# Why do anything?

# Emotion, affect and the fitness function underlying behaviour and thought.

Darryl N. Davis

Computer Science, University of Hull

Kingston-upon-Hull, HU6 7RX, UK

D.N.Davis@dcs.hull.ac.uk

## Abstract

What is mind? A straightforward answer is that which decides what to do next. How does a mind decide what to do next? A straightforward answer is by processing, acting and sensing. What does a mind sense? Everything? What is processed? Everything? How is everything processed? In every possible way? What actions are selected? Every action? Ten simple questions and two straightforward if rather mischievous answers. In this article differences in the nature and requirements of biological and synthetic minds is investigated in terms of control: control over what is sensed; control over how that is perceived; control over how those perceptions are processed; and control over how this epistemic flow leads to control over actions. No straightforward answers to any of the questions posed are presented. Rather, different perspectives on how investigations into these questions are used to present the thesis that some means of valencing the mind is necessary. In short this article considers how the economics of thought and action reside in the currency of affect.

## 1 Introduction

Control of behaviour is vital to animate biological systems. Mistakes in such systems lead to at best ineffective use of possibly scarce resources; at worst such mistakes lead to injury and death. Consider the scope of biological systems from solitary insects, insect communities through to vertebrates, mammals and primates. Many insects simply act out genetically determined behaviours, with the species surviving due to sheer number of individuals. The more sophisticated the biological system becomes, the more scope there is for adaptation, learning and error. The more sophisticated the biological system becomes the greater the range and diversity of type of drives that need to be fulfilled. Yet in every case the control mechanism within the biological system, whether cricket, ant, lizard, anteater, leopard or chimpanzee, needs, in some sense, to make a decision about what to do next. With increasing sophistication of biological system comes an increasing degree of autonomy. With the increasing degree of autonomy comes flexibility, the possibility of behaviour adaptation and learning. With the increased behavioural flexibility comes greater choice and a greater range of potential error. Without the increased behavioural flexibility, the range of behaviours triggered by any situation is more constrained, limited and sharply defined in terms of their effectiveness. The symbiotic nature of organism-niche evolution has determined (and continues to determine) the environmental scope of any given organism. The effectiveness of the evolved control mechanism(s) is self-evident in the diversity of biological organisms across individual and the many different environments. An important question for the designer of synthetic systems is whether there are levels of abstraction across these biological control mechanisms useful in the design of artificial systems. A further question is what types of commonality are there across the control mechanisms in these different biological systems? Salient answers to these and related questions will do more than simply provide useful insight into the design of artificial systems. It is within such a framework that the recent growth in research giving artificial systems emotional capabilities or qualities is questioned (Davis and Lewis 2004). This framework may provide the means by which we can advance our understanding of the phenomena that is affect (Sloman et al 2004).

This article makes a case for developing this framework in terms of affect, motivation and other control states, plus an analysis of niche and design space in terms of complexity of information processing structures. It then places recent investigations into 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 at other layers. The primary conjecture is that the design and implementation of such architectures can proceed using a systematic control language that obliviates the need for ad hoc heuristics to direct the processing within an intelligent system. This control language is grounded in affect. The aim is to try and develop a control language that is consistent across different domains, tasks and levels of processing. If and where this attempt to achieve this objective fails, the result will be a deeper understanding of the nature of the control systems necessary for synthetic (and natural) mind. The computational work is being developed with no explicit requirement for emotion but rather a reliance on affect (a valencing of and within internal processes), affordance and motivational constructs that together can be used to guide both internal and external acts.

## 2 Emotion, Affect and Theories of Mind

The philosophical foundations of cognitive science rest on a number of assumptions. One 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). In effect to understand how human (and similar) minds work, to develop theories about mind and to build computational systems capable of simulating (human) mind they should include both cognitive and affective mechanisms. Counter arguments to this latter claim do exist (Griffiths 2002). The argumentation for the counter claim bears similarities to that to be found in Sloman's research (2001, 2004a, 2004b).

