

Emotional cognitive steps towards consciousness

Will N. Browne
Cybernetics

University of Reading, Reading, Berkshire, UK
w.n.browne@reading.ac.uk

<http://www.personal.rdg.ac.uk/~sis01wnb/>

The academic journey to a widely acknowledged Machine Consciousness is anticipated to be an emotional one. Both in terms of the active debate provoked by the subject and a hypothesised need to encapsulate an analogue of emotions in an artificial system in order to progress towards machine consciousness. This paper explores why the topic of 'emotion' promotes debate - due to the contributing interacting systems and varying definitions. It then considers the inspiration that the concepts related to emotion may contribute to cognitive systems when approaching conscious like behaviour. Specifically, emotions can set goals including balancing explore versus exploit, facilitate action in unknown domains and modify existing behaviours, which are explored in cognitive robotics experiments.

1. Introduction

It may be argued that evolution has led to consciousness in human beings. However, this process has taken millennia, which is not desirable when attempting to create artificial conscious like behaviours. Thus, rather than start with a tabula rasa, it may be preferable to start with analogues of evolved human traits in an attempt to speed-up the creation of useful behaviours. One such trait is emotions, which have shown to be common across many groups of humans and are hypothesised to be important in developing cognitive abilities, e.g. Ekman [99]

There are many identified functions of emotions that could benefit the operation of a cognitive robot, see section 2.2, e.g. one such emotional function is to achieve a multi-level communication of simplified, but high impact information [Fellous 04]. Many other functions have been suggested, including resource mobilisation and conservation, prioritisation of behaviours, decoupling stimulus and response and so on [Scherer 94]. Behavioural difficulties associated with some cognitive architectures, such as

dithering, may be addressed through emotions. Emotions can act as stimuli or goals themselves, similar to internal message passing. However, as emotions are linked to the outside world there is less likelihood of parasitic behaviours emerging, which are self-sustaining without assisting the agent in reaching its goals.

In common with many aspects in the field of consciousness there is a debate on the exact scope and meaning of 'emotions'. This paper considers relevant literature, but not in an attempt to distil a hard definition. Rather, interesting aspects of concepts related to emotions, which may provide inspiration for the development of an artificial analogue to emotions, are identified. Initial cognitive robotics experiments are presented as supporting evidence.

It has been argued that emotions are not the sole priority of humans and if other animals can have emotions, then so can robots [Fellous 04]. However, due to significant differences in embodiment and sensorimotor experiences 'robot-emotion', although fulfilling similar roles, are likely to be very different to 'human-emotion'. The aim of this paper is to identify methods to utilise (robot-)emotions for generating useful behaviours in cognitive robotics.

The topic of emotions may be studied by many fields, such as Neuroscience, Psychology, Anthropology and Cybernetics. The insight provided by Cybernetics is an important aspect of this work. Wiener defined Cybernetics as 'control and communication in the animal and machine' [Wiener 48]. There are many aspects to the communication aspect with the enhancement of interaction, e.g. Human Robotic Interaction [Breazeal 04] being the most publicised. However, there are aspects of self-affirmation, including viewing the self relative to others, which are considered to be a more important role for emotions when related to conscious behaviours - it is noted that advanced communication has been

considered to be core to self consciousness [Parslow 06].

There are many levels on which to investigate emotions, from the visceral to the philosophical. Similarly, if an analogue to emotions is to be developed based on biological inspiration, this inspiration may come from one or more of these levels. The lower the level investigated, the more the model can be used to give insight into the actual biology of the system studied [see work by Krichmar 03 & 05]. However, the work presented here is based at the functional level, rather than the neuronal level. It is not intended to give insight to the actual biological mechanisms and pathways. Instead, it seeks to replicate interesting functions and behaviours. Thus, it is not constrained to use connectionist neural net-based approaches, so will use symbolic based approaches due to their transparency of operation. There are strong arguments against the use of symbolic, including production rule, based approaches for investigating the mind/brain. This paper does not seek to disprove these arguments, but show that if care is taken in the problem selection, application of methods and claims made that the symbolic approach to emotions for cognitive robotics has benefits.

