactive processes can invoked with parameters defining success and fail conditions (Find ball in ten cycles, do not hit any object, etc.) The motivational processes and associated deliberative reasoning (for example belief revision) can be subsequently invoked once any of these are met. The primary conjecture is that the design and implementation of such architectures can proceed using a systematic control language to direct the processing within an intelligent system. This control language is grounded in affect, and aims to be

consistent across different domains, tasks and levels of processing. This enables the architecture to adapt, learn and create new decision models.

The computational work is being developed with no explicit requirement for emotion type but rather a reliance on magnitude of affect (a valencing of and within internal processes), affordance (a valencing of the environment and entities within that environment) and motivational constructs to guide both internal and external acts. The generic representational schema makes use of fifteen components that reflect the nature of the most expansive of motivational control states (Davis 2001). In many instances, for example behaviours related to drives, many of these components are unused and the motivators can be manipulated by mechanisms analogous to the reactive planners of Kaelbling (1989). Where required more extensive (and computationally expensive) deliberative processes are used. An instance of this is the motivator merging, given in (Davis 2003), which makes use of mechanisms analogous to those used in teleological planning (Nilsson 1994).

Valence	Aspect	Process and Dimension Category	Affect Magnitude
Belief Indicator	Motivator	Truth values for Semantic Content and Motivator Attitude Perception » Deduction » Assumption	[0, 1]
Commitment	Motivator	Motivator Acceptance (ignored to first priority)	[0, 1]
Dynamic State	Motivator	Motivator Process (instantiated to complete)	[0, 1]
Importance	Goal	Goal Importance (low to high)	[0, 1]
Insistence	Association	BDI Association Strength (low to high)	[0, 1]
Intensity	Motivator	Motivator Strength (low to high)	[0, 1]
Urgency	Motivator	Urgency (low to high) or time cost function	[0, 1]
Decay	Motivator	Motivator Decay (low to high) or time cost function	[0, 1]
Reinforcer	Affect	Goal and Association Feedback (negative to positive)	[-1, 1]

Table 1. Valences affecting motivational constructs.

The affective valencing of processes and representational structures can be given, as a domain or task model, or the agent can adapt or learn appropriate affordances according to

its role and current environment. 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 a motivational 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 mapping between its internal and external environments. This influences the different categories of cognitive and animated behaviour. To paraphrase Mearleu-Ponty (1942), an agent is driven to learn, adapt and act in its environment by disequilibria between the self and the world. . The current task model may be adapted, with changes in motivational affordances as a result of behaviour-environment interaction. Alternatively the task-model may be extended with new belief-desire-intention associations, or an alternative task model used. The valences used in the current motivational structure (tables 1 and 2) provide the means to characterise the disequilibria. The multi-dimensional measures associated with the motivational construct, in effect, provide the fitness function for easing any such disequilibria. The problem remains how to generate these values and decide across the current stack of motivators in a manner that does not rely on ad hoc control heuristics.

Goals, associations and motivators need to be valenced in a number of different ways, for example intensity, urgency, insistence. Goals can be preventive (ensuring that certain environmental or agent states do not arise) or assertive (ensuring that the environment is changed in some way). From the total set of goals available to the agent, a subset of those not currently achieved (i.e. they do not match to the current belief set), are selected as potential sources of action. Goals that have recently failed can be retracted from this subset

if the appropriate goal selection strategy is in use. The remaining goals are then matched to currently valid associations (i.e. associations with their Basis to be found within the current Belief set). This Belief, Goal association combination forms the basis of a motivator.

Where no goal is currently valid, the currently failed goals and the belief set are examined to determine if a more weakly supported goal is viable. For example, goals rejected because the belief that negates their selection is based on an assumption, rather than deduction or perception. Alternatively the association set and past (failed) motivators are examined to determine if an alternative means of achieving a failed goal is possible. This again enables more weakly supported goal selection. Where the selection of a goal is not possible (no matter how weakly supported), the architecture uses its domain-defined default (reflexive) behaviour (for example, do nothing, explore the environment for objects, avoid collisions etc.). This behavior persists until it provides feedback to initiate a new Belief-Desire-Intention cycle. Alternatively if the meta-cognitive processes recognise a failing decision model, alternative task models can be used.

Table 2 gives examples of goals and associations defined using this model for two domains.

Goals take the form `goal (Desire, Success, Importance)` where both `Desire` and `Success` are defined in terms of predicates that may be found within the belief model.

Associations take the form `association (Belief, Desire, Intention, Insistence)` where `Basis` and `Desire` are again defined in terms of predicates that may be found within the belief model. `Importance` and `Insistence` are defined in terms of the affect model.

`Intention` is defined in terms of plans, activity or behavior (for example a reactive sub-architecture specification) that may be suitable for achieving the `Desire` given the `Belief`

basis. For some domain, task combinations we have used prior experimentation to generate success data for the association model. This data is accessed to build a competence model at run-time. Where no prior experimentation has occurred, a completely new (and *Desire-Intention* exhaustive) set of associations is generated.

Domain	Aspect	Definition
Robot Ball	Goal	goal(avoid(collisions), not(hit(Object)), low)
Robot Ball	Goal	goal(find(ball), found(ball), high)
Robot Ball	Goal	goal(find(robot), found(robot), medium)
Robot Ball	Association	association(environment(sparse), find(ball), reactive(true, true, method1), medium)
Robot Ball	Association	association(environment(dynamic), find(robot), reactive(true, true, method3), high)
Robot Ball	Association	association(environment(sparse), avoid(collisions), reactive(true, false, method2), high)
Robot Ball	Association	association(environment(dynamic), avoid(collisions), reactive(true, false, method2), low)
TileWorld	Belief	belief(environment(sparse), deduction, 50)
TileWorld	Belief	belief(near(tile), perception, 55)
TileWorld	Goal	goal(avoid(collisions), not(hit(Object)), high)
TileWorld	Goal	goal(pick(tile), hold(tile), high)
TileWorld	Association	association(environment(sparse), avoid(collisions), reactive(true, false, method2), high)
TileWorld	Association	association(near(tile), pick(tile), reactive(false, true, methodpick), high)

Table 2. Beliefs, Goals and Associations from Two Domains used in Experimentation.