There is a growing consensus among theorists and designers of complete intelligent systems (Minsky 1987, Sloman 2001, Franklin 2001) 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 1967, Norman 1980) in their discussions on the nature of cognition in biological systems (typically humans).

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 neural 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 are still very much open to debate.

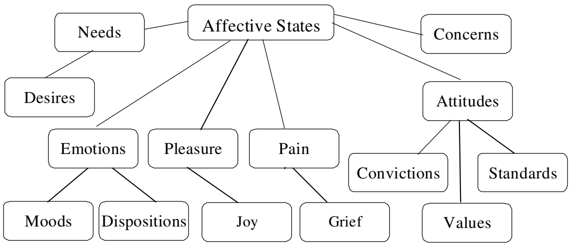

*Figure 1. An incomplete taxonomy of affective states.*

The emotions are but one type of affect among the various classes of sometimes fuzzily differentiated control states associated with mind (Simon 1967, Sloman et al 2004). Previous research has attempted to organize motivational control states in an initial taxonomy as a starting point for future research (Davis 2001b). A similar (fuzzy and very incomplete) taxonomy for affective states is shown in figure 1. The ones shown in figure 1 have been addressed, albeit in a relatively shallow manner, at a theoretical, design or computational level in earlier research (Davis 1996, 2001a, 2001b, 2002, Davis and Lewis 2003, 2004). This taxonomy and the type of categorisations made through the rest of this article are wider in scope to the conceptual analysis of emotion made in for example (Ortony et al 1992), albeit at a relatively shallow level. Section 5 of this article provides further analysis of the affective categories associated with needs and desires in terms of motivational control states.

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 many of 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. As Sloman

(Sloman et al 2004) points out, this is to be expected with any developing scientific theory.

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.

Numerous prominent researchers into intelligent systems have suggested that affect-like mechanisms are necessary for intelligence (Simon 1967; 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 (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.

## 3 A Requirement for Affect?

Norman's pivotal paper (Norman 1980) suggested emotion-like processes are 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 using a diverse range of computational mechanisms (Adamatzky 2003; Elliot 1992, Frijda and Swagerman 1987, Ortony et al. 1988; Riley and Bates 1991, Scherer 1993, Velasquez 1996, Wehrle, 1994).

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 is 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 affect. Do emotive recognisers in sophisticated interfaces require any more than the coarsest granularity in their discrimination of the possible affective state of the user?

## 4 Niches, Designs and Affect

Using the running definition as mind as a control system that decides what to do next, we now visit some alternative designs. The framework used, even if at a relatively shallow level of analysis, is the idea of niche and design space (Sloman 1995, 2001; Sloman et al 2004). Figure 2 provides a simple exemplar of alternative environmental niches, defined in terms of altitude and aquaticity, and designs for life that inhabit them.

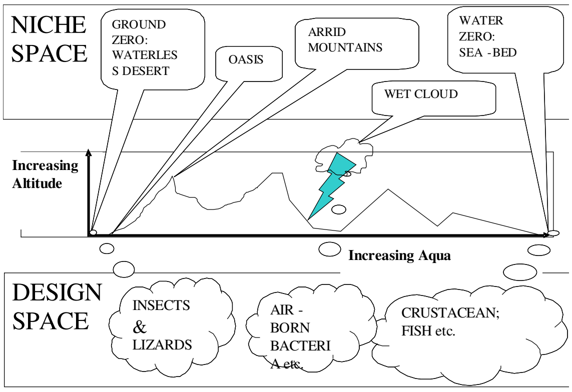

*Figure 2. Environmental niche space and associated designs.*

A different type of niche can be specified in terms of the resource and task requirements for any organism. The suggestion is that different categories of affect are associated with different levels of complexity in the structures and processes that support different classes of mind. Animal psychology and comparative ethology can help here in identifying the broad categories of mind (Davey 1989, Gibson and Ingold 1993, McFarland 1993, Toates 1998). Rolls (1999) provides four broad categories of brain complexity: mechanisms for 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. A similar continuum, in niche space, for conceptualising the increasing sophistication of mind is presented here. Along this continuum thresholds can be placed for stating the "mind" has: mechanisms of adaptation; mechanism capable of learning via the equivalent of classical conditioning; mechanism capable of learning via operant conditioning; mechanisms allowing tool use; mechanisms for tool manufacture; map formation; and the use symbols. Figure 3 shows these niche spaces and in the associated design space, examples of architectures from the animal kingdom.

