Cellular Automata, Computational Autonomy and Emotion

Darryl N. Davis
NEAT Research Group, Department of Computer Science, University of Hull
Kingston-upon-Hull, HU6 7RX, UK
E-mail: D.N.Davis@dcs.hull.ac.uk

Abstract

Central to the idea of computational intelligence is the concept of autonomy. Here the thesis that a computational equivalent to emotion may provide a robust foundation for autonomy is considered. Many computational intelligence architectures are competency-based designs related to tasks in specific domains. More general frameworks map across tasks and domains. These types of computational architectures tend to fit well with the concept of weak notion of agency; i.e. they define autonomous systems that perform specific roles within a real or abstract environment. Four foundations to the weak notion of agency are autonomy, social ability, reactivity and pro-activeness. These tend to be defined in terms of interactions between an agent, its goals and its environment. From the perspective of developing computational systems this is acceptable. However, these definitions are insufficient for more sophisticated designs (and architectures) that claim to function intelligently. There is no core to these agents other than an information processing architecture. In future computational intelligent systems, this will be problematic. In developing sophisticated computational systems, irrespective of any research goal to develop human-like reasoning or task systems, complex intra-system and environment-system interactions can arise that will compromise the system's goals. A computational equivalent to emotion may help in recognising and managing such situations. Rather than consider emotion solely as an emergent property of such systems, a synthesis of psychological models of the emotions provide design descriptors for the different forms of processing performed by the agent. The primary conjecture is that through designing and implementing such architectures, emergent properties can be harnessed in a harmonious manner. As an agent functions it is sometimes called upon to monitor its own internal interactions and relate the nature of these wholly internal functions to tasks in its external environment. The modification of an agent's internal environment, and its actions upon its external environment, is described in terms of an emotion motivated mapping between its internal and external environments. The impetus for change within itself (i.e. to adapt or learn) is manifested as a combination of emotions.

1 Introduction

This paper results from a series of experiments in the use of cellular automata in defining agent autonomy. Agent is a ubiquitous term. However it still does offer credence as a metaphor and short-hand for the sorts of system that embody some degree of computational intelligence. We use a computational analogue to emotion to provide consistent descriptors for autonomy across different layers of agent architectures. These emotions are defined so as to offer a useful model for the design of computational systems capable of identifying and autonomously monitoring and managing their internal state. This research builds on work on the modeling of motivation [1, 2] and the use of cellular automata in game playing A-life agents [3]. An impetus for this research is the conjecture [4] that sophisticated computational systems will display perturbant emotion-like qualities, as an emergent behavior when resource requirements cannot be met or ongoing system goals are thwarted. Such emergent behavior is troublesome if we consider affective computation as part of the next generation of intelligent computational systems [5].

Affective computation will necessarily require computational systems to be able to model emotion at a computational level in order to recognize and identify the affective state, and hence emotional stance, of their users and other systems to which they interface. This affects the criteria with which we view and design sophisticated computational systems.

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 [6]. In discussing the various types of autonomy that an agent can demonstrate, a distinction can be drawn between belief and goal autonomy [7]. 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 a process 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. The management of belief sets is described in terms of rationality (for example logic or fuzzy based reasoning) and the credibility of incoming information (belief updating). Autonomy can also be defined with regard to "a set of tendencies, which can take the form of individual goals to be achieved" [8:pp 9-10]. This notion of autonomy is constrained by the concept of validity, and is in effect defined in terms of epistemological events and processes. There are however other models of autonomy.

2 Why Computational Models Of Emotions?

Phenomenology supposes humans are moved to action by disequilibria between the self and the world [9]. One impetus for thought and action in human beings is a lack of cohesion in the mapping and/or understanding of the relation between internal and external environments. Emotion is a primary source of motivation, and plays a large role in initiating and providing descriptors for these types of disequilibria. Perturbant states arise in any information processing (i.e. epistemological) infrastructure where there are insufficient resources to satisfy all current and prospective goals [10]. Such a computational system must be able to regulate these affective states or compromise its autonomy. However to consider emotions solely as an emergent quality of mental life that undermines reason and rationality is short-sighted. Emotions can also be harmonious and rewarding. They play an important role in the executive aspects of cognition, i.e. judgement, planning and social conduct, and the valencing of thoughts related to emerging problems, tasks and challenges [11, 12, 13]. In short emotion has a central role in a functioning mind. There is therefore a case for a computational model of emotion in the construction of theories and designs of intelligent systems.