Section 2 investigates existing work on emotions, addressing the aim to discover inspiration for an analogue of emotions. Section 3 uses this inspiration to identify steps towards a cognitive controller for robotics experiments, investigating the worth of this analogue. It is considered that whilst nature has evolved emotions in humans, these emotions are tuned by nurture, i.e. through interaction with an environment. Section 4 highlights initial experiments where the benefits of innate emotions and the ability to nurture responses are investigated. The results from physical robotic experiments based on the ideas developed in section 3 are presented prior to discussion and conclusions.

2. Background to Emotions

Research into emotions considers how and why they arose in humans (and animals), where they are sited in the brain/mind, their functionality and their categorisation. There is little agreement to these research questions and the answers have varied over time, with new investigating tools giving more insight, but requiring refining or reassessment of the existing knowledge.

Individuals differ in their emotional responses either through pre-programming or experience. However,

individuals have a phenomenological experience of emotions so are free to give comment and judgement. Unfortunately, the phenomenological experience may not be what actually occurred and the differences in experience may make the discussions regarding emotions subjective.

Interesting analysis of emotions has been conducted for millennia, e.g. Greek philosophical study of emotions. A useful starting point to this research is Darwin's work on the similarity of emotions in different species, which supported his ideas of a common ancestor [Darwin 1871]. The 'survival of the fittest' would suggest that emotions are useful for the survival of an animal that can take advantage of them. However, it is interesting to consider whether emotions arose before consciousness or consciousness arose before emotions or whether they arose simultaneously. Thus, it may be more appropriate to rephrase Darwin's famous quote to 'the fittest survive whether they want to or not', as it is plausible that cognitive animals developed emotions as a precursor to consciousness [a discussion of desires (wants) versus needs occurs in section 3.1].

2.1 Physiology of emotions

There are broad theories that relate the order of stimulus to experienced emotion [see slides]. Thus emotions are considered to be more than just a reflection of past events.

Another over simplification is that there is just a single emotion generating site in the brain. Several brain regions are involved in a single emotion, these may be different between different emotions and the same brain region may participate in different emotions [Fellous 04]. Neurotransmitters, such as serotonin, epinephrine (adrenaline) and dopamine relate to emotion, e.g. a lack of serotonin increases aggression. Emotions may be considered as signals. These 'signals' vary in timescale (few milliseconds to hours), location and scope. Some have effects on the nervous system, immune system and endocrine systems, which are all interrelated and in turn affect other emotions. Thus to accurately and predictably model emotions it is arguable that every interacting system within the human body needs to be accurately modelled.

Haikonen [03] proposes that machine emotions are interactive combinations of systems reactions that determine the style of cognitive processes and actual responses. These autonomic system reactions are in response to elementary sensations, which are proposed as:

- good-bad

- pain-pleasure
- match-mismatch-novelty

It is noted here that elementary sensations resemble the concepts of desire, motivation and needs, see section 3.1.

Emotions are developmental, i.e. some emotions are present at birth, by nine months all basic emotions are present. Self awareness emotions (e.g. embarrassment) 18-24 months and evaluated emotions (guilt) developed by 2-3 years. Thus the trade-off between nature and nurture invites exploration.

Emotion is both complex and adaptive [Michand et al. 01]. Fellous concludes that it may be " more fruitful to focus on function of emotions not what they are", which is the focus adopted by this work.

2.2 Functionality of emotions

In agreement with the underlying physiology there is not just a single function or use for emotions, e.g. emotion is not just a communication aid. Again there are differing viewpoints and opinions on the functionality of emotions [Arbib and Fellous 04][Di Paolo and Iizuka 08][Scheutz 04][Takeno 05].

A broad overview of functionality is "Emotions are reflections of the adaptations that animals make to universal problems" [Plutchik 01], where the universal problems of adaptations are temporality, identity, hierarchy and territoriality, see section 2.3 for corresponding classes of emotion.

Theories of emotional functionality are over a hundred years old, including Darwin [1871] and James [1884]. These often reflect current ideas on the location of functionality of the brain in general. Although there is often overlap and some agreement, there are many differences in functional theories, see figures 1 and 2.

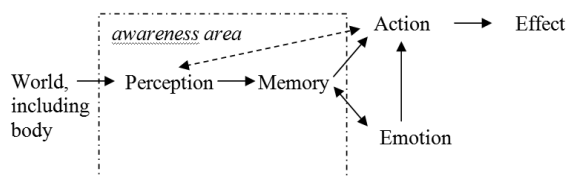


Figure 1 Aleksander and Morton [07] emotional architecture this model shows both afference and efference.

