

Multiple Level Representation of Emotion in Computational Agents

Darryl N. Davis¹

NEAT Group, Department of Computer Science, University of Hull, HU6 7RX.

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

Abstract

This paper describes current work from an ongoing investigation into the computational modelling of emotion and motivation. The impetus for this research can be found in the work on the computational modelling of motivation. The impetus for that in turn is the control state approach to modelling cognition. Recent research has focussed on the use of cellular automata to provide a foundation for an emotion engine. This emotion engine is to be placed at the core of an agent architecture. Earlier work used simple tokens, for example fuzzy-valued symbols such as angry or very angry, in motivational structures to act as referents to specific emotional states. Having revisited concepts underlying the nature of agency, a number of simple multi-agent scenarios are being investigated as experimental frameworks for developing the theory, designs and implementations.

1 Introduction

The nature of our research, and the areas that need to be addressed, is informed through the use of philosophical and psychological perspectives on natural and synthetic minds. These perspectives and their associated stances are structured through the use of niche space and design space (Sloman 1993). This paper presents the current state of our research into using emotion as the core to control states (Simon 1979; Sloman 1993) and synthetic minds. We are pursuing a line that links motivational states (for example impulses, drives, goals) and emotion. Earlier 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. Various synthetic harnesses are being used as experimental vehicles for the investigation and development of theory, design and implementations. These include Tileworld (Hanks et al 1993), simulated robo-cup (Robo-Cup 2001) and predator-prey scenarios using a-life techniques. It is a variety of the latter that is used as a descriptive vehicle in this paper.

2 Niche Space and Design Space

Niche and design spaces are discontinuous, multi-level and interacting. Niche space consists, in part, of

collections of requirements and constraints. The study of niche space provides a means to facilitate the investigation of collections of requirements and constraints that define the nature of interactions between an agent and its domain(s). They provide the focus from which to investigate alternative designs for systems that could fulfil specific functions. Hence, it is possible to define a niche point for an agent that predates on agents inhabiting alternative niche points. Points in niche space are cogent collections of requirements. Two organisms may share the same external environments but inhabit different niches. For instance in the synthetic predator-prey scenarios discussed in this paper, both classes of agents exist in the same synthetic environments. Requirements for navigating the uninhabited environment are effectively the same. The drives that compel these agents to explore different parts of the environment differ. Prey agents direct their behaviour to flee predators and seek vegetation. Predator agents seek both vegetation and prey agents. Hence, the niche spaces for these two agents overlap but differ. On the other hand design space is a collection of more or less specific methodologies, formalisms, architectures, mechanisms and algorithms. Hence, design space provides a means for the study and analysis of alternative and competing design possibilities for niche spaces. Example alternatives in design space include classical planning robots such as Shakey (Nilsson

¹ Two of my research students, Suzanne Lewis and Humberto Nunes, for their jargon-bursting efforts.

1984); pure behaviour-based agents such as used in the Pengi experiments (Agre and Chapman 1987); and evolvable agents (Husbands et al 1993). Mappings between these two spaces represent more or less well-fitted matches between requirements and designs. Any specific agent design fulfils the requirements and constraints of one specific niche point to some degree of efficacy. Furthermore, any specific agent design may function, with greater or lesser efficacy, in more than one niche space. A dragonfly is an effective insect predator. A mayfly fulfils a specific insect niche very effectively but makes an ineffective insect predator.

A specific design space is used for the agents on which the predator-prey agents are based. This design space is that of rule and object based synthetic agents in the SIM_AGENT toolkit (Davis et al 1995). The niche spaces associated with this toolkit are those of synthetic agents in simulated environments. Not all design spaces associated with synthetic agents are used in the toolkits base class agents. Only when we start to consider more sophisticated versions of these agents that include machine perception, neural networks (Sloman and Poli 1996), filter-protected architectures (Sloman and Logan 1998) and emotion-based autonomy (Davis 2000) are other parts of design space used. The design spaces used in these experiments are incrementally expanded to arrive at sophisticated agent architectures in a manner similar to the Sugarscape experiments (Epstein and Axtell 1998).