Cognition can be considered to be the source of emotion, with valence (i.e. positive-neutral-negative) and appraisal states (cognitive reflection of the valences) as the primary basis for describing emotion [14]. From this perspective emotions differ from other cognitive states as they require a valenced reaction. Emotions differ from other affective states as they require appraisal. There are basic classes of emotion focussed on events (pleased vs. displeased), agents (approving vs. disapproving) and objects (liking vs. disliking). Specific emotions are instances and blends of these types and subclasses. Emotions of the same type have eliciting conditions that are structurally similar. They reject the idea of emotions such as anger and fear being fundamental or basic emotions. This model of the emotions has been used in several computational systems [15]. The cognitive processing that appraises emotions is goal-based and resembles the type of processing and structures discussed in motivation for autonomous agents [16].

Emotion can be defined as "a state usually caused by an event of importance to the subject. It typically includes (a) a conscious mental state with a recognizable quality of feeling and directed towards some object, (b) a bodily perturbation of some kind, (c) recognizable expressions of the face, tone of voice, and gesture (d) a readiness for certain kinds of action" [17]. Others [18, 19] give similar definitions with some agreement on the physiological, expressive and semantically distinct basic emotions, for example

- Fear defined as the physical or social threat to self, or a valued role or goal;
- Anger defined as the blocking of a role or goal by the perceived actions of another agent.

Rolls [12] presents a complementary perspective on the psychology of the emotions. Brains are designed around reward and punishment evaluation (reinforcer) systems. While this can be seen as analogous to the valenced arousal states in the Ortony et al. theory, reinforcers are precursors to specific emotions. Rather than reinforcing particular behavioral patterns, the reinforcement mechanisms work in terms of cognitive activity such as goals and motivation. Emotions are states elicited by reinforcers. These states are positive when concerns (i.e. goals) are advanced and negative when impeded. Emotions have many functions (Rolls lists ten) including the priming of reflexive behaviors associated with the autonomic and endocrine system, the establishment of motivational states, the facilitation of memory processing (i.e. storage and control) and maintenance of "*persistent and continuing motivation and direction of behavior*".

3 Theoretical Framework

Thinking about, designing and building agent systems is guided by cognitive frameworks used in human mental functioning. For example, a BDI agent architecture [8] is a computational extension to human thinking about problems using logic. The use of such rational models is understandable. They provide formal systems with well-defined properties and limitations (e.g. logical omniscience). Such systems are amenable to the epistemological tasks and goals in the micro-worlds that many agents inhabit and internally represent. However, their use effectively circumscribes the depth of the foundational concepts underlying a currently adopted notion of agency. Where computational resources are constrained, goal-based agent models will necessarily have to prioritize which goals are to be pursued. This can lead to perturbant control states. Without an effective way of managing these states, such agent designs compromise their autonomy and reactivity. Even the most rational of agents can be compromised if lacking the mechanisms to cope with the side-effects of determining which among many goals are those to be pursued. Many researchers recognise this problem and attempt to circumnavigate it through the addition of a meta or reflective layer to their agent design. What seems to be missing in these designs is a consistent descriptor model for defining the current state of processing across the entirety of the agent. These meta-layers are therefore different in type to the architectures they attempt to control.