Motivators combine goals and associations, with perceptual object references, parameters for running behaviours, and are compared in terms of their *Intensity*. The *Intensity* of the motivator is given by $Insistence(Association) * Importance(Goal) * TimeStamp(Belief) / Current\ Time$. As *Time* is represented as an increasing integer, this equation favours associations with their belief basis founded on more recent events, and Goals related to recent changes in the environment. Alternative decision models are possible. Motivators are ordered according to this metric, with (in the default decision

model) the highest valued motivator selected. Hence for the Tileworld situation given in table 2, a motivator to pick a tile using `methodpick` would be chosen, with its `Intensity` value reflecting the belief basis as more recent for this goal association pairing. Where multiple motivators with identical `Intensity` are generated, the one with the highest goal importance, then highest association insistence provides a conflict resolution mechanism. If the belief basis for the Robot Ball domain associations were all true with an identical timestamp, and goals valid, the motivator to achieve the `find(ball)` goal with the `method1` association would be selected over the `find(robot)` goal with the `method3` association combination (which has an equal motivator intensity), as it has the highest goal importance. Where no means to achieve a goal is available or left untried, the competence model is extended with novel Belief-Desire-Intention combinations (with a default mid-range `Insistence` value). In all cases `Insistence` values are modified by `Reinforcers` as associations are used within motivators. `Reinforcers` are in effect numeric values generated via feedback from actions (as indicated in Figure 4). If an association succeeds, a `Reinforcer` increases its `Insistence`; failure causes a negative `Reinforcer` to be generated. This allows the architecture to adapt and modify the means by which it can achieve goals. Goal `Importance` does not normally change at run-time, unless the agent's norm model changes. As experimentation continues, and more unpredictable, novel environments are explored, this aspect of goal management will be tied in with explorations into reflective (meta-cognitive) cues for learning about norms. The current architecture modifies its current task model and adopt alternative decision making strategies at the deliberative level, with the research into meta-cognitive processes yet to be integrated.

6. Computational Architectures for Motivation and Affect

Current work on architectures for motivated agents is based on experiments in the theory, design and implementation of affect and motivation based architectures (Davis 1996, 2001). It builds on the ecological perspectives offered by Gibson (1979), and on the work of Simon's control state theory. Preliminary work (Davis 1996) centered on motivators and goals, how they come into being and how they are managed. The architectures developed in this work, and related research into a multi-level representational framework for emotion (Davis 2002). Those works made use of variations of the three column, three level architecture developed with the Cognition and Affect project (Sloman 1993, Davis 1996, Wright et al 1996, Sloman et al 2004). This led to work on agents and control states (Davis 2001), again focused on goal processing, and the adoption of a psychology based model of reasoning in children, CRIBB (Wahl & Spada 2000), in the work with Lewis (2004). Current work develops these threads further.

There is longer an unequivocally commitment to variations of a three-column, three layer architecture (Gat 1998). The research requires other frameworks to be used for a number of reasons. Figure 3, for example, shows a four tier, five column design that provides an abstract template for all our implementations. We assume that modularization, for certain cognitive processes, is feasible. The approach taken is one that merges the principles of architectural parsimony (Hayes-Roth 1993) and the conceptual ease through architectural expansion of Singh and Minsky (2003). Architectural parsimony requires that we include no layers, components or mechanisms unnecessarily, while the principle of conceptual ease allows the separation of processes and mechanism as further columns or layers if this

enables clarification of the principles under investigation or ease of design and/or implementation. In one satellite project, the reflective processes are separated into multiple layers for learning, meta-control and metacognition as this eases progress.

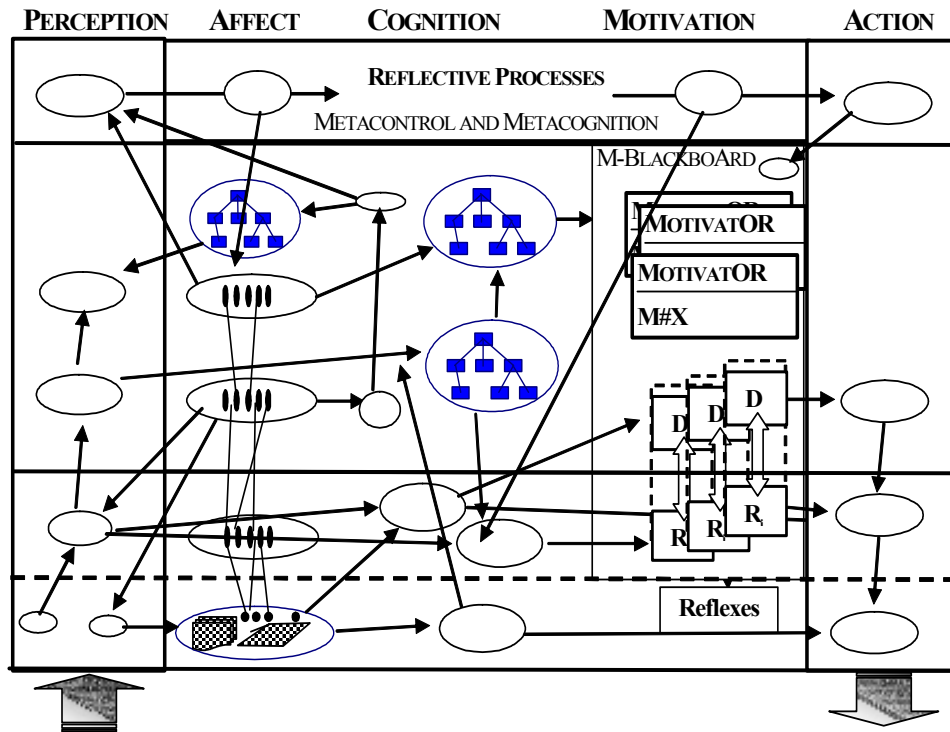


Figure 3. The Four Tier-Five Column Architecture

CogAff and most of the three layer architectures differentiate between the three towers of perception, cognition (or models) and action across all layers. This provides a useful starting point to consider the contexts for processing across an architecture. However, as Sloman and colleagues (2004) point out, this can lead to an inherent bias to simplistic left-to-right processing models based on *Sense-Think-Act* cycles. Such cognitive models may suffice for simple agents with limited tasks but hinder more sophisticated information

processing that include feedback, directed perception and sophisticated behaviour-environment interaction.

