Cognitive Science Quarterly (200?)

Motivation, Drives and Goals

Darryl N. Davis

Neural, Emergent and Agent Technology Research Group

Department of Computer Science

University of Hull

Kingston-upon-Hull

England

HU6 7RX

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

Fax: +44-1482-466666

Home page: http://www2.dcs.hull.ac.uk/NEAT/dnd/index.htm

Abstract

This paper explores the concepts of motivation, drives, goals and emotion from the perspective of computational architectures for synthetic minds. Our research into motivation and emotion using computational agents has grown out of the control state approach to cognition and affect. Initial work used filter-protected multi-layer architectures to model a variety of motivational states. The most complex of the early implemented agent architectures, although relatively naïve, used value systems to differentiate between self-centred and selfless motivational states in a simulated environment. The agent's autonomy was modelled in terms of these motivational states and made use of a shallow emotional model. More recently we have explored computational architectures for the modelling of affective states, specifically emotion. We are pursuing a line that links affect, motivational states (for example impulses, drives, goals and concerns), emotion and cognition. The initial research in this area is being revisited as we consider the effect of placing a computational analogue to emotion at the core of a model of synthetic mind. The latest directions involve multiple forms of autonomy; indeed emotion is viewed as a facet of autonomy in cognitive agents. Designs and implementations involve direct perception of an agent's environment in terms of affordances given by current, transient and mediating motivational states. The theoretical architecture involves the use of learning mechanisms that use affective states to provide the criteria for measuring the effectiveness of behaviours and affect strength to indicate importance of memory events

Keywords: Architectures, motivation, emotion, goals, drives, concerns.

Introduction

What is motive-less behaviour? It is random, haphazard and decisionless behaviour. Directed or purposeful behaviour is a result of some form of motivational state. The triggers for such motivated behaviour can be implicit or explicit. The processes responsible for managing such triggers can be conscious or subconscious. Such processes and any resultant behaviour can be described as rational or irrational. In this paper the concepts of motivation, drives, goals and emotion are explored from the perspective of theoretical and implemented computational architectures. The aim is threefold. Firstly to identify and understand the issues pertaining to these concepts. Secondly through studying alternative computational designs and architectures to further the predominantly philosophical and psychological analysis of motivation through a consideration of mechanisms capable of supporting such phenomena. Thirdly to make use of the results of these inquiries in advancing computational intelligent systems.

Our research paradigm is that of studying complete agents (Davis 2001b). This research is in effect integrative, drawing together a number of threads in cognitive science and artificial intelligence, for example behaviour, decision-making, memory, learning. Such agents are designed to posses 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, Kaebling 1989) in their rejection 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 research described in this paper differs from the behaviour-based approach in a number of important ways. Firstly an environment need not be physical – an agent can be situated in synthetic environments whether those of experimental test-beds and harnesses (Hanks et al. 1993) or information-system environments typically associated with databases (Davis 2000b) and the Internet. (Luo et al. 2001). Indeed most a-life scenarios are synthetic. Secondly there are many aspects of intelligence as studied in philosophy, psychology and cognitive science that are abstract and unrelated to an agent's environment. There are also reservations about what lessons can be learned about the nature of general intelligence from the design and implementation of no matter how many task-specific robotic systems.

The thesis that will be pursued here is that if we consider the control state perspective to cognition (and affect) is that while many control states (such as motivations and drives) are highly connected via processes internal to an agent to the agent's external environment, some are not. A complete (and intelligent) agent is *not* limited to its external environment. The pursuit of such synthetic agents at the theoretical and design level is a long-term activity. To facilitate research, decisions need to be made about the direction of any project and what is immediately achievable at theoretical and computational levels and how those results segue with long-term research goals. Hence the adoption of a variety of the broad but shallow approach (Bates et al. 1991) to investigating complete agents. In short we research complete agent architectures that exhibit specific qualities related to immediate research goals. These qualities should be based on a deep understanding and sound theoretical basis. Other qualities may or may not be present at an implementation level but are not excluded from the theoretical framework that drives any specific research project. This will mean that initial theoretical models for a complete agent architecture are laterally expansive but process shallow. The inadequacies of any such model can then be tackled in subsequent or alternative research projects. One worry with this approach is whether our research goals can be achieved in using an incremental approach to designing and building complete or cognitive agents. There always appears to be a huge distance between implemented and theoretical architectures. Decisions about which aspect to investigate more fully result in years spent on deepening the implemented architectures. This is countered by the feedback that empirical computational work has on the theory and further designs and implementations.

Control State Perspective on Cognition and Affect

The control state approach to cognition (Simon 1979, 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 other reactive. Other changes are emergent and may be unpredictable. There are a number of thorny issues that can be tackled from this perspective.