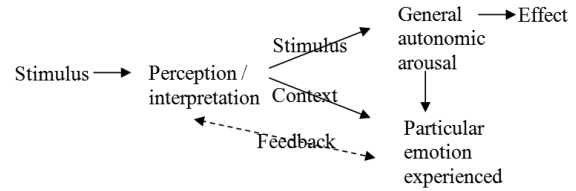


Figure 2 Schachter's cognitive theory model shows efference then afference [64]

There is an important difference between afference feedback and efference signals. Re-experiencing the emotional context of a state can *affect* the decision taken so is a form of emotional feedback, see somatic marker hypothesis [Damasio 96]. Similarly, being in an emotional state biases the decision making with this signal *effecting* the output.

A good example of the debate regarding the exact functionality of human cognition surrounds the somatic marker hypothesis. Rolls [99] argues that it has a major weakness in that if afference feedback is based through a peripheral response it would be a slow and noisy process.

The purpose of emotion is also debated, with a general categorization provided by Michaud et al [01]

- to adapt to limitations
- for managing social behaviour
- for interpersonal communication

This is not a definitive list as emotion has also been linked to the memory of facts, which is improved when the facts are learnt in connection with an emotion (to a limit) [Cahill 95][Hamann 05]. Also, there a strong link between emotion and decision making and other frontal lobe cognitive functions, e.g. working memory [Bechara et al. 00].

Alternative viewpoints exist, e.g. in relation to whether emotions require a sense of self. Some define function dependent on scope of consequences [Averill 94]

- intended versus unintended
- short versus long-term
- target to individual versus target to social
- singular versus predictable events

Further details can be expanded, e.g. Rolls [99] identifies 10 major functions

- change in autonomic and endocrine system
- flexibility of behavioural response to reinforcing stimuli
- triggering motivated behaviours
- communication

- social bonding
- improving survival
- affecting cognitive processing (mood congruence) and facilitating its continuity
- facilitating memory storage
- allowing persistence of motivated behaviours
- facilitating memory recall

This list of categorisation is not exhaustive, e.g. [Fellous 04 with 12 categories], so although there is some agreement and overlap between differing lists, there is no template for reproducing exact functionality in cognitive robots.

2.3 Classification of emotions

Similar to the differing viewpoints and multiple useful concepts for emotional functionality there are varying types, descriptions and definitions of emotions themselves

A well known classification is by Plutchik [91] whose classification is based on purpose:

Temporality: *joy/sadness*

NB joy is higher response than happy

Identity: *acceptance/rejection*

CF affection/disgust

Hierarchy: *anger/fear*

Territoriality: *expectation/surprise*

CF exploit versus explore [McMahon et al. 06]

The different levels of intensity is an important concept reflected in most classifications, e.g. from Ecstasy to Terror in Rolls [99]

Anthropology studies by Ekman [99] have identified classes of emotion with similarities across cultures. That these classes are especially true in the recognition, albeit not completely accurate recognition tests. The 1972 list of anger, sadness, happiness, fear, disgust, surprise was expanded with contempt and embarrassment and then rewritten with 15 base emotions in 1999.

2.4 Definitions of emotions

With inconsistency when investigating emotions, it is unsurprising that there is no widely accepted definition. A couple of insightful definitions are presented here:

Rolls [99] considers the essence of emotions in that emotions are states elicited by rewards (anything an animal will work to obtain) and punishers (anything an animal will work to escape or avoid), including changes in rewards and punishment.

Michaud et al. [01] contend that emotion is a complex set of interactions among subjective and objective factors, mediated by neural/hormonal systems, which can

1. Give rise to affective experiences such as feelings or arousal, pleasure/displeasure;
2. Generate cognitive processes such as emotionally relevant perception of events, appraisals, labelling processes;
3. Activate widespread physiological adjustments to the arousing conditions
4. Lead to behaviour that is often, but not always, expressive, goal directed and adaptive.