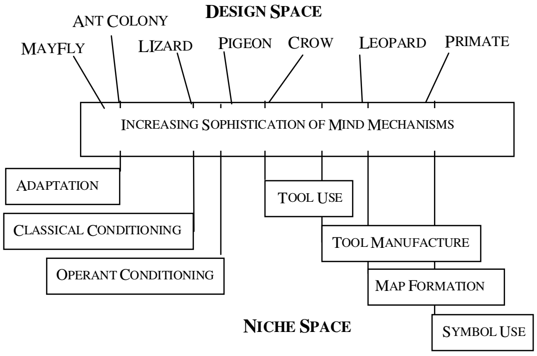

*Figure 3. A tentative niche space of increasing sophistication with associated design examples.*

From an evolutionary and anatomical perspective, there is some commonality in the mechanisms running across the dimension of figure 3. For example the chemical (hormone) and amygdala routes to behaviour in the description of the dual routes to behaviour (Rolls 1999). However while the organisms to the right of figure 3 may share the use of hard-wired finite state automata-like mechanisms, for example compare the fly-tongue reflex of the frog with the knee-jerk reflex of humans, the capabilities of the organisms to the right of figure 3 far surpass those to the left (mayflies and grasshoppers do not perform behavior rehearsal for example!).

Even a relatively trivial analysis of the opportunities offered by this perspective, shows how difficult this task is. The continuum of increasing sophistication of behaviour (and presumably mind mechanism) is neither discrete nor linear; perhaps dimension will be a better term than continuum. Consider the case of two very different organisms such as the crow and the leopard. At first it would be tempting to unequivocally suggest that the information processing requirements associated with the adaptive hunting strategies and rearing capabilities of the leopard far outstrip the information processing requirements of the crow. Yet as recent experimental evidence suggests (Weir et al 2002), the crow is capable of innovative tool creation in feeding itself, yet the leopard uses no recognisable tool. Does this place to the crow to the right of the leopard in the design space of figure 3? No! - At least not wholly to the right. The crow's behaviour while interesting is an adaptation of its natural tool making activity to support food foraging. The leopard however does use tools, for example sound, in the modifications that it can make to its hunting tactics. Typically, while stalking at night, a hunting leopard, close to a herd of prey, will typically move with retracted claws and with sometimes very slow and deliberated movement (for example fifteen metres over two hours). However it can modify this almost silent hunt, and deliberately create sound, with a pounding paw, to agitate and disorientate gazelle herds. In raising their offspring, the crow will not dwell over the loss of a brood. The leopard on the other does appear to dwell over the loss of her cubs. In short, in moving across the range of warm-blooded animals from for instance pigeons there is an information processing complexity change in moving to mammals. At that point up to the more advanced primates (for example the orang-utan) there are genera and species level partial advantages, related to fulfilling or taking advantage of specific niches and environments. The theory of affect would benefit if a similar conceptualisation as produced by ethologists were produced for affect.

## 5 Needs, Desires and Motivations

The previous section provided a tentative look at taxonomy of control mechanisms, the degree of task complexity and diversity of task repertoire. Here we look behind the behaviours to see the motivational mechanisms responsible for the behaviours. This builds on earlier work (Sloman 1990, Beaudoin 1994, Davis 2001b) on motivators. This differentiation between emotional (affective) and motivational control states is not new (Simon 1967). Here, however, previous analyses are revisited in terms of furthering the aims of the tentative analysis of affective states given in section 2.

At a very coarse grain we can differentiate between primal needs required to maintain the life-force of an individual organisms, the requirements of the species and the requirements arising from social interaction. For example, Aubé (2004) in his analysis of needs in nurturing species, differentiates between primal needs, that are related to the resource requirements of an individual organism, and second order resource requirements, that are related to requirements arising and made available through activities such as social bonding and collaborative behaviours. Aubé suggests that the affective states associated with these requirements differ too; he terms these commitments.