In the most recent design, the four represent reflex, reactive, deliberative and reflective processes. The five columns enable the separation (and modularization) of perception, affect and affordance processes, generic cognitive processes (planning, behaviour selection etc), a motivational blackboard and output to actions. Only in the most mundane of environments is the *Sense-Think-Act* cycle not interrupted. The theory requires, the design indicates and the implementations allow information flows as indicated by the arrows in figure 3. Singh and Minsky (2003) present a realm matrix approach that discriminates between, for example, the Bodily Realm, Physical Realm, Sensory realm, Psychological realm, Social Realm etc. Mappings across these realms provide for simple and complex models of information processing. For example mappings across Sensory to Bodily to Physical Realms provides for a simple *Sense-Think-Act* cycle for a mobile robot. More sophisticated processing follows. In this abstract model, there exist reflexes and reactive behaviours that allow a direct response to sensory events. These can provoke processes or being modified at a more abstract level. Other automatic processes *necessitate* the generation of deliberative control states to achieve their goals. The reflexive and reactive processes are domain and task-defined. In any one domain (for example, five-a-side soccer agents), the currently loaded reflex, and available set of reactive mechanism changes slightly as the agent adopts different roles. As the domain changes, and hence the tasks change more radically, a wholly different set of reflexes and reactive capabilities may be required. 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. Earlier experimentation (Davis & Lewis, 2003, 2004) made use of an architecture (ACRIBB) based on cognitive models of reasoning in children (Bartsch & Wellman, 1989; Wahl & Spada, 2000). The results from ACRIBB show that an affect model, incorporating perceptual affordances and affect magnitude, improved both life expectancy and task efficiency in stressed environments. This is now incorporated into the architecture described here, with the ACRIBB BDI reasoning model providing a default information flow across the columns of the architecture at the deliberative level that integrates perception, goal processing and action selection. The association of current states (for example beliefs) task descriptors (for example, goals) and the means to achieve them (intentions in the form of plans, behaviours and reactive sub-architectures) can be given, modified or learnt. The decision models in place at the deliberative layer, together with the reactive and reflex behaviours (and sub-architectures) are situated. The reasoning models and motivational framework are generic and domain independent. The reflective processes serve to monitor cognitive behaviour or control it in some other way. In one current project, the reflective layer provides the means by which the agent initiates learning activities, through the recognition of extreme affective states associated with failures to map internal models of action and behaviour to the current tasks and domain. More extreme affective states (symptomatically categorised as a loss of control or perturbation), for example persistent attempts to establish goals with actions that continually fail, are effectively clamped by means of self-regulatory processes within the architecture; again based on a theory of affect.

From a theoretical, design and implementation stance, high level and low level processes coexist and interact in a holistic manner through the use of motivation and affect. In effect, goal processing, planning, decision making and other cognitive processes are not purely abstract but exist in relation to other automatic, affective, motivational and situated processes. They are, in effect, embodied within the context of their interactions with their underlying (situated) processes and the agent's relationship(s) with its environment. In our current work the general deliberative processes are not changed for differing testbeds. The more generalized levels of the architecture interface to task-specific behaviours and reactive sub-architectures through the use of domain specific belief, goal and association models. The dynamic processes associated with these models, such as belief revision, goal management and action selection are generic. In effect, we are developing the cognitive architecture with a range of competencies (domain skills, reactive sub-architectures) and tasks (modelled as desires and goals). In some experimentation, the implementations are purposefully given the wrong task model. Where task domains are not too dissimilar, for example where there is some overlap of objects between domains, the architecture quickly adapts to achieving those goals that are possible with whatever current capabilities (i.e. reactive sub-architectures) are available, irrespective of how insignificant the goals were for their intended domain. The architecture is currently being developed with meta-cognitive capabilities enabling autonomous domain models switches on recognition of goal (desire or intention) perturbation.

The current framework extends earlier work with an architectural global blackboard for the motivational constructs. Earlier research (Davis 1996, 2001) did not separate motivational

processes from other cognitive functions. This sub-architecture allows beliefs, goals and intentions to be linked using the motivator structure detailed above. The blackboard is managed by motivational control processes, allows the activation of deliberative and reactive processes with the intention of achieving currently selected goals, and provides for feedback via a coupling of deliberative and reactive sub-architectures to achieve goals (Figure 2). Feedback to the blackboard can initiate motivator assessment, belief revision, goal management and association updating. The process history on the blackboard can ultimately trigger deliberative and reflective processes responsible for domain model management. The use of this blackboard again demonstrates the cognitive architecture relies on more than the default *Sense-Think-Act* cycle.

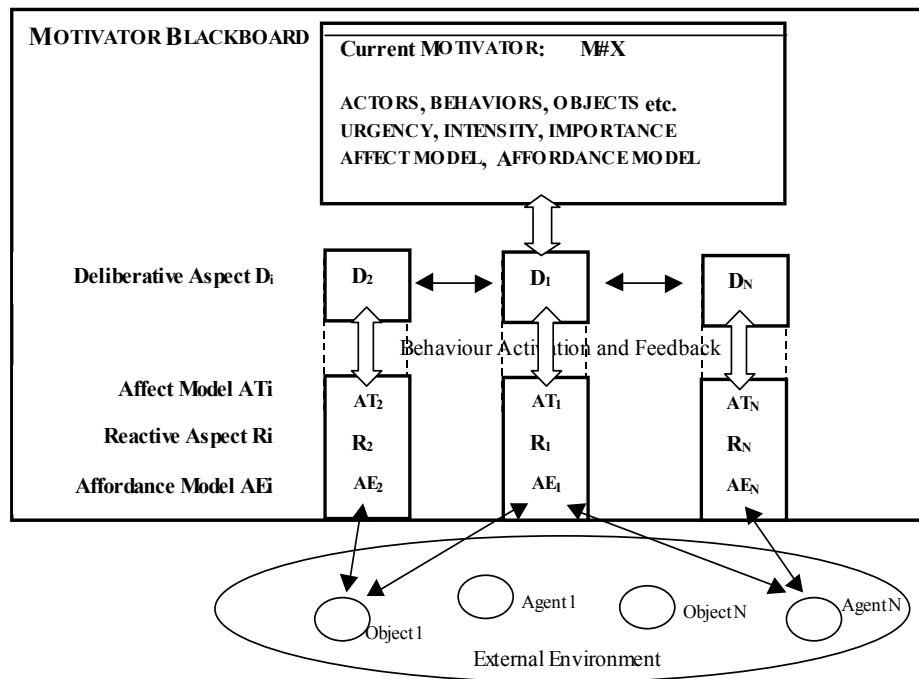


Figure 4. Motivator Blackboard Enabling the Linking of Reactive (R_i) and Deliberative (D_i) Processes through Affect (AT), and the mapping of Affordances (AE) onto Actors and Objects in an external environment

Langley (2006) identifies three architectural paradigms used for developing cognitive and intelligent systems. One, the multi-agent system approach, was characterised as flawed for cognitive systems due to its requirement for inter-agent communication. The blackboard approach resolved this problem as communication is bypassed in favour of interactions with data and knowledge on the blackboard. However Langley states that the blackboard approach is inappropriate for cognitive architectures as it fails to meet a number of requirements stated by Newell (1990) in his unified theories of cognition. In particular he makes the assumption that a blackboard system cannot meet requirements for module independence and the use of representational, process and learning schemas for memories, beliefs, goals and knowledge. We argue that the blackboard paradigm can provide a useful and viable approach for cognitive architectures that does not break constraints imposed by cognitive theory, and offers a diversity of control and representation paradigms (Englemore and Morgan 1988). The blackboard approach serves as a useful computational model for the design and implementation of cognitive models based on the global workspace perspective (Baars 1997, Franklin 2001).