Simply put, this leads to notions of emotion:

1. emotions are the causes of action
2. cognition can trigger emotions – emotions can trigger cognitive operations
3. emotions can cause expressive actions
4. emotions can affect purposive, goal-directed action
5. emotions can become goals
6. the behaviour also affects emotion

2.5 Existing emotional models

Computational models of emotion exist for both neuroscientific understanding and for cognitive robotic control [Kawamura and Browne 08][for standard architectures see Kieras and Meyer 97, Laird et al. 87]. In the former class includes emotional learning in the amygdala [Moren and Balkinius 00], which is based on the Amygdalo-orbitofrontal system expounded by Rolls and LeDoux [96]. The balance of inhibitory with excitory signals is important in the simulations [Shanahan 05]. John Taylor's group models the interaction of attention and emotion, including the enhancement of perception caused by emotional cues [Fragopanagos et al. 06].

A bridge between these types is the models of Fellous whose behavioural investigation contains an organisation where the potential for emotional control (through neuromodulation) increases with higher level cognition (from reflexes to drives to instincts to cognitions). Other advice for emotional modelling in cognitive robotics includes Clark & Grush's promotion of a minimal yet robust internal representation [99] and affective architectures [Sloman and Chrisley 05][Sloman and Logan 98].

Kawamura et al [05] develops the role of episodic memory and emotion for the cognitive robot ISAC with the emotion is based on Haikonen's System Reactions Theory of Emotion (SRTE). The relation between emotions and system reaction is predefined,

e.g. pain due to an external agent will result in an aggressive response. The DARE architecture [Macas et al. 01] again works with the double and parallel stimuli processing concept of LeDoux plus the somatic marker concept of Damasio. Campagne and Cardon [03] approach an emotion model from similar inspiration using a multiagent perspective in a massive simulation only.

3. Emotional Inspiration

If robots are to benefit from mechanisms that have a similar role to emotions it is suggested to use internal variables [Michaud et al. 01]. However, Fellous warns that an isolated emotion is simply an engineering hack, i.e. simply describing a single, isolated internal variable as an emotion could be descriptive or anthropomorphic, but not biologically inspired [04]. Instead, interrelated emotions, expressed due to resource mobilisation with context dependent computations dependent on perceived expression is more realistic.

A consequence of this is that an artificial system must have limited resources in order to express emotions. These emotions may appear different if expressed externally or internally, but are very related due to their underlying mechanisms.

Thus robot-emotions should be built from the following guidelines [ibid]:

- emotions are not a separate centre that computes a value on some predefined dimension
- emotions should not be a result of cognitive evaluation (if state then this emotion)
- emotions are not combinations of some pre-specified basic emotion (emotions are not independent from each other)
- emotions should have temporal dynamics and interact with each other.
- System wide control of some of the parameters (of the many ongoing, parallel processes) that determine the robot behaviour.

3.1 Desires versus needs

Motivational effect of rewards on behaviour is universally acknowledged by experimental psychologists [Wise and Bozarth 87]. This can either be appetitive or aversive

The standard reinforcement learning diagram from Sutton and Barto [98] is useful, but misleading as it has the reward signal entering the agent externally. Reward is considered here to be an internally generated variable (often termed pleasure) based on some external state signal. Pleasure arises through

satisfying a need [Cabanac 71], which once satisfied means there is no longer a need. Most pleasure seekers (needs) are probably inbuilt, but can learn others [Rolls 99]. However, a need requires it to be wanted (desired) before actions can be taken to satisfy it. Thus there is an important difference between wanting something and something giving pleasure as can have one without the other.

An agent is required to learn what states lead to pleasure – see efference learning

An agent is also required to have a desire to find pleasure – see efference learning.

An agent must be motivated (wanting to satisfy a need) with 'motives' acting as goals [Michaud et al 01]. Emotions fulfil this role as they are more motivational (action tendencies, biases, allocators) than behavioural (trigger specific actions), see Franklin and Ramamurthy 06][Savage 03].

Many current cognitive agents have a single goal (a single need), which is sought using an e-greedy search strategy that returns a constant reward when entering the goal state. Consequently, emotions are thus not needed to mediate between differing needs by controlling desires.

Differing needs could relate to power levels, exploration, collision avoidance or a user defined goal. These can be pre-programmed to generate reward when satisfied. It is hypothesised here that humans have a need to improve their predictive ability and this should be replicated in robots – see reduction of cognitive dissonance theory [Oudeyer 08].