An alternative (and perhaps complementary) approach is to look to develop the taxonomy of primary reinforcers that Rolls (1998:table10.1) provides. That taxonomy is differentiated primarily in terms of five sensory modalities, reproduction and a collection of diverse reinforcers related to social and environmental interactions. The relevance is that these reinforcers, either positive or negative, are mapped onto drives and affective states. In the somatosensory modality for example pain is a negative reinforcer, while touch positive. Control over action is a positive reinforcer

In accordance with earlier research (Davis 2003) needs are manifested in processing terms as drives. Drives are low-level, ecological, physiological and typically pre- or non-conscious. They provide the basis for an agent's behaviour in the world, are periodic but short-lived and are defined in terms of resources essential for an agent. Such activities for information agents include the need to gather resources and propagate information to associates in their society. In biological agents such drives include thirst, hunger, and reproduction. Nurturing sublimates some of these drives in the service of others. 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. We can model such drives relatively easily in computational agents using intrinsic control structures. Prior work (Davis 2003) used fuzzy logic models to do just that.

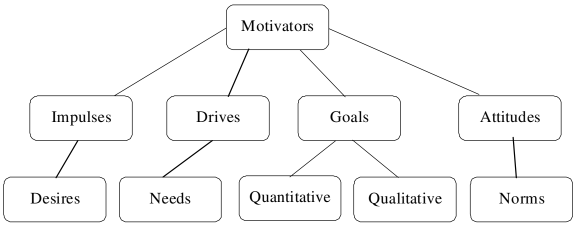

*Figure 4. Taxonomy of Motivational States*

Having established a primal motivational category, we can now look further at the types of taxonomy produced for motivational control states (Davis 2001b). Figure 4 provides four major types with, in each case, subcategories. In keeping with the theme of tentative dimensions for control states that is being used throughout this article, there is an implied ordering from left to right across figure 4. The processing requirements and representational qualities associated with these four broad categories become more sophisticated towards the right of the figure.

*Impulses* are related to spontaneous behavior, for example suddenly leaving the cinema during the screening of a film or making a rash purchase. They are associated with the instantaneous formation of an idea, perhaps unrelated to current cognitive context, and can cause a temporary or more persistent re-focus of mind. Here *Desires* are only partly analogous to their use in BDI agent architectures (Georgeff and Lansky 1987), for example *desires(agent, stateY)*.. Desires can underpin goals and other purposeful behavior. Desires and impulses are akin in that impulses may arise out of desires, and that neither need be realistic, achievable or rational. Drives and needs, as described above, do not require deliberative mechanisms and architectures capable of supporting adaptive state automata suffice to model these. Quantitative goals can encompass needs and drives but are differentiated to allow for more flexible representations and process models. These are the types of goals discussed in engineering control theory (Sontag 1998) and reinforcement learning, for example (Maes 1989, Toates 1998). Qualitative goals are the types of motivators discussed in most planning literature (Nilsson 1998). The remaining category identified here, attitudes, are pre-dispositions to respond to specific sensory or cognitive cues in specific ways. For example, an agent could generate pro-active goals to investigate a hapless agent based on an altruistic standard (an attitude) and a set of beliefs about the capabilities of that agent. The work on norms (Staller and Petta 2001) is relevant to this category. The following sections describe computational work in bringing together these analyses in terms of working models. For conceptual (and historical) reasons motivational control states are dealt with before the computational model of affect.