This extended motivational blackboard provides context for the dynamics of ongoing processing. Both architecturally generic and motivational construct specific processes can access the blackboards and in turn be influenced by their content and processes.

Experimental work with Five-Aside football scenarios (Nunes 2001, Bourgne 2003) revisited the representational structure and processes associated with motivators (Davis 1996, 2001). This laid the foundation for the use of affect and affordances to valence the motivational constructs. The association of perception, behaviour and abstract

representations about plans of actions and the relevance of actions and entities in the environment with agent internal worlds is now defined and compared in terms of common criteria. Affect magnitude (and affordance) become the means by which the agent architecture can weigh and control the economics of its processing. It provides a means whereby attention can be directed to the most relevant and/or pressing aspects of the interactions of the agent with the environment, its needs and its goals. Related work (Davis and Lewis 2003, 2004) suggests that adding a simple model of affect to cognitive desire and intention models such as CRIBB (Bartsch and Wellman 1989), result in more effective processing and task management in resource competitive environments. This now extended affect mechanism in hand with the blackboard scheme for motivation provide a sophisticated interaction between low level (reflexive), reactive, deliberative and meta-level (reflective) processes. More recent work with robotic and simulation test-beds has led to a strengthening of the reasoning mechanisms, with an improved version of the CRIBB BDI model. In this model perceptions not only update the Belief model but are also used as feedback to evaluate motivational states, and allow the (adaptable) affect model to be used for prioritising goals and belief-desire-intention associations. The current implementations work well with a variety of Intention models (as represented in Figure 3 by the action column, or through the direct coupling indicated in Figure 4) including reflexive, task-tuned reactive and planning sub-architectures in a number of domains.

7. Discussion

As the designer of artificial intelligent systems one could ask what is the biological analogue to the information processing complexity of the system being designed and developed? If it is insect, does it need affect or emotion and would not some other control criteria be more appropriate? In short is there a need for emotion in the system. The developer of multi-media interfaces (Kort et al. 2002) may require some form of affective processing in generating hypotheses about the emotive state of the user sat outside the video camera within the interface. But do all artificial systems require an emotion system for intelligent behaviour?

The reason these questions are raised is the ongoing efforts of cognitive scientists across many disciplines, philosophers, psychologists, computer scientists, neurophysiologists etc., to move the theory of affect (and emotions) forward. The roots of current theories reside in folk psychology and historical theories of affect. Are the efforts of the practitioners in giving their artificial systems emotion helping this progress? The AI and Cognitive Science literature is littered with overlapping and disjunct definitions, with the same term being used as a referent to many kinds of disparate activity. It is suggested that the theory of synthetic intelligent systems can progress without the need for emotion per se but with a requirement for 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 neurophysiology, 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 emotion types and categories. It is suggested here that in order to make more substantial progress, efforts are required to provide the means by which we can categorise the types of information processing systems in existence and being developed, whether natural or synthetic. A means of providing the discriminatory criteria necessary to perform such conceptual analysis, built from the substantial work of Sloman and others (2004) has been given here.

In negating the use of emotion some alternative is required, not just simply to mirror the fact that natural minds use emotion but because some form of motivational control language is required to do anything associated with mind. Consider activities such as sensory attention, behaviour selection, goal maintenance and the learning of new skills. There needs to be some metric or fitness function, whether explicit or implicit, to provide a means of conflict resolution. For example given two contrasting percepts, both of which are equally viable for an agent to act on, but which require mutually exclusive processing, how does the agent determine which to attend? Without the appropriate criteria to choose between two equally plausible activities, the agent in effect will have to choose at random. Many artificial systems in the past have used ad hoc control heuristics to solve prioritization of activity or heuristically defined domain parameters. Here we suggest that at a theoretical, design, architectural and implementation level a consistent valencing and

control language across processes and mechanisms, described in terms of affect magnitude, may offer more in the pursuit of synthetic intelligent systems.

Merleau-Ponty (1942) considered that humans are moved to action by disequilibria between self and the world. If a descriptive model can be provided for these disequilibria then it may be possible to use it as a design framework for synthetic minds. The phenomenological approach to the analysis of human behaviour implies a distinction between an agent's internal and external environments. For any specific agent, there may be no differentiation in the extent to which either of these is real. One implication is that reasoning and behaviour is activated by a need expressed in terms of descriptors for drives, concerns and goals. Excepting the most spartan of environments, the potential available information and the associated combinatorics in a perception, cognition and action sequence can tax even the most powerful of symbol processing agents. An agent requires some means of filtering this information. Reactive architectures with behaviors activated by specific sensory signals provide one means of solving this problem. Such architectures are limited to quite narrow domains and range of tasks. Combining multiple reactive architectures to provide more general purpose capabilities is possible. Such a system requires the means to decide between the available reactive architectures. Here a general purpose cognitive motivational model is used to resolve this problem.

Consider a highly modular architecture for a synthetic mind. Within this framework exist many vertical and horizontal modules, some highly specialized and responsible for specific activities and processing, some generic, some very localised and others more global. It can be argued that that there is a requirement for global mechanisms to provide context and

allow the integration of modules. The global (contextual) mechanisms may be based on ideas such as computational chemistry (Adamatzky et al, 2005), global workspaces (Baars 1997, Franklin 2001), blackboards (Hayes-Roth 1993) or some combination. Should and how can the control architecture make consistent decisions across these different modules and mechanisms? We have used multiple-level representation based on the idea of affect magnitude and taxes. For example in integrating behaviours (whether innate, adapted or acquired) into a skill sequence for a particular context, affective dissonance, a negative feedback, provides a fitness function to be minimized. At the individual module level, a fitness function mapping input to output (for example as an affordance and accordance over the requisite sensori-motor mapping) is required. At a more abstract level, multiple valued schema as local blackboards, for reasoning about motivations, goals and other forms of control states, are used. Again varieties of affect magnitude provide a consistent valencing mechanism across control states, behaviours and architecture levels.