3.2 Opposites and attractors

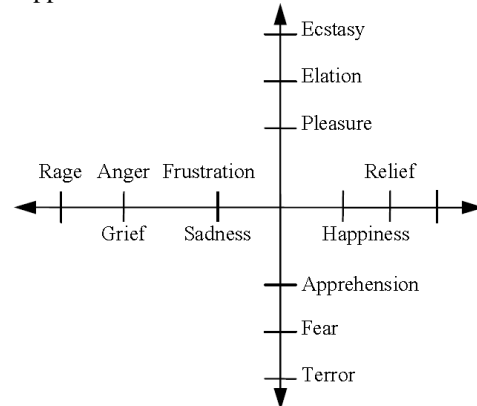


Figure 3 Bipolar ranges of emotion, based on Rolls [99].

A common theme in biologically inspired models of emotion is a range of emotions between opposites [Rolls 99] [Plutchik 91], e.g. the two-dimensional

bipolar model of four emotions joy/sadness, anger/fear [Michaud et al. 01].

Rolls visualises these as a linear graph, see figure 3, but dynamical systems theory suggests otherwise. It is unlikely, considering the brain as a dynamic system with multiple causes interacting over nested timescales [Smith 03], that emotional structure will be linear. An alternative model is attractor states, as proposed for neurotransmitters rather than set levels from zero [Fellous 04]. These states may be considered stable limit cycles rather than set point attractors or chaotic variance. This enables emotional states to be stable for certain perturbations, but very rapidly switch to alternative emotional states due to other input signals - an interesting property for artificial cognitive systems.

3.3 Multidimensional and multimodal

There are strong arguments that single production rules cannot work as a basis for emotional cognitive control, but to dismiss all 'if ... then ...' symbolic systems based on these arguments appears premature.

To explore some arguments against production rules, consider efferent emotions where state s_1 evokes emotion e . However, it may not be derived from state to action (a) to reward (r) to emotion $s_1-a_1-r_1-e$ as the episode may have been $s_1-a_1-s_2-a_2-s_3-a_3-s_4-a_4-r-e$ or alternative complex sequence. Similarly, it cannot now be stated that 'if s_1 then e ' as the next time s_1 is presented the system reasons on afferent as well as efference signals s_1 & e , which could produce a completely different action-reward-emotion sequence, e.g. if a need has been satisfied.

From section 2 the types of emotions are limited. Each emotion may be considered as a dimension, which consists of a series of limit cycles from one extreme reflecting aversive rewards to the other representing appetitive rewards. 'Negative' emotions are just as essential for the survival of the agent as 'positive' emotions as they help to avoid aversive states. It is to be decided how these emotional values are changed, but it is considered that states, actions, rewards and even emotions themselves may influence the emotional level. Although it is wrong to have a production rule that always says s_1-e , it is acceptable to have a production rule that says s_1 & e evokes e and $s_1 - \Delta e$, where the evoking and changing of e is not unique to s_1 .

3.4 Affectors and effectors

A production rule model for emotions could be $s&e-a$ evokes e and Δe . However, this can be complemented by direct rules, e.g. s_1-a or s_1-e and that can be

inhibited other rules where these are hardwired or learnt.

Haikonen [03] identifies five stages of consciousness with step three relevant here. Emotions are considered to be for:

- attention control
- motivation
- shortcut template is the style of action
- affect learning

Thus selected states invoke emotions and the totality of emotions corresponds to rewards.

Forward modelling is necessary for cognition with efference copies being required to predict the world. It is considered that emotions are likely to be evoked when the model and world do not match. It may be considered that the world is its own best model [Brooks 85], but there is still a requirement for an internal model on how best to use the external model with sensors [check Holland reference]. The evolutionary computation based production rule system of Learning Classifier Systems (LCS) was originally proposed by Holland as a cognitive system [75]. Decades of research have enabled LCS to become an effective machine learning technique [Lanzi 02][Butz and Wilson 02]. Recently computed predictions [Butz 04] have been introduced into LCS where the output is a function of the input state. Not only has this improved the representational capabilities, but represents a method by which the world acts as an interpreted model that can be learnt through environmental interaction.

The world may be considered as continuous, but the symbolic architecture considered above is discrete. However, human senses often discretise the continuous world, e.g. human eyes being limited to 20 Hz, so the morphology [Pfeifer and Scheier 99] of a cognitive robot will aid the design of a cognitive robot. Another important consideration is the parallel nature of the human brain versus the serial operation of a computer. With multi-core processors and multithreaded languages becoming more available the artificial architectures may better reflect parallel functionality.