The AI and Cognitive Science literature is littered with overlapping definitions, with the same term being used as a referent to many kinds of disparate activity. Consider the term goal. 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 of a problem is defined as the goal state. From this perspective, problem solving is

- the transformation of an initial state into the goal state where there is no obvious path or method of achieving a solution; otherwise the *problem* (sic) is trivial;
- a cognitive behaviour and therefore an internal, not directly observable, behaviour;
- a process involving the manipulation of knowledge typically referred to as symbolic, but need not necessarily be so - see (Kosslyn 1975);
- is directed; i.e. there is an *intention* to produce a solution or transform the problem in some way;
- a personal experience, dependent upon existing knowledge.

Aspects of this model can be seen in many artificial intelligence models of problem-solving. From a broader and complementary perspective problems can be categorised (Greeno 1978, Eysenck, 1990) in a number of ways as

- Routine problems, with well defined paths from given to goal state;
- Well described problems, where a clear representation is possible;
- Inducing structure problems where patterns are generalised from examples;

- Transformation problems in state space, for example the Towers of Hanoi;

- Arrangement problems, requiring the arrangement of mainly pre-existing problems and partial solutions, for example (Hayes-Roth et al. 1979);

- Ill defined problems, where the problems are described (if possibly at all) in quite vague terms.

There would seem to be a continuum (perhaps in something other than linear space) from the trivial (non-problem) problem to the most ill defined of problems, where the cognitive behaviour necessary to further the given state (if it can be described as such) is contingent upon creativity. Where representation is possible, problems can in effect be said to be understood (Polya 1945), if not necessarily solved, or even currently solveable, and may be represented in terms of some arrangement of sub-goals. These issues will be returned to at a latter point in this paper.

A starting point for this analysis is that presented by Sloman (1987) and others (Beaudoin 1994, Davis 2001b). This provisional (and admittedly incomplete and fuzzy) taxonomy provides us with five different forms of control state (see figure 1). One of the major types (*motivators*) consists of four further categories. Briefly these control states, which are categorisations of mental phenomena, can be described:

- Beliefs are internal models of the world, possibly inferred from perceptual acts or from information arising from other control states; these need not have a rational basis;

- Images are control states using mental images, whether pictorial, diagrammatic or iconic. These images may relate to any (number) of perceptual modalities in typical or atypical ways (e.g. synaesthesia);

- Imaginings are control states that embody alternative ways of constructing internal worlds. These can be related to directed problem-solving and therefore motivators or more whimsical (e.g. day-dreaming).

- Motivators are a disposition to assess situations in certain ways; i.e. a `glass' through which internal and external `reality' are viewed, involving perception of problematic events and states, representations and paths to modified states of affairs;

- Reflexes are ballistic mappings from input (i.e. perception) to output (i.e. behavioural response) and can involve single actions (e.g. knee jerk) or multiple actions (e.g. balancing on a uni-cycle).

Motivators can take several sub-forms:

- Goals whether quantitative, qualitative or hybrid. Quantitative goals are those type of goals talked about in control theory, and tend to involve negative feedback. Qualitative goals are similar to most artificial intelligence goals (especially in the planning literature) and can involve relations, predicates, states, behaviours etc.. Hybrid goals are some mix of these two sub-types;

- Attitudes are predispositions to respond or act (either internally or externally) to specific (perceptual or cognitive) cues and can involved intricate collections of beliefs, motivators, etc. Standards are prescriptive or relative attitudes that embody ethical, social or personal rules.

- Emotion drives many thought processes and actions on the external world. These transient states give rise to and are also a product of moods. These more persistent states can be viewed as emergent states that pervade the entirety of cognitive processing or a side-effect of other control states. Certain moods favour certain motivators and inhibit others; i.e. they are closely related to predispositions and attitudes.

- Desires are related to beliefs, attitudes, emotions and imaginings. Desires can relate to objects, agents or events in an agent’s internal world that may relate to plausible states in the external world. Desires may be irrational and relate to an alternative model of the world that is implausible. Impulses are transient desires which when acted on give rise to non-transient changes in the internal, external or both worlds.

An alternative taxonomy of mental phenomena (control states) is given in figure 2. Wollheim (1999) differentiates between mental events and states in terms of the degree of consciousness associated with that event or state. Sub-conscious events and states remain so. However they may give rise to events and states which are ultimately manifested as conscious states. An agent is aware of and may deliberately provoke conscious states and events. Preconscious events and states can give rise to and be invoked as conscious phenomena. However no mental event or state can move between these different levels without some form of transformation. A preconscious mental state in being invoked by the conscious mind is transformed and gives rise to an *analogous but different* mental state. The mind focussing its attention on an ongoing event is qualitatively different to the mind invoking a memory of that mental event. Some aspects of that mental event are effectively lost to the subconscious. These different categories of mental states coexist and interact but never be identical. For the migration of information and knowledge across these phenomena, conscious, pre-conscious and subconscious mental states must overlap in a manner analogous to the dove-tail joint in carpentry.