An alternative stance is place at the core of agent processing a control model defined in terms of the agent’s main design criteria (for example emotion). This provides an agent with a model of self that maps across different levels and types of processing. Emotion provides an internal source of autonomy and a means to valence information processing events. In the remainder of this paper the influence of environment, both internal and external, upon this (agent-self) model of autonomy is considered. This emotion core gives rise to episodic states, trajectory states and (semi-permanent) endogenous states. These control states provide an agent with an internal model it can use to valence many aspects of its behavior. Through the development of an appropriate control model, an agent can regulate itself in terms of qualitatively different types of motivation.

The theoretical framework presented here revisits an earlier computational architecture and emphasizes the interplay of cognition, emotion, appraisal and motivation. Emotions are appraisal states with supporting and causal processes. This provides a regulatory framework for the different forms of emotion inducing events, whether designed, emergent, harmonious or perturbant. In moving towards a computational viable model of emotion, the extent of the model will be (initially) minimized for reasons of ontological parsimony and design tractability. For this model five basic emotions (anger, disgust, fear, happiness, and sadness) suffice. In humans these are physiologically, expressively and semantically distinct. In fact this can be simplified to a four emotion model, if the valence (i.e. positive-neutral-negative) and appraisal (goal and motivation driven reflection of these valences) of each of the five are considered. Sadness and happiness are antipathetic, being reflections of each other, or extremes on one dimension, sobriety. This four dimensional model provides a framework with which to investigate computational models of emotion and its relation to autonomy.

Earlier research on agents [1, 2] focussed on an architecture that supports motivation. A salient feature of the working definitions of emotion is that they are described in terms of goals and roles. The architecture (sketched in its simplest form in figure 1) emphasizes four distinct processing layers: a reflexive layer containing processes that mirror purely autonomic systems, a reactive layer of semi-autonomic behaviors, a deliberative layer of structure manipulation and reformulation and a reflective or meta-control layer. A consistent descriptor and control model enables high and low level processes to co-exist and interact in a holistic manner. The higher level processes tend to remain dormant or frozen, protected by filters and are activated only when sufficiently required. The agent’s processing exists in relation to the agent’s environmental stance; i.e. which objects and agents exist and what events are occurring in the environment and how they affect the logistics of goal satisfaction. Motivator processing, planning and other deliberative processes are not merely abstract but exist in relation to an agent’s long term goals and immediate circumstance. The agent is autonomous to the extent that it determines how these long-term goals (the reason for its existence) are to be achieved. The extent of its autonomy is governed by the nature of its design and epistemological and environmental skills. Different trajectories (goal-achieving behaviors) are possible for any move between actual and desired control states. Some trajectories while impossible are supported or attended to for any number of reasons. The possible trajectories between control state sub-spaces depend on an agent’s design. The preferred trajectory between these spaces depends on autonomous preference for specific aspects of its internal landscapes. An agent is autonomous to the extent that it can choose to pursue specific motivational trajectories. An agent is rational to the extent that it follows feasible (or achievable) trajectories. An agent's internal landscape is the epistemological space that underpins an internally consistent understanding of external events, objects and agents – this need not nor is it always rational. Goals and drives may be implicit or require explicit representations. Drives and motivations embody some representation about desired agent and environmental states and relate short-term emotive states to temporally global processes [20]. Temporal-global drives are those associated with the agent's overall purpose and related to its current, possible and desired control state spaces. Temporal-local drives are related to ephemeral states or events within the agent's environment or itself. These can give rise to more enduring motivational states that may be acted on.