From a broader perspective cognition can be viewed as problem solving and behaviour selection. Problems can be categorised (Greeno 1978, Eysenck, 1990) in a number of ways. Routine problems, with well defined paths from given to goal state, can be solved using reactive or reflex behaviours. Well described problems, where a clear representation is possible, can be solved using a combination of behaviours in sequence, but requiring deliberation to determine appropriate behaviour sequences. Arrangement problems (Hayes-Roth et al, 1979), requiring the arrangement of mainly pre-existing problems and partial solutions, require further adaptation in the cognitive models available. Representational forms capable of supporting these problem solving capabilities can be symbol or signal

based. Signal based forms may be reactive and/or connectionist in nature. Learning enables new solutions to existing problems and the categorisation of new problems as an exemplar of (an) existing problem(s). The perspective presented here is that problems can be defined in terms of beliefs and desires. Solutions to problems can be defined in terms of goals and intentions. Motivations provide the representational structure to link problems and possible solutions.

8. Conclusion

This paper has presented an architectural approach to cognition as the linking of perception and action with a theoretical basis grounded in the work on affect and control states. This research is a synthesis of concepts based on an analysis and investigation of how different perspectives on autonomy, affect and motivation map onto computational frameworks. It has offered insights into the nature of homogeneous mechanisms across different processing, representational and knowledge levels, and their explanatory role in describing mental phenomena. The described architecture and its implementations demonstrate how artificial life and emergent behaviours can be combined with the more abstract decision making processes associated with cognitive agents. This work is a step towards a better theory of cognition and affect

There are controversies about the terminology used in this research. Motivation, drive, goal and affect are used to refer to and mean a number of different things. There is no universal definition of these terms across (or even within) the fields of philosophy, psychology, cognitive science and artificial intelligence. The framework presented here attempts to

harmonise these differences in a coherent manner across transient, mediating and permanent control states. Affective events and environmental behaviors are examples of transient states. Goals, plans, motivations and reactive subsystems are examples of mediating control states. Long term dispositions (norms), domain and task models are examples of more permanent cognitive states.

We confronted the now widely held requirement for emotion in intelligent systems on a number of grounds. The starting thesis is that overall the theory of emotion is currently too disorganised to be of much use in the design of synthetic intelligence, and that many computational models of emotion are ad hoc, not furthering the purposeful investigation of the subject. More pointedly, emotion type is not really a requirement for many forms of synthetic intelligence, and that the more straightforward concept of affect magnitude with taxes or reinforcers, can be used to enable effective decision-making. Our model draws on the work of Rolls (1999), and others, studying motivational behaviour in biological and synthetic agents. The affect model provides the basis for adaptation, learning and control across qualitatively different classes of processes in a cognitive architecture. Affect and affordance become the means by which a sophisticated agent architecture can weigh its processes and control the economics of its processing. It provides a means whereby attention can be directed, resulting in more effective processing and task management. It also provides a consistent solution to some of the problems associated with Belief-Desire-Intention reasoning models (for example, preference operators). The competence model enables the cognitive architecture to focus attention on important objects, goals and the means by which goals can be achieved. We have explored naïve learning and adaptation over the affect model

guided by motivations, with affect magnitude as a useful “fitness function”. The investigation of deeper learning capabilities, for example developing perceptual symbol symbols using connectionist ideas (Barsalou, 1999) will test this idea more fully and extend the link between motivation, perception and action. Continuing research, in robotic and simulated environments, will determine how complex are the states arising from the adoption of this model, and demonstrate that cognitive architectures are both generic and situated.

References

- Adamatzky, A., De Lacy Costello, B and Asai, T. *Reaction-Diffusion Computers*, Elsevier, 2005.
- Agre, P. and D. Chapman. PENGI: An implementation of a theory of activity. *Proceedings of the Sixth National Conference on Artificial Intelligence (AAAI-87)*, 268-272, Seattle, WA, 1987
- Baars, B. J., *In the Theater of Consciousness*. Oxford: Oxford University Press, 1997.
- Bartsch, K. and Wellman, H. Young children’s attribution of action to beliefs and desires. *Child Development*, 60, 946-964, 1989.
- Barsalou, L. W., Perceptual symbol systems. *Behavioral and Brain Sciences* 22:577–609, 1999.
- Bates, J., Loyall, A.B. and Reilly, W.S. Broad agents, *SIGART BULLETIN*, Vol. 2, No. 4, 1991.
- Bickhard, M.H. and Terveen, L, *Foundational Issues in Artificial Intelligence and Cognitive Science*, Elsevier, 1995.
- Bourgne, G., *Affect-based Multi-Agent Architecture (for a 5-aside football simulation)*, Thesis, Computer Science, University of Hull, 2003.
- Bratman, M. E., *Intentions, Plans, and Practical Reason*. Harvard University Press, Cambridge, MA, 1987.

- Brooks , R. A., *Cambrian Intelligence: The Early History of the New AI*. Cambridge, Mass. MIT Press, 1999.
- Charland, L.C. Emotion as a natural kind: Towards a computational foundation for emotion theory, *Philosophical Psychology*, Vol. 8, No. 1, 59-85, 1995.
- Darwin, C. *The Expression of Emotions in Man and Animals*. London. Murray, 1872.
- Davis, D.N., Reactive and motivational agents, in *Intelligent Agents III*, J.P. Muller, M.J. Wooldridge and N.R. Jennings, Eds., Springer-Verlag, 1996.
- Davis, D.N., Control states and complete agent architectures, *Computational Intelligence*, 17(4):621-650, 2001.
- Davis, D.N. Computational architectures for intelligence and motivation, *IEEE Systems and Intelligent Control*, Vancouver, Canada, 2002.
- Davis, D.N. Architectures for cognitive and artificial-life agents, in *Intelligent Agent Software Engineering*, V. Plekhanova (eds.), IDEA Group Publishing, 2003
- 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), 2003.
- Davis, D.N. and Lewis, S.C. Affect and affordance: architectures without emotion. *AAAI Spring Symposium*, Stanford University, California, USA, 2004.
- Duffy, E., *Activation and Behavior*, Wiley & Sons, 1962
- Ekman, P. *The Nature of Emotion: Fundamental Questions*. Oxford University Press, 1994.
- Englemore, R. and Morgan, T., *Blackboard Systems*, Addison-Wesley, 1988.
- Epstein, A., Instinct and motivation as explanations for complex behaviour, in *The Physiological Mechanisms of Motivation*, D.W. Pfaff, Ed., Springer, New York (1982)