This current research looks to simplify these taxonomies and disambiguate the linguistic complexity associated with the above models. Drives, concerns and goals are specific forms of motivational control states. It is possible to differentiate between these three on the basis of their processing and control requirements. Other terms for motivational control states, for example desires and attitudes, are to be excluded until required (if at all). One further piece is required for this jigsaw arising out of the existentialist school of thought.

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 within 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 valenced descriptors for drives, concerns and goals. Hence disequilibria are a form or source of motivator.

Drives are low-level, ecological and physiological and typically pre-conscious. They are periodic but short-lived and include such things as the need to find food, energy sources etc. Goals require some form of planning (for example, state-based reasoning) and an explicit differentiation between an agent's model of its external environment, the external environment itself and the agent's wholly internal environment. Goals may arise out of other control states but should be thought of as related to conscious states of mind. Concerns are motivators that arise from belief sets about other entities in the agent's world that are of importance to the agent. Concerns when active are a facet of conscious mind; when inactive or dormant a facet of preconscious mind. Consider a predatory agent that may have drives to hunt prey. It may develop a goal to explore its external environment when no prey agent can be sensed. The same agent may develop a goal on how to reach a certain prey agent before another predatory agent or before the prey agent reaches safety. A concern to such an agent reflects the events, objects and agents associated with any goal. The conjecture provided here is that emotion suffices as the phenomenon that underpins these control states. The representation of such drives, goals and concerns in motivational structures is returned to in the section on empirical work.

Motivation and Emotion

There exist a number of overlapping and disjoint models for the emotions in philosophy and psychology. Some more easily relate to computational models of cognition. In doing so they afford descriptors at multiple levels for motivation and other control states. In short emotion and cognition are highly inter-linked. There is however little agreement across these fields on a definition of emotion. One possibility is to discard the term emotion but are there alternatives to the very fuzzy, ambiguous and misleading term emotion? Duffy (1962) considers the use of these terms 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 now such a volume of research in this area that a significant academic revolution would be required to pursue such a path with any success. Unfortunately these many perspectives gives rise to some confusion in the nature of the phenomena described by the word emotion. By addressing different philosophical analyses and psychological theories of the emotions we may able to more clearly define what is meant by emotion, and delineate between emotions of different categories, emotional events and emotional states. One way in which we can do this is by considering whether any of these perspectives are open to computational modelling. However unlike some analyses of mind that define emotion as an emergent or perturbant mental state, the developing theory presented here places emotion at the core of mind and not secondary to rationality. If emotions are vital to cognition, this raises the question of whether the computational modelling of human-like minds is possible without a synthetic analogue to human-like emotions? A further question is whether the development of any form of synthetic mind is possible without a computational equivalent? In many analyses emotion is seen as perturbant. The model being pursued here is influenced by Kanzei research. Kanzei draws on a Shinto perspective on activity that involves the essence or spirit of the agent wholly engaged in the activity. Kanzei emphasises harmony and thus offers a more positive approach to research in affective states. A brief perspective on this model allows primary emotions (at the drive level in motivational terms) to be considered in terms of alarms and valences. These are preconscious triggers for emotions related to an agent's current set of activities and expectations in terms of at low-level feedback associated with the results of actions and behaviours. A second level associated with motivations relates to concerns, or reasons why emotions may be initiated at a conscious or deliberative level. Concerns are related to goals. A further level of emotions related to our computational architectures are tertiary emotions. These are meta-control states such as perturbation and harmony.

Architectures for Cognition and Affect

Our research into motivation and emotion using computational agents has grown out of the control state approach to cognition. Initial work (Davis 1997, 2001b) made use of a filter protected multi-layer architecture to model a variety of motivational states. The most complex of the early implemented agent architectures, although relatively naïve, used value systems to differentiate between self-centred and selfless motivational states in a simulated hazardous environment. The agent's autonomy was indeed modelled in terms of these motivational states and made use of a very shallow emotional model. More recently we have explored computational architectures (see figure 3) for the modelling of other affective states, for example emotion (Davis 2000a, 2001a). We are pursuing a line that links affect, motivational states (for example impulses, drives, goals), emotion and cognition. The initial research in this area is being revisited as we consider the effect of placing a computational analogue to emotion at the core of a model of synthetic mind. The latest directions involve multiple forms of autonomy; indeed emotion is viewed as a facet of autonomy in cognitive agents. The latest designs make use machine perception including direct perception of an agent's environment in terms of affordances given by current, transient and mediating motivational states. The theoretical architecture involves the use of learning mechanisms that use affective states to provide the criteria for measuring the effectiveness of behaviours and affect strength to indicate importance of memory events (Rolls 1999). Various synthetic harnesses are being used as experimental vehicles for the investigation and development of theory, design and implementations. These include simulated robotic factories (Davis 2001b), Tileworld (Hanks et al 1993), simulated robo-cup and predator-prey scenarios using a-life techniques. One justification for using a-life techniques comes from (Wheeler, 1997); another from the computational experiments into social processes (Epstein & Axtell, 1997).