What are the control patterns that stabilize this model? An agent of a specific design will concentrate on certain tasks that favor specific aspects of the possible internal landscape as external agents, objects and events affect emotionally valenced goals. Emergent and possibly perturbant control states arise from the interaction of current temporal-global roles (the favoring of certain aspects of task space) and temporal-local drives that reflect the current focus of the agent. Emotional autonomy means an agent maintains an ongoing (globally-temporal) disposition with a range of (emotional) responses to specific events and their effect upon agent's overall emotional stance. The nature of this disposition can be temporarily modified through the adoption and modification of current goals and motivations. Over time events occur that modify, stall, negate or satisfy goals. Such events can impinge on all layers of the architecture, affecting current dispositions and can lead to a re-calibration of motivator preference. These events give rise to reinforcers across the agent architecture. The emotion(s) and hence epistemological and control states they reinforce depends on interactions across the breadth and depth of the agent architecture. Reinforcers and the low-level (reflexive) valences can be modeled using the interval {-1,1} - this interval need not be linear. A discrete version of this interval maps onto the three tokens: negative, neutral and positive. Non-emotion low-level drives can be associated with reinforcers and be valenced. They can also be associated with motivators and goals across the agent architecture. The management and success (or otherwise) of these drive-generated motivations can give rise to emotions. A salient feature of the given definitions of emotion is that they are described in terms of goals, roles and expressive behaviors. This enables emotions to be defined over different levels of the architecture using different aspects of motivational behaviors and provides a cogent core to the agent's autonomy.

If emergent behaviors are to be recognized and managed then there must be a control synergy across the different layers of the architecture. Processes at the deliberative level can reason about emergent states arising from anywhere in the architecture using explicit (motivator or goal) representations. Emotions can be instantiated by events both internal and external at a number of levels of abstraction, whether primary (reflexive or reactive drives) or by events that require substantive deliberative processing. Emotions can be invoked through appraisal of agent, object or event related scenarios. The postponement (or abandonment) of a goal may cause an agent to experience emotion states related to unwanted dispositions. To move to a preferred aspect of the possible emotional landscape, an agent may need to instantiate other motivators and accept temporarily unwanted dispositions. An agent with emotional autonomy accepts temporary emotional perturbance if more acceptable control states result. In the model in figure 2, intense emotions effectively override the emotion filter causing the forced deliberative consideration of the emotional state. Similar filters are used in earlier work on motivator generactivation [1]. The deliberative appraisal of the emotion then activates laterally at the deliberative layer, affecting memory management, filters and motivator management. Changes to an agent's beliefs are possible through external influence, as in the Castelfranchi model of autonomy [7] but here this process is mediated by the emotional autonomy of the agent.

4 Agent Design

We have investigated the design and implementation of a number of instances of this model. We are pursuing the conjecture that computational processes based on cellular automata provide a convenient modeling environment for the different micro-foundations of decision making. Earlier work on GO playing agents used sets of homogeneous CAs to landscape the decision-making space associated with that game. In the present work an insect community metaphor was initially employed, using sets of heterogeneous cellular automata. This provides a foundation for the agent model. The behavior of any CA, and any community of CAs (or *hive*), is directed by default behaviors. Other non-default behaviors can be chosen by the reasoning process responsible for managing any specific hive. Other processes are responsible for unifying the global behavior of the agent. The hives define the agent's reflexive capabilities and are automatic. They provide the computational foundation to the agent architectures described here. All other processes in these agent architectures are activated in response to extrema either in specific individual CA or across hives. These other processes are filter-protected.