3 Motivation and Emotion

Merleau-Ponty (1942) considered that humans are moved to action by disequilibria between self and the world. If a descriptive model can be provided for these disequilibria then it may be possible to use it as a design framework for synthetic minds. The phenomenological approach to the analysis of human behaviour implies a distinction between an agent's internal and external environments. For any specific agent, there may be no differentiation in the extent to which either of these is real. One implication is that reasoning and behaviour is activated by a need expressed in terms of valenced descriptors for drives, concerns and goals. 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. Drives, concerns and goals are specific forms of motivational control states. Other terms exist for motivational control states, for example desires and attitudes. For the sake of clarity, this paper will be confined to just these three. It is possible to differentiate between them on the basis of their processing and control requirements. 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 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 can coexist and interact but never be identical. An in-depth analysis of this differentiation is outside of the scope of this paper. The niche spaces for conscious, pre-conscious and subconscious mental states may overlap but are disparate.

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 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. A predatory agent may have drives to hunt prey agents. 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.

There exist models for the emotions, for example (Moffat 1993), that easily relate to some current computational models of cognition and afford descriptors at multiple levels for motivation and other control states. In short emotion and cognition are highly inter-linked. 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. There is therefore a

disagreement here with for example Plato, who degrades emotion as distorting rationality, and Darwin who considers emotions in adult humans to be a by-product of evolutionary history and personal development. Modern psychology and philosophy of the mind have taken a more positive stance on emotion. Artificial intelligence, cognitive science and popular science similarly so (Picard 1997). Numerous computational systems now include emotions, whether as a reasoning system (Scherer 1993), a connectionist emotion synthesiser (Velásquez 1997) or a low-level control process (Frankel and Ray 2000). There is however little agreement across these fields on a definition of emotion. 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 be 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. 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?

Consider three types of psychological theory of the emotions. The OCC model (Ortony et al 1988) is defined in terms of valence (arousal) and appraisal with primary emotions and further subtypes defined in terms of events, objects and agents. The goal-based model of the emotions, for example (Oatley and Jenkins 1996), also makes use of events, objects and agents but relates these external phenomena to internal environments of goals and roles. This leads to a definition of a small number of basic emotions, such as anger, fear, surprise etc., in terms of these internal and external phenomena. A third approach is to consider the functional role of emotions and how they reinforce internal and external behaviours and relate to motivation and memory (Rolls 1999). Again in these theories emotion is described across multiple and distinct processing levels.

Consider two philosophical approaches to the emotions. Sloman (1998) differentiates between three classes of emotion and relates these to a three-tier three-column architecture of the mind. Primary emotions arise from reactive level processing. Secondary emotions arise from deliberative processing, while tertiary or perturbant emotions require reflective processing to manage them as potential out-of-control mind states. Wollheim (1999) in

his analysis of mental phenomena differentiates between mental states and mental dispositions in terms of intentionality, subjectivity, and consciousness. It is in highlighting the difference between predominantly conscious mental states and pre or unconscious mental dispositions that Wollheim provides an analysis which results in distinguishing conscious emotional states which are related but different to the emotions per-se that exist as pre-conscious and subconscious mental dispositions.

For the purpose of developing computational models that draw on this psychological and philosophical research a tentative definition of emotion is needed. Emotion can be defined as a quality of life arising from valenced expectations and reactions to events in the world and an agent's perceptions of those events and other related categories of cognitive acts. The current work at the implementation level falls short of such definitions. We are using minimal agent designs to investigate the implications of pursuing this line of thought. The objective is to determine how useful the metaphor of emotion is in thinking about and designing computational agents. An agent has valenced expectations when pursuing co-operative goals, for example, "if agent X performs action Y with material Z, then this is good as work necessary for goal W is then accomplished". Perceptions of and reactions to events (e.g. that was good, or that was bad) relate to some form of differentiation (whether implicit, explicit, conscious or preconscious) between an agent's model of its external environment, the external environment itself and the agent's wholly internal environment. Such differentiation gives rise to disequilibria and appraisal scenarios – in terms of motivations. It is the entire gestalt of this processing that we consider to be the emotive state of the agent in relation to specific goals, events, external objects and other agents. An emotion is not a representation or a signal but a multi-layer multiple representational valenced set of processes. In the synthetic predator-prey scenario the focus of such a dynamic processing framework is evident; in more sophisticated architectures and biological agents this may not be so.

4 Mind Processes and Motivation

It is possible to classify mental processes and phenomena as transient, mediating or permanent. Such categorisations are relative to the lifetime of an agent and/or the perceived implications of an agent's actions over its lifetime. Examples of transient phenomena include impulses, reflexes and similar short-lived events. Emotional states fit within this category of temporal mental events. Mediating phenomena are non-

instantaneous, more enduring phenomena such as goals and concerns. Mental states and dispositions are associated with these. Permanence, however illusory, relates to long-term dispositions and motivations. Long-term motivations are associated with dispositions but can rise to many mental (control) states, which may be emotionally valenced in a number of ways. The dominance of motivations over these time-spans need not be temporarily ordered. Transient events can affect long-term motivations, and vice-versa. A highly valenced emotional state gives rise to a long-term motivation, which may give rise to further emotional states. The originating emotional state however is transient. In considering how to design cognitive agents with affective capabilities, particularly where these are highly inter-linked with motivations, an analysis of the temporal nature of mental phenomena is necessary.