Empirical Work

Empirical work on the above issues makes use of computational agents running various designs based on the architecture shown in figure 3. This is described in depth in other papers (Davis 1997, 2001a, 2001b). Here a brief description of how motivational structures are used in computational experiments into simple social interactions is presented. These experiments make use of multi-agent worlds but with no explicit co-operation. Agents need to feed to sustain themselves (all actions expend energy) and can reproduce with other agents of the same type. Some agents prefer the company of others. Some agents feed on vegetation (static objects in their environment) while others are predatory. The agents perceive their environment through facilities available in the agent toolkit (simulated perception) or from "energy" images. As agents consume energy it is left as a pixel trail in this image which can be sampled. Different types of object and agent leave different trails. This colour image is used for experimentation with valences and direct perception using active retinas modelled as adaptable cellular automata.

Drives makes use of signal strength based on two factors – an internal measure (energy) and an external measure (distance from some object or agent). The signal strength based on internal energy (used in the food and reproduce drive) is heuristically defined (see figure 4) and subject to experimentation, particularly with regard to the various thresholds. While physiological models could be used, they are computationally more complex and offer no advantage to the research goals being pursued. The signal strength associated with the Hunger drive makes use of the current energy level (E) and three thresholds as given in figure 4. A similar model can be used for the drives associated with the agent's perception of its external environment. These are the control and process paths depicted at the base of figure 3 and labelled as reflexes. Very simple agents can function and perform tasks based on this (impulse) control loop with no further computational infrastructure. Decisions are in not made and the agent performs according to the weighted sum of its drives.

As the reactive architecture is added the agents can respond to both internal and external pressures but with the impulses weighted according to their signal strength. Alternative decision making mechanisms allow the selection of impulses with the highest signal as a primary goal and then any other that does not conflict with this behaviour. For this test harness all agent behaviours result in a movement in two-dimensions. A conflict in behaviour is simply a vector that distracts from the direction vector of the primary goal. Again there are parameters that define a maximum deviation to the nominal direction vector. Agents with the reactive and reflexive architecture in place display drives, implicit goals (the nominal direction vector of the drive with strongest signal) but no explicit motivation. Any affective qualities are an emergent property of the agent and its environment. Maundering, the switching between two goals or behaviours, is one example. Consider a simple experiment where a predominantly static predatory agent is flanked by two food items of benefit to a further agent. While the energy levels of the agent are high it will move away from the predatory agent. As these drop, the signal strength to move towards a food item increases. At some point the signal strength associated with moving towards the food becomes sufficient that it is adopted as the primary goal.

The agent therefore moves towards one or both of the energy sources (these are not antithetic behaviours). It then moves towards the food items and hence the predatory agent. The move away from predator behaviour is then given a higher signal strength and the agent moves in its original direction away from the predator and hence the food. Typically the behaviour of the agent fluctuates until it runs out of energy or is consumed by the predator. If vector summation is used instead, the agent oscillates around its original position but again gradually moves towards the predator with the same result. This agent is in effect caught and maunders between fear and attraction. The agent can possibly reach the energy items through planning. A reactive planner with precompiled plans will not produce the non-direct path to the required items and will therefore result in the same behaviours at greater computational expense. The agent needs to be able to perform planning that incorporates models of its internal states and external environment. These are deliberative processes and are associated with explicit motivations.

As further parts of architecture are added, explicit motivation processing at the deliberative level allows a more sophisticated processing of the drive model. The signal strength associated with a drive means that calls to explicit motivational constructs for specific drives can have a quantitative (signal) and qualitative (type) component. This allows agents to develop actions based upon their internal agendas which reflect current personae and not as mere reactions to internal and external pressures. Consider the maundering scenario described above. Motivational structures can be instantiated that correspond to the following three drives:

- Flee From Predator with signal strength 0.75
- Move to Energy Source One with signal strength 0.5
- Move to Energy Source Two with signal strength 0.5

A pure reactive agent will either flee from the predator (Signal strength Model) or move towards the mean position of the three entities (Behaviour Sum Model). An agent with deliberative qualities has other possibilities. The full motivator structure is shown in table 1. Not all the components of this computational object need to be instantiated. Table 2 gives examples of this structure as created for the

Flee From Predator drive in the current situation with the initial values given in column 2. Similar structures are created for the other two (Move to Energy Source) drives. The deliberative processes associated with motivation management are initiated by an alarm call from the posting of these structures to the agent's motivator database. In this simple architecture all these motivators are given an initial appraisal, leading to the changes in value as shown in column 3 of table 2. The motivator database is then ordered on the basis of the importance value for each motivator. In some of our design architectures these processes may not be complete, as time and processing limits can be associated with processes internal to an agent. In the simplest motivation architectures the top-most item from the database would be selected and plan set 5, flee adopted as behaviours to be activated. Here a more sophisticated processes look to see how motivators can be combined, in a similar manner to teleoreactive planning (Nilsson 1994). A planning module is used to find plans that combine plansets associated with pair-wise combinations of motivators. The ordering on the motivator database dictates the combination order. Motivators that can be combined are subsumed into the highest motivator with the second deleted. Any change to the database results in further motivator appraisal. Hence the fourth column in table 2 shows the result of combining the Flee Predator with GoTo energy source 1 motivators with a suitable new planset. This motivator is adopted, made active and its plan set pursued as a series of movements in the environment that allow the agent to stay way from the predator but still consume the energy source.