## 6 Motivated Architectures

Current work on architectures for motivated agents is based on experiments in the theory, design and implementation of affect and emotion based architectures (Davis 1996, 2001a, 2001b). 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. This led to work on agents and control states (Davis 2001b), again focused on goal processing. It addressed how goals need to be valenced in a number of different ways, for example intensity, urgency, insistence (see table 1). Motivators in these architectures were representational structures generated at a reactive level. The generic representational schema made use of fifteen components that reflected the nature of the most expansive of motivational control states. In many instances, for example behaviours related to drives, many of these components were unused and the stack of motivators could 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 2003a), which made use of mechanisms analogous to those used in teleological planning (Nilsson 1994).

| **Valence** | **Process and Dimension Category** |
|---|---|
| Belief Indicator | Function over Truth values for Semantic Content and Motivator Attitude |
| Commitment Status | Fuzzy Model (ignored to first priority) |
| Dynamic State | Fuzzy Model (instantiated to complete) |
| Importance | Fuzzy Model (low to high) |
| Insistence | Fuzzy Model (low to high) |
| Intensity | Fuzzy Model (low to high) |
| Urgency | Fuzzy Model (low to high) or time cost function |
| Decay | Fuzzy Model (low to high) or time cost function |

*Table 1. Valences within a motivational construct.*

The architectures developed in this work, and related research into a multi-level representational framework for emotion (Davis 2002), made use of variations of the three column, three level architecture developed with the Cognition and Affect project (Beaudoin 1994, Davis 1996, Sloman 1990, 1995, Sloman et al 2004, Wright et al 1996).

We continue to use variations of a three-column, three layer architecture but are not unequivocally committed to such architectures, if the research requires other frameworks. Figure 5, for example, shows a four tier, five column instance. Some experimentation (Davis and Lewis 2003, 2004) makes use of an architecture based on cognitive models of reasoning in children (Bartsch and Wellman 1989, Wahl and Spada 2000). The approach taken is one merges the principles of architectural parsimony (Hayes-Roth 1993) and the conceptual ease through architectural expansion of Singh and Minsky (2003).

In the three-layer 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 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. The reflective processes serve to monitor cognitive behaviour or control it in some other way. The more extreme affective states (symptomatically categorised as a loss of control or perturbance) are effectively clamped by means of self-regulatory processes within the architecture. This model is quite general. The effect of altering the relative size and importance of the layers is an open issue. 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 and motivational 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.

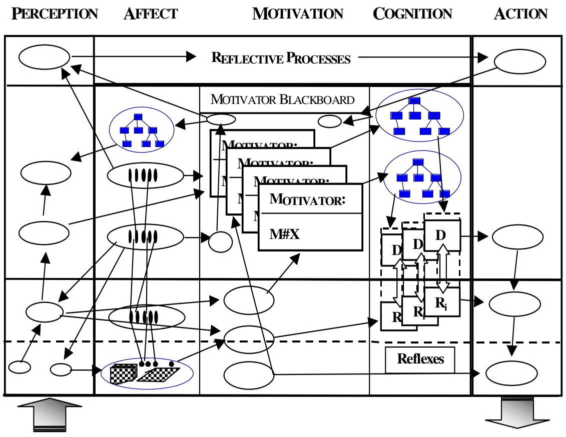

*Figure 5. The Four Tier-Five Column Architecture*

The most recent design work sketched in figure 5 shows a four tier five column architecture. The four tiers represent reflex, reactive, deliberative and reflective processes. The five columns separate perception, affect and affordance processes, the motivational blackboard, generic cognitive processes (planning, behaviours etc) and output to actions. This framework extends earlier work with the architectural global blackboard for the motivational constructs. Earlier research (Davis 2001b) did not separate these processes from generic cognitive functions. This architecture makes use of the extended motivational constructs as blackboards that provide the context for ongoing (and most other dynamics of) processing. The representational structure that is the architecture can use one or more motivational constructs concurrently. Both architecturally generic and motivational construct specific processes can access the blackboards and in turn be influenced by their content and processes. The emotion engine of earlier research (Davis 2001a) is now superceded by the affect processes column. The work on multi-level representations of emotions that run over semi-autonomous cellular automata models is being revisited in the light of current thoughts on the nature of affect (as outlined in this article) and the work of Adamatzky (2003) on computational chemistry models of affect. The latter in hand with the blackboard scheme for motivation provide a sophisticated interaction of very low level, reactive and deliberative processes in a multiply valenced framework.