Other important criteria in the design of synthetic minds with motivations and emotions arise from how an agent can consider its actual, possible and desired niche(s). In designing an agent for a specific environment, an agent's natural (sic) niche is considered. Any particular agent design is capable of moving between and performing tasks (to certain levels of competency) in neighbouring environments. In some environments an agent will flourish, in others perish. In moving between environments an agent need not change niche space. A mobile computational agent can inhabit different computational platforms performing identical tasks. Such an agent is still attempting to fulfil its original niche space, even though its environment may have changed. However it is possible to consider an agent that inhabits one external environment but modifies its design space. To fulfil its functional criteria such an agent may need to modify its design as its niche space is redefined. Some natural (biological) agents do this by transforming their physiology. Certain tropical fish modify their physiology to change gender when their (local) population requires it. Certain synthetic agents can modify their particular instantiation in design space to meet the niche criteria of specific environments - evolvable and adaptable agents for example. Other agents in moving to different environments modify their design to fulfil criteria associated with a different (typically) closely related niche space. A caterpillar in transforming to a moth changes design, environment and niche spaces. The goals or motivations that drive these changes are implicit in the design of these agents. It is unlikely that any of these agents use explicit motivations to transform their location in design or niche space.

Realistic motivations and goals enable an agent to adapt and survive as their environment changes. Dependent

upon the design and nature of the agent, a motivated move to other possible niche spaces may also be realistic. If an agent's environment is transformed dramatically, the only possible solution for its survival is to transform its design or niche space. The achievement of these types of goals allows an agent to realise harmonious affective states. The failure to achieve such goals provides cause for an agent to experience antithetic affective states. This applies whether the environment is external or internal to the agent. In sophisticated (cognitive?) agents, the transformation of an external environment may cause serious problems for the agent in managing its internal environment. Such states may be sufficiently valenced for the generation of further motivators. Learning to cope with, for example grief (Wright et al 1996), requires a change in a (human) agent's internal environment and the way information and events are thought about (a design space change?). This can precipitate a change in niche space as the (human) agent determines to change their life. Not all desired niche or design space changes are achievable. The nature of what is achievable may change over an agent's lifetime. This can give rise to perturbant agent states. If sophisticated human-like synthetic cognitive agents are to be designed then such effects need to be included in the design brief.

Perturbant states can occur in even the simplest computational agent system. For example, a behaviour-based agent can maunder - turning left then right so it effectively circles around some positional locus. If synthetic agents capable of responding to the affective state of their human interactors are to be designed, a similar brief is again required. As design complexity increases the possibility of perturbant states arising increases. If sophisticated agents capable of cognitive acts are to be designed, such effects need to be included in the design brief. It has been suggested elsewhere that a reflective (Sloman 1993) or meta-level control (Ferber 1999) layer is required to recognise and control such situations. The perspective offered by the current research is that such mechanisms are insufficient. Maundering and perturbation occur because an agent has autonomy. If the descriptive framework used in sophisticated agent designs changes over the separate processing layers of an architecture, internal consistency is jeopardised. Through incorporating an analogue to emotion as the basis for an agent's autonomy a consistent framework is ensured. A dynamic model of emotion embodied within the control architecture of a sophisticated agent (or agent community) may allow perturbant affective states to be recognised or if inherent in the agent system mediated in terms of temporal processing. This however means that emotion must be modelled as a multiple-level

representation, using processing structures appropriate to the category of processing performed at any specific level. If emotions are to be used to valence events, expectations and decision making at a deliberative level, then some form of equivalent processing structure must be used in the remaining levels of the agent architecture.

To move beyond this type of analysis we need to consider the architectural structures and mechanisms involved in the production and processing of motivators and similar (intentional) phenomenological states, and whether they are linked to emotions. And if so how deep is this type of link? We have been developing agent architectures for the heterogeneous modelling of a number of mental phenomena using concepts such as control states (Davis 1997, 2001). This development now places emotion at the core of cognitive processing, providing multi-dimensional valences at a number of representational levels that underpin motivation. Such architectures consider emotion as a multi-layer multi-faceted descriptor that informs learning and motivational processes. Emotion should therefore form the basis for agent autonomy and subsequently inform processes for motivation and learning through a synthesis of processes across different representational frameworks in a single architecture.