In the initial experiments, four categories of CAs are used: data filters; information gatherers; emotion cells (four per hive); and hubs (hive center-cells) which reflect the emotion state of the hive. All are 4 connected, or potentially so. In some experiments the hives were three dimensional with each CA being six connected with four lateral, one dorsal and one ventral connection. The nature of the CA processes is modeled at the reactive level in terms of eleven sets of behaviors. Six sets of behavior mirror the possible connection combinations of CA used in the design and the nature of their communication (types of cell connections), four sets are used to model the internal behavior of the automata, while the last is used to model hive interactions. At any one time, one or more behaviors from any of the behavior sets will be active; the others remain dormant. It is the process responsible for managing a hive that determines which are the currently extant sets of behaviors for the cellular automata. For simplicity, we initially modeled the valence of each of the 4 emotions as negative-neutral-positive. A discrete four-element vector can then be used to represent a specific emotional state. A hive hub is, in effect, modeled as a 3^4 state cellular automaton. Each emotion cell is responsible for maintaining the state of one emotion; one for anger, disgust, fear and sobriety. These can be in one of three discrete states. Information gatherers are two state automata; they hold a piece of data or they do not. Filters accept or refuse communication with their neighbors – they act as a fine-grain attenuator. In communities of homogeneous cellular automata, relatively simple rules govern the behavior of any cell. Such simple rules will not work here, as neighbor cells are not necessarily homogeneous. Even if such simple rules were applicable, we require more adaptive behavior. Hence, we need a range of behaviors (or rules) for determining the next state of a cell when connected to cells of a specific type. This set of behaviors is extended to cope with the nature of communication. One specific behavior requires that a three-state CA read a two-state cell as positive (ON) or neagtive (OFF), while a two-state cell read a three-state as ON when positive (ON), otherwise OFF (when neutral or negative). Further combinations are possible. This results in ten possible communicative behaviors for such combinations. Similar analyses define the remaining cell communication behavior sets (a further 31 behaviors). Further behavior sets define the state change functions of the automata. Qualitatively different rule sets define more global behavior, for example the nature of the interaction between a hive and mobile. A mobile is any CA community that wants to communicate with another. Such connections can be refused or condoned by hive management processes with information flow blocked within the hive by the filter cells. Communication across the layers of the agent architecture works differently; typically in terms of filter protected activations. The reflexive hives are sampled periodically. If they are in extrema, a control message activates the deliberative process responsible for the hive. The sampling rate, and the nature of an extrema are defined by the deliberative process and can be changed according to the overall needs of the agent. The processing context of the deliberative-hive couple determines what happens next. The deliberative process may not respond and simply accept the extreme state, despite its regular activation prompts, as an unwanted but necessary side-effect of mediating motivations. The deliberative process may change the set behavior of the reflexive cells. The deliberative agencies may read the internal state of any or all cells in any hive they manage. The deliberative process may activate laterally at the deliberative level informing the attention mechanism of the emotive state so it can be correlated to other ongoing processing in the agent. Upward moving messages provide emotional state information, while downward moving messages act as behavioral directions in emotion space. Some of our ongoing experiments are investigating the effect of provoking an emotional perturbation at one level and seeing how that effects the agent. For example, if we arrange for the agent to be emotionally neutral (across all the emotions and different layers), what happens if the agent experiences anger at just one level? Does it make a difference if this perturbation occurs at the reflexive, deliberative or reflective levels? The answer is not straightforward and depends on the overall context of the agent. If the reflective (and other) layers experience a common emotion extreme, the agent as a whole will turn its attention to the cause. An extreme state across a deliberative-reflexive couple causes the agent to attend to the cause in addition to other concerns. With an extrema in a reflexive agency alone nothing else may happen.

Experimentation continues with the development of causal models and the parameterisation of control states, with a view to enabling the agent to recognise and manage emotional perturbation. At a more global level, we can consider how these different processes and layers interact and their effect on the activities of the agent as a whole. By modeling emotion in a locally stochastic manner at the reflexive layer and allowing the behaviors at that level to be modified through interactions across the different levels of the agent architecture, a dynamic non-linear control model is realised. Not all the agencies (semi-autonomous integrated processes within the overall agent) need be active or exist all the time. The agent can create new (deliberative-reflexive) agencies on demand to cope with new scenarios, and then disassemble them when no longer needed. To do this the agent needs to be aware of its ongoing processes. At some albeit superficial level, the agent needs to be conscious of its own internal state. The reflective layer provides this internal state classifier. Through the use of the reflective-level emotional model, the agent can classify its current emotional state as a specific vector in emotion space.

Figures 1 and 2 present a four-layer processing model of the emotions. The ongoing automatic processes (Emotion:A) present a base for the model. The reactive behaviors (Emotion:R) define the behavior of all Emotion:A processes. The currently extant behaviors are set by deliberative processes (Emotion:D). The meta-management reflective process (Emotion;M) monitors the deliberative appraisal of the Emotion:A processes and the state of the attention filters. The output from Emotion:M provides guidance to the attention management, Emotion:D and the Emotion:A processes. The agent manages its emotions through the development of these five modules. Other aspects of the emotion engine are deliberative motivator processes, and memory management.