- Epstein, J. M., Learning to be thoughtless: Social norms and individual competition, *Computational Economics* Vol. 18, 2001.
- Eysenck, M.W. *Blackwell Dictionary of Cognitive Psychology*, Blackwell Publishers, Oxford, 1990.
- Fishbach, A., Shah, J.Y. and Kruglanski, A. W., Emotional transfer in goal systems, *Journal of Experimental Social Psychology*, Volume 40, Issue 6, 723-738, 2004.
- Fodor, J., *The Modularity of Mind*. Cambridge, MA, A Bradford Book, 1983.
- Franklin, S.P., Autonomous agents as embodied AI, *Cybernetics and Systems*, Vol. 28, No. 6, 1997.
- Franklin, S., Conscious software: a computational view of mind. In *Soft Computing Agents: New Trends for Designing Autonomous Systems*, V. Loia, and S. Sessa. Ed. Berlin: Springer 2001.
- Freeman, W.J. & Skarda, C.A. Representations: who needs them? In: *Brain Organization and Memory*. J.L. McGaugh, N.M. Weinberger & G. Lynch eds. Oxford University Press. 1990.
- Frijda, N and Swagerman, J. Can computers feel? theory and design of an emotional system. *Cognition and Emotion*, 1(3):235--257, 1987.
- Gat, E., Three-layer architectures. in *Artificial Intelligence and Mobile Robots: Case Studies of Successful Robot Systems*, Kortenkamp, D., Bonasso, R. P., and Murphy, R., Eds. pages 195-210. MIT Press, Cambridge, MA, 1998.
- Georgeff, M.P. and Lansky, A.L., Reactive reasoning and planning. *Proceedings of the Sixth National Conference on Artificial Intelligence*, 2, pp. 677-682, Seattle, WA: AAAI, 1987.
- Gibson, J.J., *The Ecological Approach to Visual Perception*, Boston: Houghton Mifflin, 1979.
- Greeno, G., Natures of problem solving abilities. in *Handbook of Learning and Cognitive Processes*. (Vol. 5), .K. Estes, Ed. Erlbaum, 1978.
- Griffiths, P. E., Is emotion a natural kind? in *Philosophers on Emotion*. R. Solomon, Ed. Oxford University Press, 2002.

- Hanks, S., Pollack, M.E. & Cohen, P.R., Benchmarks, test-beds, controlled experimentation, and the design of agent architectures, *AI Magazine*, 14(4):17-42, 1993.
- Hayes-Roth, B., Intelligent control. *Artificial Intelligence*, 59:213—220, 1993.
- Hayes-Roth, B, Hayes-Roth F., Rosenschein, S. and Cammarata, S. (1979), Modelling planning as incremental opportunistic process. *Proceedings of the Sixth IJCAI (IJCAI-79)*.
- Hebb, D.O. Emotion in man and animal: an analysis of the intuitive process of recognition, *Psychological Review*, 53, pp. 88-106, 1946.
- Iglesias, A. and Luengo, F. New goal selection scheme for behavioral animation of intelligent virtual agents, *IEICE Transactions on Information and Systems*, Vol.e88–d, no.5, 2005
- James, W. What is an Emotion? *Mind*. 9, 188-205, 1884.
- Kaelbling, L.P., An architecture for intelligent reactive systems, *Readings in Planning*, J. Allen, J. Hendler and A. Tate Eds. Morgan Kaufmann, 1989.
- Konidaris, G. and Barto, A. An adaptive robot motivational system *FROM ANIMALS TO ANIMATS 9 The Ninth International Conference on the SIMULATION OF ADAPTIVE BEHAVIOR (SAB'06)*, CNR, Rome, Italy, 2006.
- Kort, B., Reilly, R. and Picard, R., Analytical models of emotions, learning, and relationships: towards an affective-sensitive cognitive machine, *Proceedings of the Intelligent Tutoring Systems Conference (ITS2002)*, pp.955-962, France, 2002
- Kosslyn, S.M. Information representation in visual images, *Cognitive Psychology*, 7, 341-370, 1975.
- Langley, P. Cognitive architectures and general intelligent systems. *AI Magazine*. 27(2),, 33-44, 2006

- Lewis, S. C. *Computational Models of Emotion and Affect*, PhD Thesis, Computer Science, University of Hull, July 2004. (at: <http://www.dcs.hull.ac.uk/NEAT/dnd/agents/agents.html>)
- Maes, P., How to do the right thing, *Connection Science*, Vol. 1, 1989.
- McFarland, D., *Animal Behaviour*, Addison Wesley Longman, 1993.
- Merleau-Ponty, M., *The Structure of Behaviour*, Methuan:London, (ET 1965), 1942.
- Minsky, M.L., *The Society of Mind*, William Heinemann Ltd., 1987.
- Minsky, M.L., *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. Simon & Schuster, 2006.
- Newell, A., *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press, 1990.
- Newell, A. and Simon, H.A. *Human Problem Solving*. Prentice-Hall, 1972.
- Nilsson, N.J, Teleo-reactive programs for agent control, *Journal of Artificial Intelligence Research* 1:139-158, 1994.
- Nilsson, N.J., *Artificial Intelligence: A New Synthesis*, Morgan Kaufmann, 1998.
- Norman, D.A., Twelve issues for cognitive science, *Cognitive Science*, 4:1-33, 1980.
- Nunes, H.A., *Investigation of Motivation in Agents Using Five-Aside Football*, M.Sc. Thesis, Department of Computer Science, University of Hull, 2001.
- Oatley, K. *Best Laid Schemes*. Cambridge. Cambridge University Press, 1992.
- Ortony, A., G.L. Clore and A. Collins, *The Cognitive Structure of Emotions*. Cambridge University Press, 1988.
- Pfeifer, R. Building fungus eaters: design principles of autonomous agents, in *From Animals To Animats 4: SAB'96*, Complex Adaptive Systems series. A Bradford Book, October 1996.
- Plutchik, R. *The Psychology and Biology of Emotion*. Harper Collins. New York, 1994.