5 Architectures and Experiments

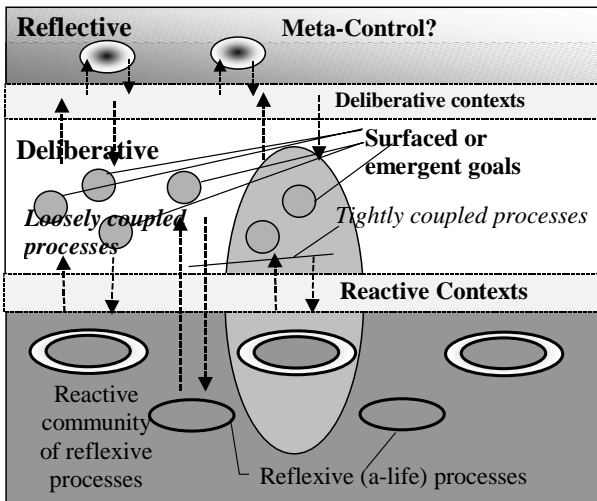


Figure 1. Abstract architecture sketched in terms of goals

Consider an agent architecture that models motivations in terms of drives, concerns and goals. Figure 1 sketches this architecture in its most abstract and distributed form. Three layers are heavily delineated. A reflective layer is responsible for meta-control. In effect this layer navigates the agent as a whole through the types of spaces it can inhabit. The deliberative layer consists of processes that

manipulate explicit representations about current and possible spaces that the agent can inhabit. The third layer contains two categories of processes. Reactive processes that respond to events (internal and external) in the agent's current space, and reflexive processes that run automatically regardless of the agent's space. The behaviour of the reflexive processes and what form the current reactive processes depends upon the agent as a whole. At any particular instant processes across the agent architecture can active or dormant. Separating these three layers are contexts or filters for control and information messages. Processes can be loosely or tightly coupled across the lower two (and a half) layers of the deliberative and reactive (plus reflexive) levels. Processes become tightly coupled as they process information on a shared context. When that context changes, or becomes defunct, tightly coupled processes become uncoupled. The now loosely coupled processes may subsequently become dormant at the deliberative and reactive layers.

5.1 The emotion engine

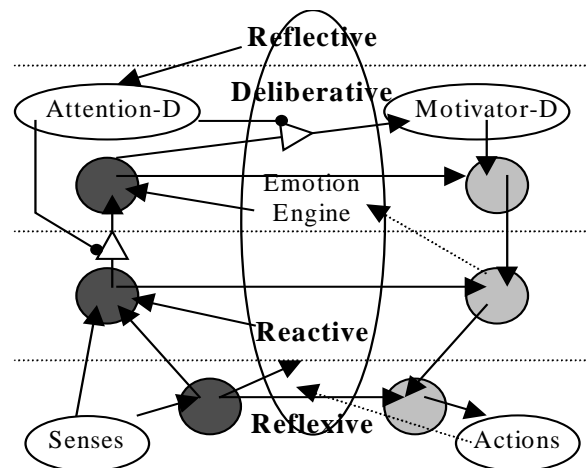


Figure 2. A four-layer architecture with the emotion engine as the core. Dark grey circles represent information assimilation. Light grey circles represent mappings onto internal and external behaviours. White triangles are filters. Dashed lines represent feedback.

The emotion engine driven architecture shown in figure 2 is one instantiation of the design sketched in figure 1. This architecture with a central core of tightly-coupled filter-activated processes was initially implemented as a domain-less test-bed. Experimentation with this domain-less model addressed the nature of control across the different levels of the emotion engine, and is described in several places (Davis 2000a 2000b). Here the design is revisited in terms of the preceding sections, and one of a number of domain experiments. The emotion engine is

being developed to support expressive graphical avatars, multiple agent co-operation models, interactive entertainment and predator-prey game playing scenarios. The emotion engine is a data and information processing infrastructure designed to support multiple-level representations of emotion. It is designed around an analysis and comparison of emotion in human agents and autonomy in insect and computational agents. Elements of this analysis have been introduced above. The emotion engine makes use of a set of filter-protected processes across the layers of the agent architecture. Only the reflexive processes, represented as one or more communities of cellular automata, run continuously. The various layers communicate in terms of their underlying emotional descriptors. This communication is filter protected both hierarchically and laterally. The communication of an emotive state, at or across whatever levels, can affect the entire architecture of the agent. Alternative design and implementation experiments are proceeding which provide a loose and much tighter couplings of the processes associated with this communication of valenced information.