Synchronisation of senses, episodes, emotions, cognitive control and motor effecting then becomes a critical problem. Plausibly the artificial replication of 'brainwaves' could assist the coordination of signals [Heraz et al. 07]. It has been observed that different EEG frequency bands responded differently to three types of emotional film content (aggressive, sad, neutral).

Although emotional learning occurs in many of the brain regions, a primary centre is the amygdala Moren and Balkenius [00]. Emotions are a reinforcing signal that is a reaction to the presentation of a primary stimulus or alternatively a reaction to a stimulus that has an intrinsic emotional charge - second order, where a stimulus has been associated with either previous stimuli is possible. Emotions can provide a rough and quick classification of sensory stimuli, e.g. a state evokes an immediate emotional response rather than requiring a chain of complex cognitive processing to determine an appropriate action.

3.5 Nature and nurture

To provide the agent with robot-emotion it is necessary to form each emotion – realised as a dimension. This could be from a tabula-rasa to determine every emotional dimension which is useful for a robot. However, this is likely to take many iterative cycles even if the embodiment and problem complexity could be developed appropriately. Thus an assumed robot-nature is required with predetermined dimensions, including attractor states. Primary needs require setting, and emotional responses to states with associated pleasures may require initial bootstrapping, e.g. increasing happiness by satisfying hunger or evoking fear when approaching states associated with pain.

The nurture of the agent should allow it to tune the limit cycles, especially the transitions between cycles and how the changing emotions are updated. Importantly, the afference needs to be learned and then the associated efference determined.

3.6 Generalisation, abstraction and anticipations

Learning requires memorisation of perceptions from the environment in order to store useful behaviours.

1. Irrelevant information from a perceived state must be removed. Attention must be focused on the important components of the state, which may be accomplished through learning generalisations.
2. Higher order patterns must be abstracted from learnt episodes so that rules may be applied to similar situations. Affordances may then be calculated [Guazzelli et al. 98]
3. Anticipatory models must then be built up linking future states to existing states with plausible actions

LCS generalise by denoting irrelevant conditions in a state as 'don't care' or by removing them from the production rule itself. Initial work has shown abstraction is possible and beneficial in appropriate environments [Browne et al 08], but further work is

required. This is also true for the determination of affordances, which is assisted by the matching of rules to states by LCS including forward planning. Anticipatory LCS implementations include ACS, ACSII and AgentP [Zatuchna 05] where *s-a-s* rules are autonomously formed.

3.7 Memory

Many types of memory have been classified in humans, including short-term, long-term and working memory [Baxter and Browne 08]. However, much debate exists regarding the underlying biological mechanisms for the observed functional differences [Phillips and Noelle 05].

The ability to generalise and abstract does reduce the required memory, say compared with a Q-learning state table. However, to accurately model a practical world would still require much memory of rules leading to potentially slow searching, accessing and maintenance. Efficient matching by removing irrelevant conditions also assists, whilst improved anticipations could enable pre-searching of likely rules that might be activated in the near future. Importantly, the ability to associate emotions with rules could greatly reduce the search space, e.g. a sad agent may ignore many rules associated with joy.

3.8 Episodic, semantic and procedural knowledge

It is argued that emotions assist storage of knowledge in memory. This requires more than storing knowledge about instantaneous state-actions as episodic, semantic and procedural knowledge are also important for a cognitive agent. Only the emotional components of these facets will be discussed now.

An episode may be defined by an unchanging element through a duration of time, e.g. achieving a goal, an unchanging condition or constant action. Therefore, episodic memory may be triggered based on emotional states, e.g. an episode associated with a happy memory. This includes pleasure gained from a goal achieved, which assists in storing procedural knowledge.

Although knowledge may be unemotional, the meaning of certain knowledge has strong emotional context. For example, procedural mathematics may not be emotional - although it is noted that an emotional response to solving (or not) complex problems is possible. Conversely, the meaning of 'rose' is often considered to require emotional context.

3.9 Emotional communication

The previous analysis has considered *control* through feedback from the environment for cognitive learning,

but Cybernetics also informs us that *communication* is an important consideration. It is argued here that emotions play a significant role in communications above and beyond assisting human-robot interactions.