Discussion

Why should an agent perform internal activities (such as problem solving) or carry out behaviours in its external environment if it is truly autonomous? A computational agent can be given tasks but if autonomous must accept those tasks as worthwhile in some sense. The inference is therefore that motivation is the foundation of autonomy. Motivation is highly inter-linked with drives essential to the well-being of an agent. Such drives have affective qualities. The thesis is that affect is the global drive of cognition. Problem-solving only occurs when the problem solver is motivated. Things are only thought about because they are of interest. Affect provides the means by which these processes are valenced and by which affordances are placed on events and activities.

This paper has presented a theoretical model linking cognition and affect through the use of motivational structures. Through the use of a number of analogous perspectives on mind we are building a taxonomy of control states (and similar phenomena) associated with mind. At the theoretical level we consider motivation to provide the key to integrating these control states with current analyses on the nature of consciousness. Computational experiments using the agent-based approach are used to investigate the theoretical and design models. In short autonomy can be defined in terms of motivation. An intelligent agent (whether biological or synthetic) should feel no compunction to perform any external or internal activity unless motivated. Our experiments into this and the nature of emergent states continues.

References

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

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

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

Bickhard, M.H. and Terveen, L (1995), Foundational Issues in Artificial Intelligence and Cognitive Science, Elsevier.

Brooks , R. A. (1999), Cambrian intelligence : the early history of the new AI. Cambridge, Mass. MIT Press.

Davis, D.N. (1997), Reactive and motivational agents. In: *Intelligent Agents III: Agent Theory, Language and Architectures*, Springer-Verlag.

Davis, D.N. (2000a), Agents, Emergence, Emotion and Representation, *IEEE IECON2000, Nagoya.*

Davis, D.N. (2000b), Agent-based decision support framework for water supply infrastructure rehabilitation and development, *International Journal of Computers, Environment and Urban Systems*, 24, 1-18.

Davis, D.N. (2001a), Multiple Level Representation of Emotion in Computational Agents, *Emotion, Cognition & Affective Computing, AISB2001-Agents and Cognition, University of York.*

Davis, D.N. (2001b), Control States and Complete Agent Architectures, *Computational Intelligence, Vol. 17, No. 4.*

Duffy, E. (1962), *Activation and Behavior*, Wiley & Sons.

Epstein, J.M. and Axtell, R. (1996), *Growing Artificial Societies*, MIT Press.

Eysenck, M.W. (1990), *Blackwell Dictionary of Cognitive Psychology*, Blackwell Publishers, Oxford.

Franklin, S.P. (1997), Autonomous Agents as Embodied AI, *Cybernetics and Systems, Vol. 28, No. 6.*

Greeno, G. (1978), Natures of problem solving abilities. In: W.K. Estes (Ed*). Handbook of Learning and Cognitive Processes. (Vol. 5),* Erlbaum.

Hanks, S., Pollack, M.E. & Cohen, P.R. (1993), Benchmarks, test-beds, controlled experimentation, and the design of agent architectures, *AI Magazine, 14(4):17-42.*

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).*

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

Kaebling L.P. (1989), An architecture for intelligent reactive systems, *Readings in Planning,* Morgan Kaufmann.

Kosslyn, S.M. (1975), Information representation in visual images, *Cognitive Pyscholology, 7, 341-370.*

Luo, Y., Davis, D.N. and Liu, K. (2001), Combining KADS with Zeus to develop a multi-agent E-commerce application, *International Journal of Electronic Commerce Research*, Kluwer, (To Appear)

Merleau-Ponty, M. (1942), *The Structure of Behaviour,* Methuan, London.

Newell, A. and Simon, H.A. (1972) , *Human Problem Solving.* Prentice-Hall.

Nilsson, N.J. (1984), Teleo-reactive programs for agent control, *Journal of Artificial Intelligence Research* 1:139-158.

Polya, G. (1945), *How to Solve It.* Princeton University Press.

Rolls, E.T. (1999), *The Brain and Emotion*, Oxford University Press.

Simon, H.A. (1979), Motivational and emotional controls of cognition, In: *Models of Thought*, Yale University Press.

Sloman, A. (1987), Motives mechanisms and emotions. *Emotion and Cognition, 1(3):217—234.*. Reprinted in M.A. Boden (ed), *The Philosophy of Artificial Intelligence*, `Oxford Readings in Philosophy' Series, Oxford University Press, 231--247, 1990.

Sloman, A. (1993), The mind as a control system, In: *Philosophy and the Cognitive Sciences, C. Hookway and D. Peterson (Editors)*, Cambridge University Press.