A variety of homogeneous and heterogeneous cellular automata based processes have been used to model the autonomic basis for four emotions (anger, disgust, fear, and sobriety). Here sobriety is used to mean a moderate state with extremes related to sadness and happiness and has little or nothing to do with alcohol! In an initial set of experiments a heterogeneous cellular automata model was used. All four emotions were modelled in one community using four varieties of cellular automata and an insect hive metaphor. This model is very efficient but also volatile in terms of its response to small changes in input. Subsequent investigations found that the use of four separate communities, one for each emotion, of homogeneous cellular automata gave a more predictable control model. It is this latter model that is used here. Further experiments do not model these emotions at the reflexive layer but use them to provide reactive valences for sensory information. Each cell can take a value in the range $\{-1,1\}$ and communicates within (a three-dimensional extension to) a Von Neumann neighbourhood. We have so far experimented with integer values in this range. Future work will investigate real-valued cell states. The valence of the automata communities provides an initialisation of the processing events that give rise to the generation motivators at a more sophisticated representational level. The emotional descriptors are vector signal strengths based on the summed valence (and summed absolute valence) of each cell in a hive. Hence emotional descriptors are initiated at the lowest level in the architecture, mirroring the values

of automata communities and/or individual automata. In a fully developed architecture, emotional descriptors may be initiated at a number of sources across the architecture. Where the vector meets the constraints of the protective filters, reactive or deliberative processes are activated in response. Where these motivators give rise to appraisal and deliberative processing complex information processing structures can arise. These can be communicated between co-operating agents wholesale as deliberative contexts (Davis 1998). Once initiated control processes assess the valence of the reflexive communities in terms of current motivations and concerns. If response strength of the reflexive communities is overly strong, the sensitivity of the communities can be reduced. In addition to this decision, the deliberative processes may also provoke a reappraisal of currently active motivators (i.e. goals). In situations where the sensitivity of the reflexive communities has already been changed and they are still responding inappropriately, the reflexive processes are activated. As a result the agent may change its preference for which part of emotion space is currently inhabits. This means that the initial patterns on the automata cells change, that the context and values on filters are modified and the deliberative processes are given an alternative preference order on the motivations they manage. In this architecture, explicit motivations are responses to reactive processes acting on an agent's internal representation of its world and preferred niche stance. The latter can change in response to the activation of the reflexive layer of the emotion engine.

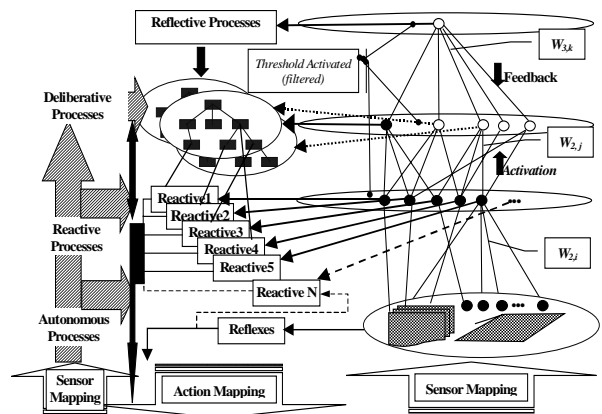


Figure 3. The full emotion engine.

Figure 3 paints a more complex version of the emotion engine. This architecture follows on from the sketch given in figure 2 and the initial experiments with the emotion engine. Three information pathways are highlighted in this architecture. One based on the emotion engine with similar information processing strategies to those just described and the other a hierarchical synthesis

of agent sensory information and motivational information. The latter is similar in kind to earlier reactive and motivational agents (Davis 1996). The third pathway relates to the mapping of intentions onto actions in the agent's environment. This requires that the agent be situated in some domain. Simple versions of this architecture are being investigated in predator-prey scenarios, using an incremental design approach.

In this architecture, internal autonomy is based on the reflexive aspects of the emotion engine (the cellular automata) and the reactive motivational level as responses to sensory processing. Processes at the reactive and deliberative are activated by means of connections over a hybrid, filter based architecture. Permutations of valenced automata at the base level are distributively connected to higher levels. After the initial experiments the plan is to investigate neuro-connectionist models. As cells at the various vertical layers in the emotion-engine become valenced they activate connections to other cells. As a cell (or combinations of cells) becomes activated they also provoke behaviours at that level. The information content of these behaviours is largely domain dependent and can also be activated by hierarchical pathways on the left of figure 3. The nature of the lateral activation requires investigation, particularly as the left and right-hand pathways may at times be competitive. Emotion in this architecture is a multi-layer multi-faceted descriptor used to inform domain dependent and control processes. Emotion therefore forms the basis for agent autonomy and subsequently informs processes for motivation (and learning) through a synthesis of heterogeneous processes across different representational frameworks in a single architecture.