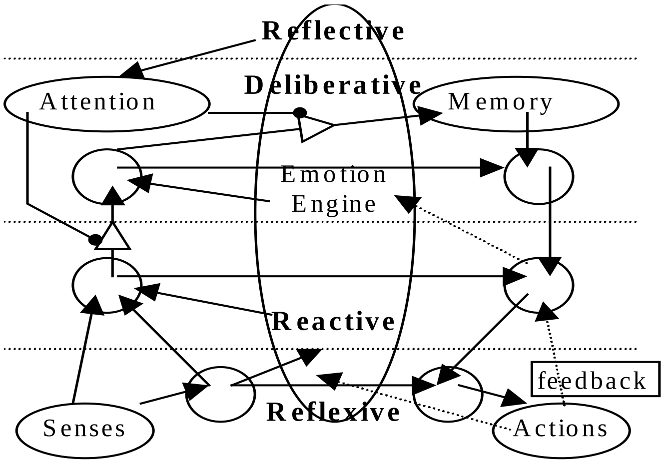

Fig. 1. *The simplified four-layer architecture with emotion as the core. Circles on the left represent information assimilation and synthesis processes. Circles on the right represent information generation processes that typically map onto internal and external behaviors.*

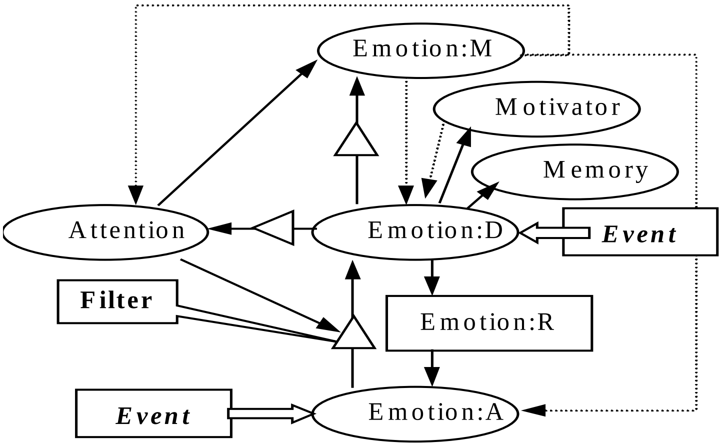

Fig. *2 The emotion engine. The core of the design sketched in figure 1.*

The four emotion-cell one hub model has been expanded to use four communities of three-state automata; one to represent each of the four dimensions of our model – the basic emotions of anger, disgust, fear and sobriety. Each emotion is discretely valenced in each of the cells as positive-neutral-negative. Further hive types represent reinforcers - valenced pre-emotive events. The behaviors associated with Emotion:R are those that govern the internal behavior of single cells, the communication between adjoining cells in communities and inter-community communication. This is the basis for the emotion engine, and ultimately underpins the notion of autonomy in the agents. Communication between different hives (and events outside of the emotion engine at the reflexive and reactive level) is by means of further CA communities. The currently set behavior from these dispositions and disposition-event couplings is selected by a deliberative (Emotion:D) process responsible for asynchronously monitoring these communities in response to intense hive states and to guidance from Emotion:M. Early experiments found that from a given state, the CA communities rapidly achieved a steady (continuous or oscillating) state. By changing the currently extant behavior set or by communicating with another hive (through the use of a mobile CA community) transitions to the same or other steady states always occurred. Approximately 20,000 transition possibilities exist. Rules are used to select different CA community dispositions and transitions. Through the modification of the internal state of a small number of cells the emotion engine moves to a closely related state. The deliberative processes change their emotional context (the temporally-global aspect of emotions) and hence the currently extant behavior set for their hive in response to the reflective processes. The deliberative processes also disturb the motivator management processes with their current emotive state – as part of an asynchronous appraisal mechanism. Memory management (Memory:D) also responds to the Emotion:D processes in order to provide emotional context to the storage of memories about external events, objects and agents. The attention filter processes also make use of the Emotion:D-Emotion:A complexes to provide semantic context for motivator filters. The quantitative emotion filters are set directly by Attention:D mechanisms. The intensity levels of these filters are set in response to the Emotion:D mechanisms and the reflective component of the emotion engine.