Wheeler, M. (!997), Cognition's coming home: The reunion of life and mind, In: *Husbands, P. and Harvey, I. (Eds.) Fourth European Conference on Artificial Life*, MIT Press, pp 10-19.

Wollheim, R. (2000), *On The Emotions*, Yale University Press.

Castelfranchi, C. Guarantees for autonomy in cognitive agent architectures. Wooldridge, M. and N.R. Jennings (Eds), *Intelligent Agents.* Springer-Verlag, 1995: 56-70.

Davis, D.N., Sloman, A. and Poli, R., Simulating agents and their environments. *AISB Quarterly, 1995*

Davis, D.N., Synthetic Agents: Synthetic Minds? *Frontiers in Cognitive Agents, IEEE Symposium on Systems, Man and Cybernetics,* San Diego, 1998.

Davis, D.N., Minds have personalities - Emotion is the core, *AISB2000, University of Birmingham, 2000*

Ferber, J. *Multi-Agent Systems,* Addison-Wesley, 1999

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

Husbands P., I. Harvey, and D. Cliff, An evolutionary approach to situated AI. In *A.Sloman, D.Hogg, G. Humphreys, D. Partridge, A. Ramsay, Prospects for Artificial Intelligence, IOS Press, pp 61-70, 1993.*

Moffat D., R.H.. Phaf, and N. Frijda, Analysis of a model of emotions, In: *A.Sloman, D.Hogg, G. Humphreys, D. Partridge, A. Ramsay (eds.) Prospects for Artificial Intelligence, IOS Press, 219-228, 1994.*

Nilsson, N, Shakey the Robot, *Technical Note 323, SRI Internal, Menlo Park, CA, 1984.*

Oatley, K. and Jenkins, J.M, *Understanding Emotions, Blackwell Publishers, 1996.*

Ortony, A., G.L. Clore and A. Collins, *The Cognitive Structure of Emotions. Cambridge University Press, 1988.*

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

Scherer, K.R., Studying the emotion-antecedent appraisal process: An expert system approach, Cognition and Emotion, Vol: 7, 325-355, 1993.

Sloman, A. and Poli, R., SIM_AGENT: A toolkit for exploring agent designs, In: *Intelligent Agents Vol II (ATAL-95), Eds. Mike Wooldridge, Joerg Mueller, Milind Tambe, Springer-Verlag 1996 :392-407*.

Sloman, A. and Logan, B., Cognition and Affect: Architectures and Tools, *European Conference on Cognitive Modeling, 1998.*

Velásquez, J.D., Modelling emotions and other motivations in synthetic agents, In: *AAI97:10-15 1997.*

Wright, I., Sloman, A. and Beaudoin, L., Towards a Design-Based Analysis of Emotional Episodes, *Philosophy Psychiatry and Psychology, Vol. 3 no 2, 1996.*