5.2 Scentworld : a predator prey scenario

Scentworld is a test-harness being developed in SIM_AGENT for the experimentation of implicit and explicit motivational states and computational models of autonomy. Autonomy (defined as operating "...without the direct intervention of humans or others, and have some kind of control over their actions and internal state") is one of four foundations for a weak notion of agency (Ferber 1999). Castelfranchi (1975) discusses the various types of autonomy that an agent or robot can demonstrate, and in particular draws a distinction between belief and goal autonomy. If an agent is to be autonomous, then it must set and have control over its goals. These goals can be simple (e.g. switch process P_a on) or complex (e.g. maintain optimum process management as resources become scarce) but in essence define the current computational focus of the system.

Other (external) agents, whether biological or computational, should not have direct control over the setting of an agent's goals. They can only influence these through the provision of information that may affect an agent's belief set. Ferber (1999) associates autonomy with "a set of tendencies, which can take the form of individual goals to be achieved" (pp9-10).

The simplest perspective on this experimentation into agent spaces is to view it as predator-prey game that permits incremental sophistication. The simplest agents are no more complicated than the regulatory agents used for example in the Pengi type of experiments. A regulatory agent is defined as an integrated (computational) entity with intentionality and some degree of autonomy (Davis 2001). Such agents provide a base class (stage1) capable of some function in a specific synthetic environment. For example we can utilise a specific niche space associated with regulatory agents and produce a base class of agents that navigate around a synthetic environment on the basis of simulated vision. To survive these agents need to find food using two senses. They can use simulated vision to detect food objects within their visual range. Such objects provide large energy packets but are in short supply and are competitively sought by other agents. Smaller packets of energy are also available. These are scattered across an image that the agents can sample. The sample size is variable but tends to be small and local to the agent. Processing of this image sample allows the agent to directly detect and consume the small energy packets. These energy packets can also be viewed as some form of scent or trail. An agent uses its energy as it moves. This is left as a pixel trail in the image – the more energy consumed over an epoch the larger the trail. The two sense types are used to inform behaviour selection process by activating (Do Move-Ahead) or inhibiting (Not Move-Ahead) specific behaviours. The agents make use of an internal behaviour (defined as a set of rules) that then decides between the remaining competitive behaviours (such as *Go-Right*, *Go-Left*, *Stop*, *Turn-180°* etc.). An agent-oriented programming paradigm, making use of a combination of object and rules, is used to implement these designs. Other possible designs make use of neural networks and the agents then learn to make use of their classifier systems. The implementation technique for stage1 agents is mainly irrelevant.

Two further but alternative niche spaces can then be defined as refinements of the niche space for the base class of agent. The first niche space closely resembles the original niche space but defines what types of energy packets or trails can be consumed and what

environmental figures agents to follow or flee. A further niche space defines the omnivorous predator. These agents survive on different categories of energy packets, pursue prey scent trails and consume prey agents. Predators prefer not to follow other predator trails, but do not actively flee other predators. No explicit goals structures are necessary to define the design space for these agents. The drives and concerns of these agents can be modelled as rule-definable behaviour sets. The agent determines what is the most appropriate current behaviour on the basis of what behaviours are currently disallowed and then the weighted sum of the remainder. These designs result in stage2 agents. The different stage 2 agents make use of similar agent architectures but the information content of their left-hand pathway (in figure 3) differs.

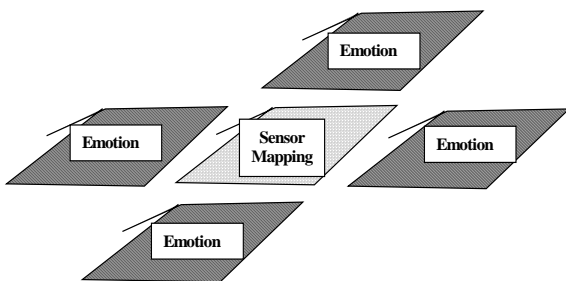


Figure 4. Plus sign topology. Subtypes have the emotion communities interchanged

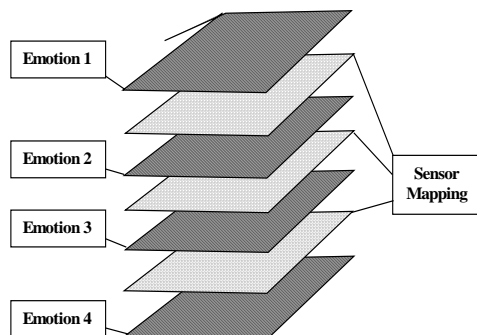


Figure 5. Wafer topology – the connections between the successive layers depend on the nature of the wafer