The feed-forward model for the cognitive agent with learning and emotional feedback has been sketched above. However, the model is currently egocentric with no explicit interaction with other agents and no sense of self. It has been considered that an agent is by nature aversive to uncertainty, which includes the uncertainty associated with poor models. An environmental state may include other agents that must be modelled including their intentions for future action. In order to make accurate predictions the internal states of these other agents and their ability to make intelligent decisions are needed. These properties are not readily accessible to the agent.

Emotions are good for quick, high impact information transfer [Fellous 04]. Thus if an agent recognizes an emotion in another agent (due to having a similar nature), then it can postulate based on mirroring its own rules associated with this emotion to determine plausible internal states and likely actions of the other agent.

Communication through modelling emotional states facilitates the ability to place other agents in an 'out there' world. By mirroring these rules the agent can then place itself in the 'out there' world too. These abilities form part of the arguments by Owen Holland for building a consciousness [Holland 03].

4. Results from Cognitive Robotics Experiments

Prior to embarking on the journey (outlined above in section 3) to develop a cognitive architecture that includes the multiple facilities of emotion, two important questions were addressed. Firstly, is the nature of artificial emotions useful to a robot, such that it provides functionality not easily obtained by other means? Secondly, can emotions be tuned by nurture through interaction with a given environment in order to improve their usefulness to an agent?

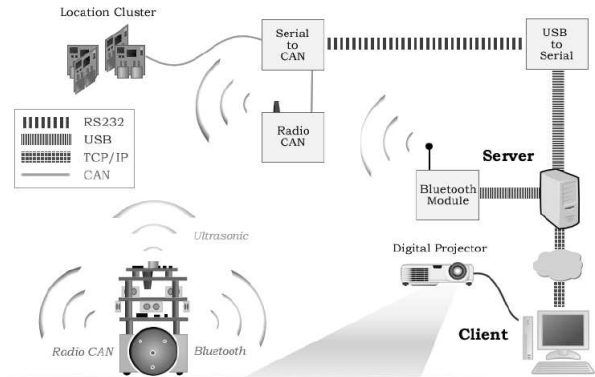


Figure 4: The development platform

The experimental platform used in both experiments is outlined in figure 4. The server handles all sensory and motor commands, with the client representing the perceptions, reasoning and learning from the environmental interactions.

4.1 Nature

The full details of the experiment outlined in Browne and Tingley [06]. Essentially, a static rule-base was corrected manually including both afference and efference rules but no direct state-action links, see figure 5. Multiple states could affect an emotion and each of the five modelled emotions could effect an action.

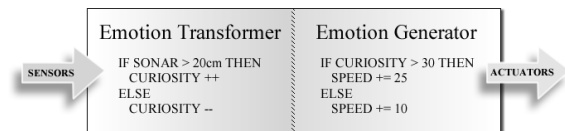


Figure 5: An example of an emotion transformer and associated transformer.

Superimposed runs of an explore task are shown in figures 6 and 7. A non-emotional benchmark architecture produced almost identical paths when starting from the same position, which is common in deterministic systems. Introducing randomness to the controller was ineffective as the robot makes little forward progress. However, when considering the paths generated by the emotional system, both runs were completely different, but equally effective – ~90% exploration compared with ~70% for the benchmark system.

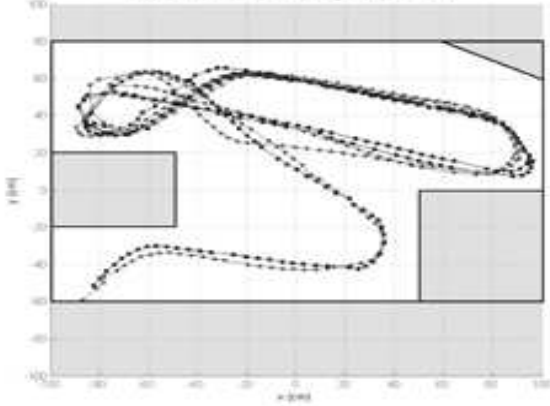


Figure 6: A non-emotional agent architecture exploring a complex domain for three minutes.