An experimental harness was developed using an agent toolkit in which the emotion engine is trained to prefer specific combinations of emotions. Artificial scenarios are then provided in which the hive(s) are set in specific and random configurations. As the different agent “personalities” prefer different aspects of emotional space, the engine modifies itself so that a preferred emotional state arises. This allows experimentation with motivators of varying transience and sophistication in a number of domains. Experimentation continues with the investigation of how the different control mechanisms interact. We are finding that any particular initial combination of cell states and reactive behaviors rapidly maps onto a steady state at the hive level. As the deliberative agencies change the currently active behaviors, minor changes in the steady state occur. The effect of coupling these agencies with the reflective classifier is to be investigated shortly. The relation of these internal states cognitive behaviors such as goal management is ongoing research.

Conclusion

The current work is part of an ongoing research program into how to design and build sophisticated agents and makes use of past research efforts in a cyclic agent design framework [11, 2]. This approach allows us to consider for example a single cellular automaton as an agent in itself. We then select and combine a number of these agents to build a community of cellular automata, which are added to the set of prototype reflexive agents. We then build reactive agents that comprise of behaviors suitable for different categories of cellular automata communities. An agent of this class can then be combined with a cellular automata agent to produce yet another prototypical agent class. Having produced prototypical agent classes for all the agencies required for our design, we then combine and refine their synthesis in a holistic manner. If at any point, any of the prototypical agent classes fail to meet their design requirements, particularly at the integration stages, that class of agent must be redesigned and the design process reiterated.

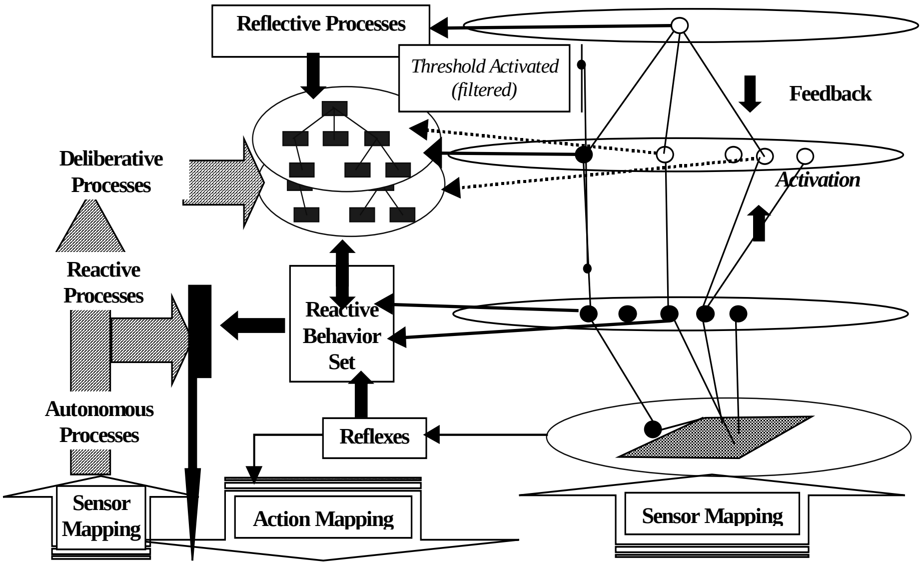

Figure 3. A more sophisticated architecture making use of the emotion engine.