The stage2 agents are then extended to include internal models of “emotional” autonomy, and so define the stage3 agents. The niche space for stage3 agents does not differ from the predator and prey stage2 agents. Their design space does. In current experiments this gives rise to multiple descriptors for autonomy at the reflexive and reactive layers. The agent needs to be able to decide between alternative actions across these multiple levels. This is particularly so where conflicts occur in behaviour selection between the more orthodox reactive behaviours of the base class processes (as described above for the

stage1 agents) and the behaviours selected on the basis of valenced emotive states. The computational model used in the domain-less emotion engine has been extended with machine perception processes (as shown in figure 3) that permit an affordance interpretation of the agent’s environment. Such interpretations are mapped onto the same internal structure as used to represent the agent’s internal low-level automatic states that underpin the emotion engine. Communication between these agent-internal processes providing these multiple descriptors is possible. This communication tends to be in terms of extended computational processing analogous in type to the processing performed at any layer.

The niche space for each agent defines how any specific pixel in an image sample (a combination of red, green and blue image values) is interpreted. For a prey agent a predator trail pixel is negatively valenced, for a predator a similar trail is neutrally valenced. Prey trails are positively valenced for both. The image sample is thus mapped onto automata community, and provides the initial valences for all cells in that community. This percept processing unit provides input to the four communities representing the four modelled emotions. The topology of the extended model can be changed by the user in experimentation or by the agent to change the nature of the processing performed at the reflexive level. The default topology resembles a plus sign with the percept community at the centre; various subtypes are possible (see figures 4 and 5). A fully connected wafer, a border connected wafer and convoluted wafer topologies have also been used in experiments. Plus sign topologies are essentially two-dimensional processing models. Wafer topologies are three-dimensional processing models. Wafer topologies sandwich the percept community between the four emotion communities. A fully connected wafer has each cell connected vertically to the geometrically equivalent cell in the adjoining communities. A border connected wafer has just the outer cells of each community connected to adjoining communities. Convoluted wafers have the left border cells connected to the right border cells of the adjoining communities, and the right border cells connected to the left border cells of the adjoining communities; with the top and bottom cells similarly convoluted. Experimentation has shown that the different topologies are distinct in terms of their level of sensitivity to differently valenced input from the sensor mapping community. Control mechanisms activated by the reactive and deliberative levels can change the reflexive topologies according to the needs of the agent. This minimal model of the right hand side of figure 3 includes the reflexive or automatic level and the control

mechanism that increase and decrease the sensitivity of the automata community processing.

The left-hand and right-hand pathways converge at the deliberative level in stage3 agents with the two pathways providing information about beliefs (left pathway) and emotive valence (right pathway). Where highly valenced emotive vectors arrive and no beliefs about motivationally important figures or events are present, the reflexive processes are desensitised. Where belief predicates contain motivationally important information but no emotive vectors are present the reflexive communities have their sensitivity increased. Where the content of emotive vectors is inappropriate to the content of the belief predicates and there are no other indications of perturbant control states the topology of the reflexive communities is changed. The simplest possible change is to reconfigure the geometric nature of a plus sign topology so the four emotion communities are more appropriately placed. Experimentation continues into the control state significance of the topologies and community sensitivity models. In any case where emotive vectors are available at the deliberative level, they provide valence qualities for use in motivational processing at that level. They can be used in determining the importance of a motivator as defined in earlier work (Davis 1996) or to valence beliefs and store them as memories. The information arriving at this deliberative stage provides different interpretations of the same events in the agent's external environment. These different interpretations can be in agreement. If this agreed interpretation is at odds with the agent's current motivations, the agent will be required to re-evaluate these motivations. If this causes no change, then the agent will not suffer any control state perturbations. If this is not the case and motivators are changed then the agent may begin to experience perturbations in its motivation control states. When the information arriving from the two sources is not in agreement, similar motivation control states can ensue. Stage3 agents have no reflective layer, and so no means of controlling these perturbations is possible according to the design. Further experimentation is required before the fullness of the architecture shown in figure 3 is designed and implemented for this domain.