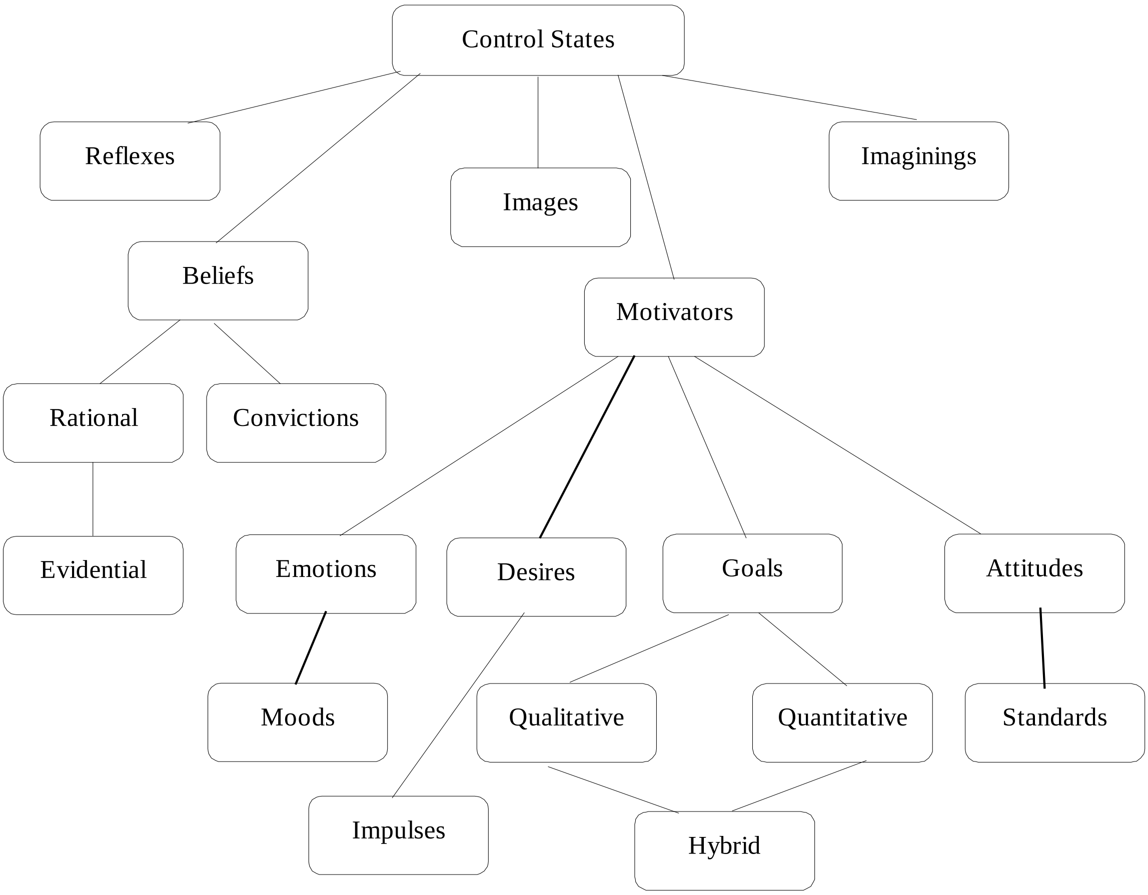

Figure 1. Taxonomy of control states based on Sloman (1993).

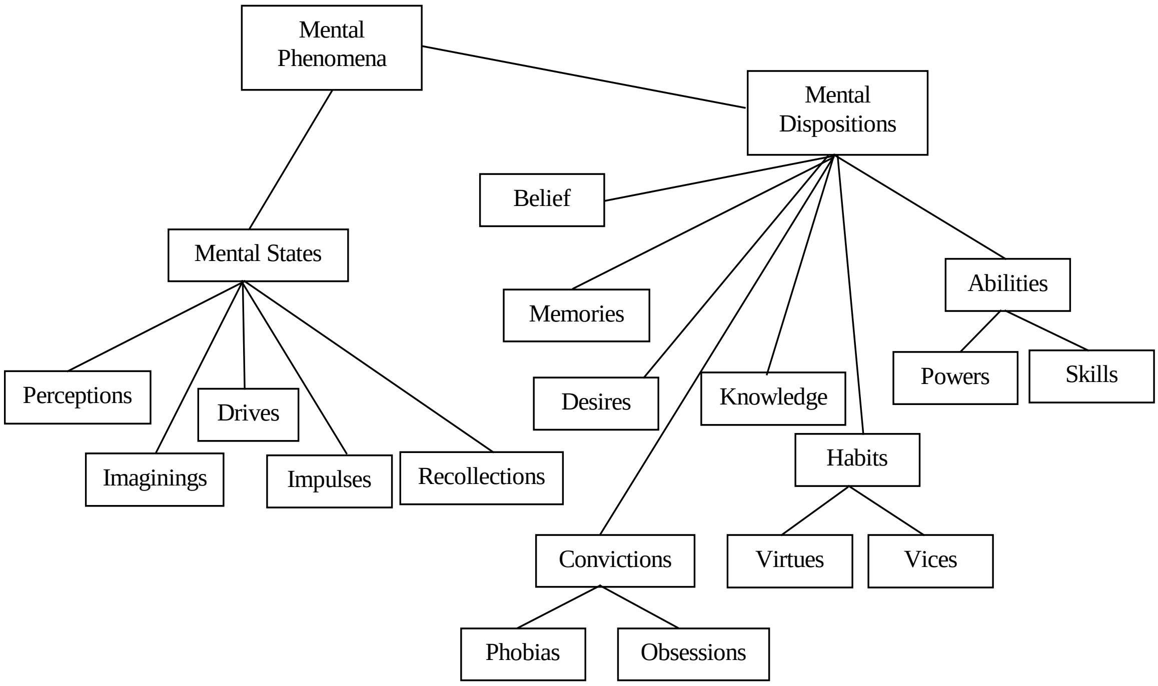

Figure 2. Taxonomy of Mental Phenomena Based on Wollheim 1999.