The affective 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. It forms 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 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 valences used in the current motivational structure (table 1) 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.

## 6.1 Affect, Affordance and Motivation

Previous research (Davis 2001a) has used emotional models that include basic emotions. The current stance is that basic emotions are unnecessary in a theory of emotion. A number of emotion theories use the concept of basic emotions; Scherer (1994) 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. However the range of states in a modal model is far more diverse. A salient feature of many theories of emotion is that they are described in terms of goals and roles. Emotion in the goal-based theories, for example (Oatley 1992), can be described as "*a state usually caused by an event of importance to the subject*". This involves mental states directed towards an external entity (attitudes, motivations, expectations etc.), physiological change (increased heart beat, hormone release etc), facial gestures and some form of expectation. Scherer (1994) defines emotion as *"a sequence of interrelated, synchronised changes in the states of all organismic subsystems (information processing, cognition, support, ANS, execution, motivation, action, SNS, monitoring, subjective feeling) in response to the evaluation of an external or internal stimulus event that is relevant to central concerns of the organism"*. These 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. This differentiation of processes can be easily mapped onto the architectural model of figure 5. While still accepting the validity of developing a computational theory of emotion, there is a very important adjunct. Emotions are considered unnecessary for most synthetic systems, and that the case for including emotion in a synthetic system should be based on an analysis of the demand for emotions in the developed system. Given the motivational framework outlined in the previous sections, the requirement is that some model of affect is required. This may not necessarily involve the emotions, and may be simpler in its requirements than the mechanism necessary for a fully developed implementation of the emotions.

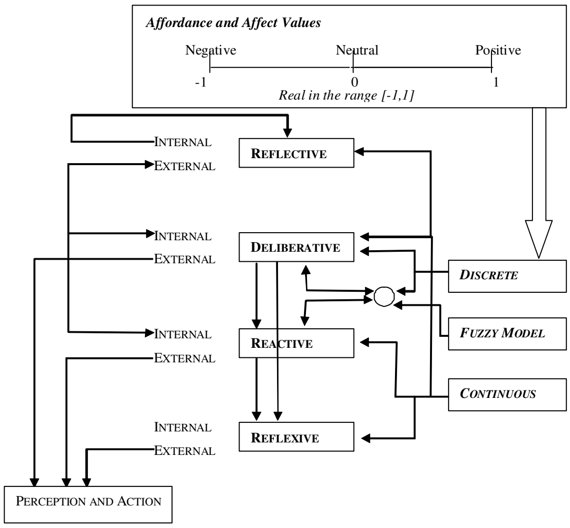

*Figure 6. Affect Model (in second column of figure 5)*

We are developing a theory of affect that draws on themes such as control states and motivators (Simon 1967; Sloman 1990; Davis 2001b) and affordances (Gibson 1979; Davis 2001a). The overlap between the goal-based and modal response theories provides for a coherent hybridization and defines the bare bones of the affect model used here. We define affect in terms of reinforcers over processes and representational structures. It is qualitatively defined over negative, neutral or positive values, as in the work of Rolls (1999), or numerically over the interval (-1.0,1.0). Future work will look to develop the fine details of a fuzzy (and/or neural) valued processing model that maps across these measures at different levels of the architecture (Figure 6). This will build on the research on the emotion engine and also relate to the eight valences for the currently developed motivational construct (Table 1). Hence, affect forms a basis for a control language for agent architectures. It allows external events and objects to take valenced affordances, and allows the results of internal mechanisms to be prioritised and compared via 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, and particularly extreme and disruptive control states, 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 fashion.

## 7 Discussion

This paper has 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. More pointedly, emotion is not really a requirement for many forms of synthetic intelligence, and that more straightforward affective means, based on something as straightforward as the concept of affective taxes or reinforcers, can be used to enable effective decision-making. Elsewhere, it has been suggested (Davis and Lewis 2004) that a direction given by the less semantically overloaded term affect is more appropriate for synthetic intelligence. The problem is however that the phenomena covered by affect are even more diverse and currently less well specified than emotions! Future research will determine how complex are the states arising from the adoption of the simple model outlined here.