Current research is developing the model presented here in a number of test-beds and experimental frameworks. This includes extending the architecture to allow more sophisticated motivation management. The architectures vary with the test-bed and the short term research goals associated with each experimental framework. For example, we are investigating the nature of motivation across teams of co-operative and competitive agents using simulated Robo-Cup. This addresses the work described in [1, 2] but aims to provide a more detailed analysis of how the computational structures associated with motivation and its management can be distributed across teams of agents. One follow-on experiment to the work presented in this paper is in the use of cellular automata as perceptual mechanism for agents in predator-prey scenarios. Predator-prey is used because the drives, goals and motivations of the agents are easily described (stay alive and reproduce!). We can then allow agents to use current concerns (such as herd or predate or escape) to drive their perceptual processing. As scene sampling (i.e. machine vision) is being used as well as simulated sensory apparatus from within the agent toolkit, the images are modelled as active pixel arrays of cellular automata. The rules that govern the processing of these images are emotion-driven and adapt to the ongoing concerns of the agent. For example, image pixels representing food items become positively valenced when the agent is hungry, neutral otherwise. Hence objects and agents in the external environment give rise to valenced perceptual reactions at the image level. Between sampling the environment these arrays of CAs function as described above with extreme states (defined using adaptive thresholds) causing activation of further processing in the agent. Figure 3 sketches this architecture.

The primary reason for the *preliminary* research described above was to gain a better understanding of the relations between emotion, cognition and mind. Although earlier research on the computational modeling of motivation looked promising, there was a psychological implausibility with the motives behind motivators. If synthetic agents are going to experience emotions because of the nature of multiple-goal processing, then the computational infrastructure of those agents needs a representational framework in which these emergent qualities could be harnessed. Through the development of stronger notions of agencies based on psychological models and philosophical analyses, better-founded concepts underlying the weak notion of agency ensue. We are using these to drive a computational model of synthetic mind that can monitor and manage emergent control states.

References

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

[2] Davis, D.N., Control States and Complete Agent Architectures, *Computational Intelligence, 17(4) 2001*.

[3] Davis, D.N., T. Chalabi and B. Berbank-Green, Towards an architecture for artificial life agents: II, In: M. Mohammadian (Editor*), New Frontiers in Computational Intelligence and Its Applications*, ISO Press, 1999.

[4] Sloman, A. and Croucher, M. Why robots will have emotions. *IJCAI7*, 197-202, (1987).

[5] Picard, R., Affective Computing, MIT Press, 1997.

[6] Wooldridge, M. and N.R. Jennings (Eds), *Intelligent Agents*. Springer-Verlag, 1995.

[7] Castelfranchi, C. Guarantees for autonomy in cognitive agent architectures. In [18]: 56-70.

[8] Ferber, J. *Multi-Agent System*s, Addison-Wesley, 1999

[9] Merleau-Ponty, M., *The Structure of Behavior*, Methuan:London, 1965.

[10] Sloman, A. Architectural requirements for human-like agents both natural and artificial, In *Human Cognition and Social Agent Technology, K. Dautenhahn*, Benjamins, 1999.

[11] Davis, D.N. Synthetic Agents: Synthetic Minds, *IEEE SMC-98, San Diego, 1998*.

[12] Rolls, E.T., *The Brain and Emotion*, Oxford University Press, 1999.

[13] Spaulding, W.D. (Editor), *Integrative Views of Motivation, Cognition and Emotion*, University of Nebraska Press, 1994.

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

[15] Elliot, C., I picked up catapia and other stories: A Multi-modal approach to expressivity for 'emotionally intelligent' agents, *First Int. Conf. On Autonomous Agents, 1997.*

[16] Beaudoin, L.P. and A. Sloman, A study of motive processing and attention, In: *Prospects for Artificial Intelligence,* Sloman, Hogg, Humphreys, Partridge and Ramsay (Editors), IOS Press, 1993.

[17] Oatley, K. and Jenkins, J.M, *Understanding Emotions*, Blackwell, 1996.

[18] Frijda, N., *The Emotions*, Cambridge University Press 1986.

[19] Power, M. and T. Dalgleish, *Cognition and Emotion*, LEA Press, 1997.

[20] Davis, D.N, Agents, Emergence, Emotion and Representation, *IEEE IECON2000, Nagoya, Japan 2000*.