- Pylyshyn, Z. *Computation and Cognition: Toward a Foundation for Cognitive Science*, Cambridge, MA, A Bradford Book, 1984.
- Reilly, W S. Bates, J. Emotion as part of a broad agent architecture. In *WAUME 93: Workshop on Architectures Underlying Motivation and Emotion*. University of Birmingham, England, 1993.
- Rolls, E.T., *The Brain and Emotion*, Oxford University Press, 1999.
- Savage, T. The grounding of motivation in artificial animals: indices of motivational behaviour *Cognitive Systems Research*, 4, 23-55, 2003.
- Scherer, K.R., Studying the emotion-antecedent appraisal process: An expert system approach, *Cognition and Emotion*, Vol: 7, 325-355, 1993.
- Scherer, K. Appraisal considered as a process of multilevel sequential checking. In *Appraisal Processes in Emotion*. K. Scherer, A. Schorr, T. Johnstone, Eds. Oxford University Press. New York, 2001.
- Simon, H.A., Motivational and emotional controls of cognition, In: *Models of Thought*, Yale University Press, 29-38, 1979.
- Singh, P. and Minsky, M., An architecture for combining ways to think, *Proceedings of the International Conference on Knowledge Intensive Multi-Agent Systems*, 2003.
- Slovan, A. Motives mechanisms and emotions. *Emotion and Cognition*, 1(3):217—234. 1987. Reprinted in *The Philosophy of Artificial Intelligence*, ` M.A. Boden (ed), Oxford Readings in Philosophy, Oxford University Press, 231--247, 1990.
- Slovan, A. The mind as a control system, In: *Philosophy and the Cognitive Sciences*, C. Hookway and D. Peterson, Eds, Cambridge University Press, 1993.
- Slovan, A., Beyond shallow models of emotion, *Cognitive Processing*, 2(1):177-198, 2001.

- Sloman, A. and M. Croucher, Why robots will have emotions. *Proceedings of IJCAI87*, 197-202, 1987.
- Sloman, A., Chrisley, R. and Scheutz, M., The architectural basis of affective states and processes, in *Who Needs Emotions?*, J-M Fellous and M. Arbib, Eds. Oxford University Press, 2004.
- Sontag, E.D., *Mathematical Control Theory: Deterministic Finite Dimensional Systems*. Second Edition, Springer, New York, 1998
- Spier, E. *From Reactive Behaviour to Adaptive Behaviour: Motivational Models for Behaviour in Animals and Robots*. PhD Thesis, Balliol College, University of Oxford, 1997.
- Staller A. and Petta P. Introducing emotions into the computational study of social norms - a first evaluation. *Journal of Artificial Societies and Social Simulation*, 4(1), 2001
- Thangarajah, J. Padgham, L. and Harland, J. Representation and reasoning for goals in BDI agents. in *Proceedings of the Twenty-Fifth Australasian Computer Science Conference (ACSC 2002)*, M. Oudshoorn, Ed., Melbourne, Australia, 2002.
- Toates, F., *Control of Behaviour*, Springer Verlag, 1998.
- Velásquez, J.D., Modeling emotion-based decision-making, in *Emotional and Intelligent: The Tangled Knot of Cognition*, D. Canamero, 164—169, 1998.
- Wahl, S. and Spada, H., Children's reasoning about intentions, beliefs and behavior, *Cognitive Science Quarterly*, 1, 5-34, 2000.
- Wehrle, T., New fungus eater experiments, In: *From Perception to Action*, P. Gaussier and J.-D. Nicoud, Eds. pp 400-403, IEEE Computer Society Press, 1994.
- Wheeler, M., Cognition's coming home: the reunion of life and mind, In: *Fourth European Conference on Artificial Life*, Husbands, P. and Harvey, I., Eds. MIT Press, pp 10-19, 1997.

Wright, I.P, Sloman, A, and Beaudoin, L., Towards a design-based analysis of emotional episodes,
Philosophy Psychiatry and Psychology, 3(2): 101-126, 1996.

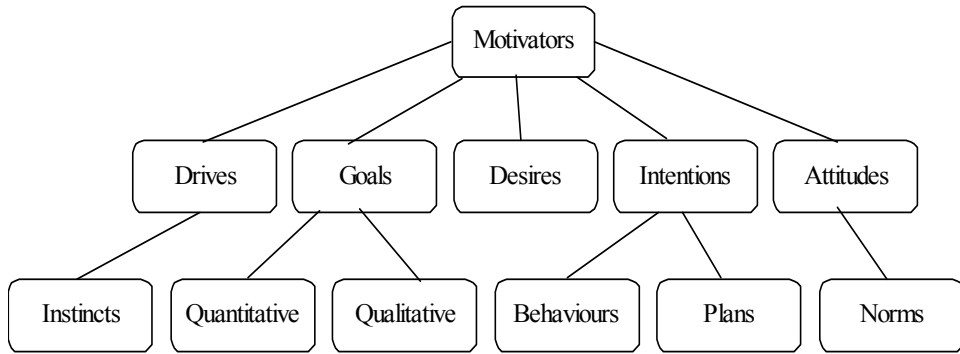


Figure 1. Taxonomy of Motivational States

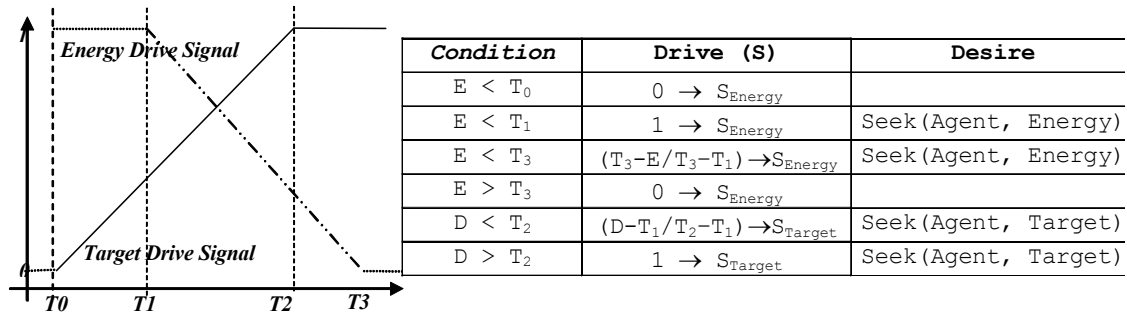


Figure 2. Model for Energy and Target Drives (with Desire Equivalents)