If our research agenda is slightly different and pursues the theory, design and building of artificial systems that sometimes work analogously to human mind does this requirement for emotions still hold? In negating the use of emotion some alternative is required, not just to simply 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 valence or fitness function associated with these, whether explicit or implicit. Some means of conflict resolution is required. 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 (see for example Englemore and Morgan 1988). Here we suggest that at a theoretical, design, architectural and implementation level a consistent valencing and control language based may offer more to the pursuit of synthetic intelligent systems. That this language at times bears similarities to the language used to describe emotion and affect should not be surprising.

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. There should exist some global mechanisms that at least provide for context and integration of modules. It matters not for the time being whether the global mechanisms for context are based on ideas such as computational chemistry (Adamatzky 2003), or global workspaces (Baars 1997, Franklin 2001) or blackboards (Hayes-Roth 1993) or some combination or neither. Should and how can the control architecture make consistent decisions across these different modules and mechanisms? We suggest the use of multiple-level representation based on the idea of affective taxes. This will bear some similarity to aspects of a number of theories of emotion where they serve useful satisfaction for system requirements. For example in integrating behaviours (whether innate, adapted or acquired) into a skill sequence for a particular context, affective dissonance provides a fitness function to be minimized. At the individual module level, we require a fitness function mapping input to output (for example as an affordance and accordance over the requisite sensori-motor mapping). At a more abstract level, we are using a representational schema (Davis 2001b) as local blackboards for reasoning about motivations, goals and other forms of control states. Again we look to provide a consistent valencing mechanism across control states, behaviours and architecture levels.

The theory of synthetic intelligent systems can therefore progress without the need for emotion per se 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 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 emotions. Furthermore most theories of emotion involve body attitude or facial expression changes that are typically inappropriate for machines. As yet, there are no machines that rely on body posture or facial expression for inter-communication other those affective systems that attempt model the emotive state of their user (Picard 1997). Even there the interactive system needs only to model the emotive or affective state of its user, and not function in terms of emotion.

## 8 Conclusion

Recent experimental work (Nunes 2001, Bourgne 2003) has revisited the representational structure and processes associated with motivators (Beaudoin 1994, Davis 1996), but made use of affect and affordances to valence the motivational constructs. Associated with motivational structures are attitudes to classes of events and entities relevant to that motivator. These are valenced in the same way that affordances and affect are valenced. 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 can now be defined and compared in terms of a common criteria. Affect and affordance become the means by which an 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 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.

Returning to the theme of the introductory paragraph, 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 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 does the designer of the computational equivalent to a grasshopper? Albeit a grasshopper that can manipulate text?

The reason these questions is 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? 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 work of Sloman and others (2004) has been given here.

## References

*Adamatzky, A., Affectons: automata models of emotional interactions. Applied Mathematical Computation, 146(2), 579-594, 2003.*

*Aubé, M., Beyond needs: Emotions and the commitments requirement. In: Davis, D.N. editor, Visions of Mind: Architectures for Cognition and Affect, Idea Group Inc., Hershey, PA, USA, 2004.*

*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.*

*Bates, J., Loyall, A.B. and Reilly, W.S., Broad agents, SIGART BULLETIN, Vol. 2, No. 4, 1991*

*Beaudoin, L.P., Goal Processing in Autonomous Agents, Ph.D. Thesis, School of Computer Science, University of Birmingham, 1994.*

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

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

*Damasio, A. Descartes' error: Emotion, reason and the human brain. New York: Grosset/Putman, 1994.*

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

*Davey, G., Ecological Learning Theory, Routledge, 1989.*

*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., Multiple Level Representations of Emotion in Computational Agents, Emotion, Cognition and Affective Computing, AISB2001: Agents and Cognition, University of York, 2001*

*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, Copyright 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):159-165, 2003.*

*Davis, D.N. and Lewis, S.C. Affect and Affordance: Architectures without Emotion. AAAI Spring symposium 2004, Stanford University, California, USA*