6 Discussion

This research addresses questions about situated action arguments and the use of artificial life mechanisms in cognitive science (Wheeler 1997). This paper has skirted over an analysis of mind with mentions of a temporal analysis of motivation and emotion. Transient events at

reflexive level can cause transient arousal and perhaps appraisal in both individual agents and agent societies either by design or as an emergent property of changes in one agent's motivational processing and subsequent behaviour. Mediating actions may be shared between small groups, while long-term concerns are modelled across the whole society. Agents that accept motivations to achieve short-term goals that are necessary but at odds with their long-term goals (i.e. they exhibit a small degree of sophisticated flexibility) will experience perturbant control states. Such agents need an internal model of autonomy that is capable of defining such perturbant states in motivational terms if they are to differentiate between expected and unexpected perturbant states. A multiple-level representation of emotion is one such model

Accepting the doctrine of architectural parsimony (Hayes-Roth 1993) means that in designing and implementing these agents, no process is incorporated unless necessary. In short, there is no duplicity of processing mechanisms and a minimal configuration is produced for any specific design target. What happens when we create teams of such agents when some goals, drives or concerns are shared across the agents? Do each have explicit structures for those concerns? Can we make use of an alternative agent architecture that not only models the common concerns (goals or motivations) but is used by all the agents to perform relatively expensive deliberative processing associated with motivation? Furthermore how do these agents communicate to each other about these motivations, given that the motivations may require multiple levels of representations? We are addressing such questions by means of agent models that permit multiple forms of autonomy.

Our experiments are incomplete. We are currently investigating the relation between motivation, emotion and autonomy in a number of related projects. For example the designs described here appear promising in the design of single agents in an environment of other agents. However there remains the question of how communities of agents are able to share motivators. For this form of communication to be meaningful across co-operative cliques of agents, each agent may need to be part of a shared emotional model that underlies their individual and collective autonomy. In effect we are looking to build a distributed emotion engine, and see if such mechanisms can result in more effective agent behaviour. It is too early to draw even tentative conclusions. However any attempt to seriously model emotion as an add-on to a current computational model will result in a shallow system. The broad but shallow approach (Bates et al 1991) has influenced this research

for a number of years. Such an approach in adding emotion to a current model of mind is flawed. Emotion is at the heart of every-day human life. If it is to be used in computational models of the mind or in agents capable of sophisticated motivations it should be at the heart of their autonomy.

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. A.B. Loyall & W.S. Reilly, Broad agents, *SIGART BULLETIN*, Vol. 2, No. 4, 1991.
- Beaudoin, L. *Goal Processing in Autonomous Agents*. Ph.D. Thesis, Computer Science, University Of Birmingham, 1994.
- 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., Reactive and motivational agents. *Intelligent Agents III: Agent Theory, Language and Architectures*, Springer-Verlag, 1997.
- 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*
- Davis, D.N. Agents, Emergence, Emotion and Representation, *IEEE IECON2000, Nagoya, 2000*.
- Davis, D.N., Control States and Complete Agent Architectures, *Computational Intelligence*, 17(4) 2001.
- Epstein, J.M. and Axtell, R. *Growing Artificial Societies*, MIT Press, 1996.
- 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
- Hanks, S., Pollack, M.E. and Cohen, P.R., Benchmarks, Test-beds, Controlled Experimentation, and the Design of Agent Architectures, *AI Magazine*, 14(4):17-42, 1993.
- Hayes-Roth, B., Intelligent control. *Artificial Intelligence*, 59:213—220, 1993.
- 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.
- Merleau-Ponty, M., *The Structure of Behaviour*, 1942.
- 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.
- RoboCup Federation, RoboCup Official Site, <http://www.RoboCup.org/02.html>, 1998-2001
- Rolls, E.T., *The Brain and Emotion*, Oxford University Press, 1999.
- Scherer, K.R., Studying the emotion-antecedent appraisal process: An expert system approach, *Cognition and Emotion*, Vol: 7, 325-355, 1993.
- Simon, H.A. Motivational and emotional controls of cognition, *Models of Thought*, Yale University Press, 1979.
- Sloman, A., The mind as a control system, In: *Philosophy and the Cognitive Sciences*, C. Hookway and D. Peterson (Editors), Cambridge University Press, 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.
- Wheeler, M. 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, 1997, pp 10-19.
- Wollheim, R., *On The Emotions*, Yale University Press, 1999.
- Wright, I., Sloman, A. and Beaudoin, L., Towards a Design-Based Analysis of Emotional Episodes, *Philosophy Psychiatry and Psychology*, Vol. 3 no 2, 1996, pp 101--126.