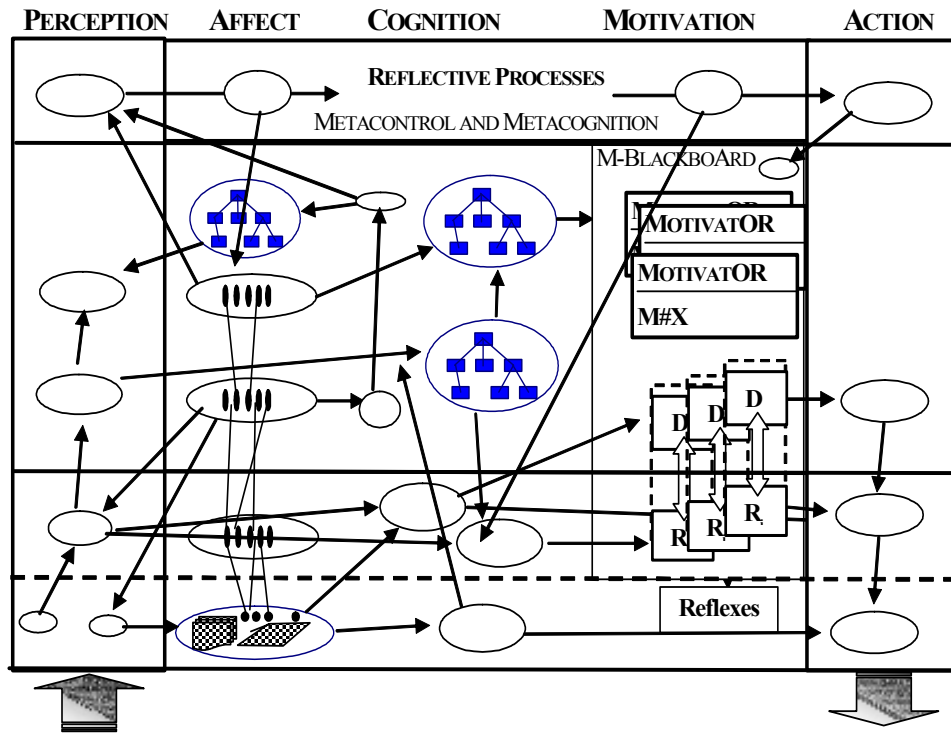


Figure 3. The Four Tier-Five Column Architecture

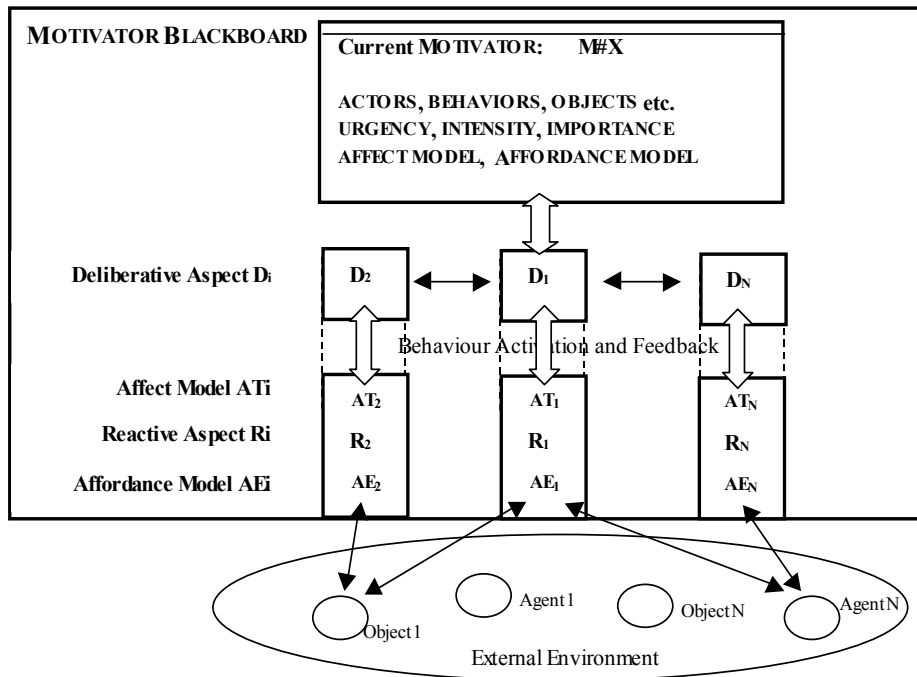


Figure 4. Motivator Blackboard Enabling the Linking of Reactive (R_i) and Deliberative (D_i) Processes through Affect (AT), and the mapping of Affordances (AE) onto Actors and Objects in an external environment

Valence	Aspect	Process and Dimension Category	Affect Magnitude
Belief Indicator	Motivator	Truth values for Semantic Content and Motivator Attitude Perception » Deduction » Assumption	[0, 1]
Commitment	Motivator	Motivator Acceptance (ignored to first priority)	[0, 1]
Dynamic State	Motivator	Motivator Process (instantiated to complete)	[0, 1]
Importance	Goal	Goal Importance (low to high)	[0, 1]
Insistence	Association	BDI Association Strength (low to high)	[0, 1]
Intensity	Motivator	Motivator Strength (low to high)	[0, 1]
Urgency	Motivator	Urgency (low to high) or time cost function	[0, 1]
Decay	Motivator	Motivator Decay (low to high) or time cost function	[0, 1]
Reinforcer	Affect	Goal and Association Feedback (negative to positive)	[-1, 1]

Table 1. Valences affecting motivational constructs.

Domain	Aspect	Definition
Robot Ball	Goal	goal(avoid(collisions), not(hit(Object)), low)
Robot Ball	Goal	goal(find(ball), found(ball), high)
Robot Ball	Goal	goal(find(robot), found(robot), medium)
Robot Ball	Association	association(environment(sparse), find(ball), reactive(true, true, method1), medium)
Robot Ball	Association	association(environment(dynamic), find(robot), reactive(true, true, method3), high)
Robot Ball	Association	association(environment(sparse), avoid(collisions), reactive(true, false, method2), high)
Robot Ball	Association	association(environment(dynamic), avoid(collisions), reactive(true, false, method2), low)
TileWorld	Belief	belief(environment(sparse), deduction, 50)
TileWorld	Belief	belief(near(tile), perception, 55)
TileWorld	Goal	goal(avoid(collisions), not(hit(Object)), high)
TileWorld	Goal	goal(pick(tile), hold(tile), high)
TileWorld	Association	association(environment(sparse), avoid(collisions), reactive(true, false, method2), high)
TileWorld	Association	association(near(tile), pick(tile), reactive(false, true, methodpick), high)

Table 2. Beliefs, Goals and Associations from Two Domains used in Experimentation.