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

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

*Elliott, C., The Affective Reasoner: A Process Model of Emotions in a Multi-Agent System. PhD Thesis. Northwestern University, USA, 1992.*

*Englemore, R. and Morgan, T., Blackboard Systems, Addison-Wesley, 1988.*

*Fodor, J., The Modularity of Mind (Cambridge, MA, A Bradford Book), 1983.*

*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.*

*Franklin, S., Conscious Software: A Computational View of Mind. In Soft Computing Agents: New Trends for Designing Autonomous Systems, ed. V. Loia, and S. Sessa. Berlin: Springerm 2001.*

*Frijda, N and Swagerman, J.. Can computers feel? theory and design of an emotional system. Cognition and Emotion, 1(3):235--257, 1987.*

*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, E. and Ingold, T (Editors), Tools, Language and Cognition in Human Evolution, Cambridge University Press, 1993.*

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

*Griffiths, P. E., Is emotion a natural kind? In Solomon, R., editor, Philosophers on Emotion. Oxford University Press, 2002.*

*Hayes-Roth, B., Intelligent control. Artificial Intelligence, 59:213—220, 1993.*

*Hebb, D.O. Emotion in man and animal: an analysis of the intuitive process of recognition, Psychological Review, 53, pp. 88-106, 1946.*

*James, W. What is an Emotion? Mind. 9, 188-205, 1884.*

*Kaelbling, L.P., An Architecture for Intelligent Reactive Systems, Readings in Planning, Morgan Kaufmann, 1989.*

*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*

*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.*

*Nilsson, N.J, Teleo-reactive programs for agent control, Journal of Artificial Intelligence Research 1, 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., Clore, G.L. and Collins, A., The Cognitive Structure of Emotions. Cambridge University Press, 1988*

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

*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.*

*Scherer, K. Studying the emotion-antecedent appraisal process: An expert system approach, Cognition and Emotion, 7, 325-355, 1994.*

*Scherer, K. Toward a Concept of 'Modal Emotions'. In P. Ekman and R. Davidson (Eds): Nature of Emotion, Oxford University Press. New York, 1994.*

*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.*

*Simon, H.A., Motivational and emotional controls of cognition, Originally 1967, Reprinted 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.*

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

*Sloman, A., Motives mechanisms and emotions. In: M.A. Boden (Editor), The Philosophy of Artificial Intelligence, Oxford University Press, 1990.*

*Sloman, A. Exploring design space and niche space, Proceedings of the 5th Scandinavian Conference on AI, Trondheim, Norway, IOS Press, 1995.*

*Sloman, A., Beyond shallow models of emotion, Cognitive Processing, 2(1):177-198, 2001.*

*Sloman, A. What are emotion theories about? Multidisciplinary workshop on Architectures for Modeling Emotion at the AAAI Spring Symposium at Stanford University in March 2004.*

*Sloman, A., Chrisley, R. and Scheutz, M.The Architectural Basis of Affective States and Processes In: 'Who needs emotions?', edited by Jean-Marc Fellous and Michael Arbib, 2004.*

*Sontag, E.D., Mathematical Control Theory: Deterministic Finite Dimensional Systems. Second Edition, Springer, New York, 1998*

*Staller A. and Petta P.: Introducing Emotions into the Computational Study of Social Norms - A First Evaluation. In: Edmonds B. & Dautenhahn K. (Eds.),* Starting from Society: the Application of Social Analogies to Computational Systems, *Special Issue of the* Journal of Artificial Societies and Social Simulation*, 4(1), 2001*

*Toates, F., (Editor), Control of Behaviour, Springer Verlag, 1998.*

*Velasquez, J. D. Cathexis – A Computational Model for the Generation of Emotions and their Influence in the Behaviour of Autonomous Agents. PhD Thesis. MIT, 1996.*

*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: P. Gaussier and J.-D. Nicoud (Editors) From perception to action, pp 400-403, IEEE Computer Society Press, 1994.*

*Weir, A.A.S., Chappell, J., and Kacelnik, A., Shaping of hooks in New Caledonian crows. Science, 297:981, 9th August, 2002*

*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.*