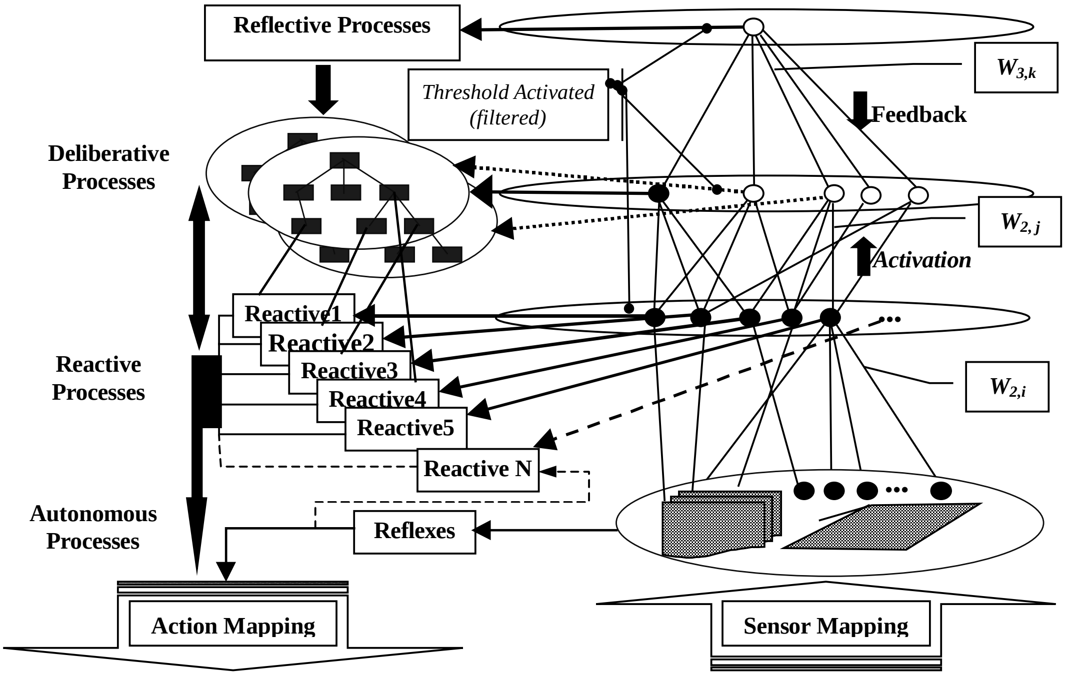

Figure 3: Design sketch for an integrated architecture for synthetic mind.

IF E < ET_0 *THEN* 0 → $Signal_{Hunger}$ (Dead)

IF E < ET_1 *THEN* 1 → $Signal_{Hunger}$

IF E < ET_2 *THEN* $(ET_2\text{-}E/ET_2\text{–}ET_1)$ → $Signal_{Hunger}$

IF E > ET_2 *THEN* 0 → $Signal_{Hunger}$ (Satiation)

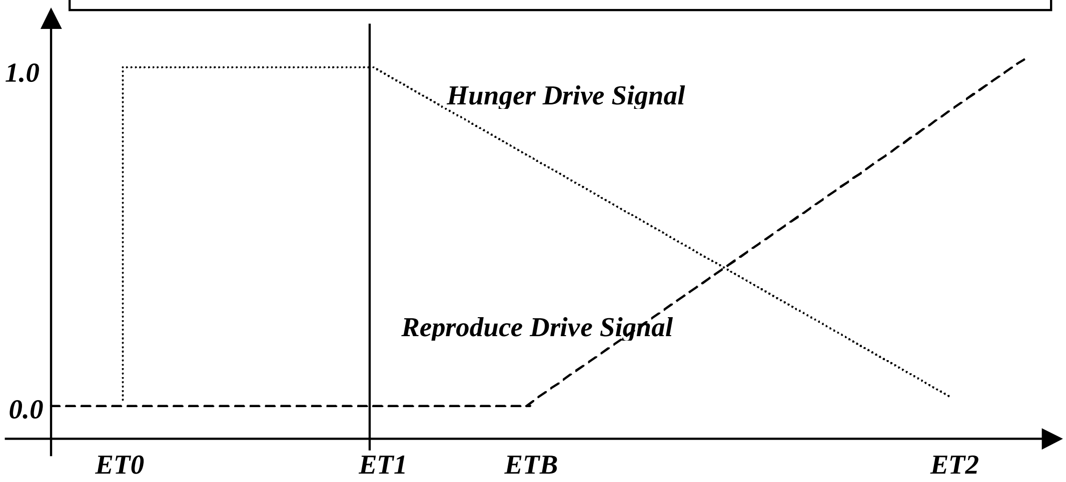

Figure 4. Model for internal energy based drives. The four energy thresholds are subject to experimentation and could be optimised using for example genetic operators.

Table 1: The components associated with motivator structures.

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

Table 2: The instantiated motivator structure for Flee Predator.

Component	Value (initial)	Value (appraisal 1)	Value (appraisal 2)
Actors	Predator1	Predator1	Predator1, Food1
Belief Indicator	True	True	True
Status	Unconsidered	Considered	Adopted, Merged
Decay Function	Null	Null	Null
Dynamic State	Null	Postponed	Active
Emotion	Fear	Fear	Fear, Hunger
Importance Value	Null	High	High
Insistence Value	0.75	0.75	0.75
Intensity	Null	Medium	Medium
Management	Null	Null	Combined(M1,M2)
Attitude	Keep true	Keep true	Keep true
Plan Set	Null	Null	Plan(set10, bypass)
Rationale	Drive:Fear	Drive:Fear	Drive:Fear, Drive:Hunger
Content	Flee(Predator1,X,Y)	Flee(Predator1,X,Y)	Flee(Predator1,X,Y) GoTo(Food1,Xf,Yf)
Urgency	Null	Null	High