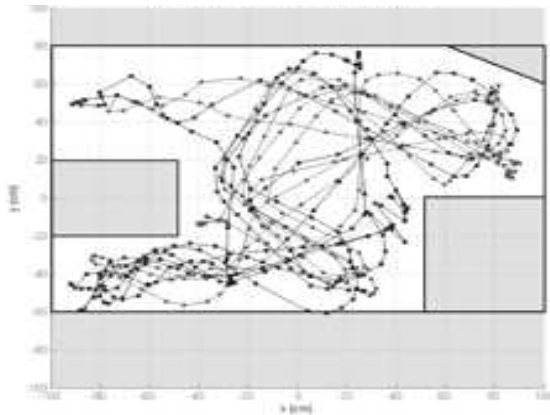


Figure 7: An emotional-based agent architecture exploring a complex domain for three minutes. Note the non-deterministic nature of the two runs.

5.2 Nurture

The emotion analogues selected for this problem domain are: Happiness (P+), Sadness (P-), Curiosity (I+), Anger (I-), Hope (D+), and Fear (D-). Where each emotional signal is related to satisfying the need by a proportional (P), integral (I) or derivative (D) relationship and maybe appetitive (+) or aversive (-). An LCS is used to learn the effective use of emotions, i.e. best action for a given emotional state. Again the task was to explore, initially in a simple domain. Experiments showed that as the training progressed the area explored in a given time interval increased, see figures 8 and 9.

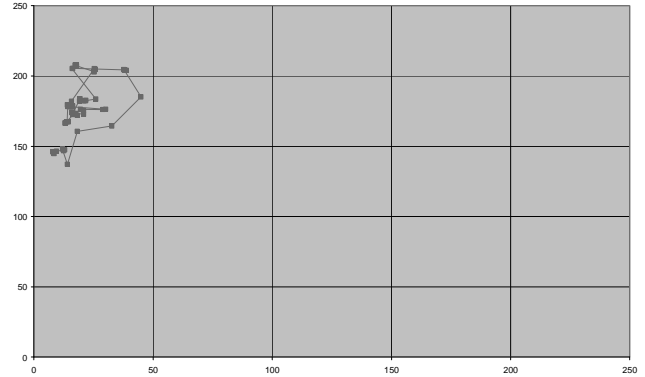


Figure 8: Plot of the map data from LCS run 4

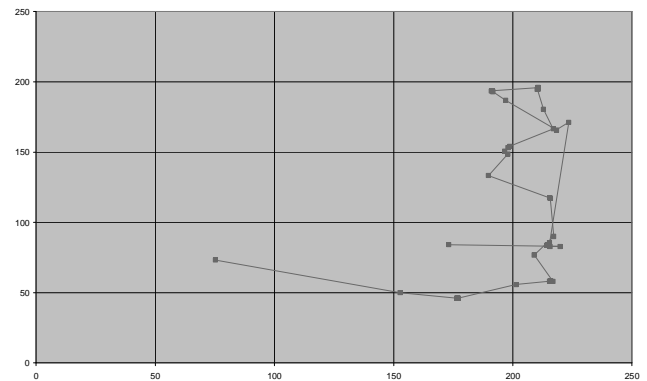


Figure 9: Plot of the map data from LCS run 9

Examining the rules produced showed plausible learning had occurred. The autonomously identified fittest rule learnt:

```
111 - If ((18<=Happiness<=88) & (6<=Sadness<=23)
& (12<=Curiosity<=81) & (17<=Anger<=85) &
(17<=Hope<=85) & (8<=Fear<=52))
{ Action = 1; }
```

This corresponds to the behaviour: agent is in the open space, then go forwards slowly.

As the motor actions were set arbitrarily, this suggests that the top range of speeds were too fast for the domain leading to collisions with the walls.

Due to the accuracy based nature of the LCS used, it also formed aversive rules. The behaviour to be most avoided is: if agent is trapped, then turn left fast [note that trapped corresponds to low values from the ultrasonic sensors].

5. Discussion

There is a wide variety of proposed causes, functionality and classification of emotion. Hence there is no widely accepted definition of emotion or a rulebook to follow when replicating the useful facets

of emotions in robotics, but there are useful guidelines.

There is a danger that internal variables for control or aspects of communication will be anthropomorphized and termed 'emotions' when only loosely inspired by natural phenomena. The multiple and interconnected nature of emotions were shown to produce useful behaviours in initial experimentation, i.e. non-linear, non-deterministic (directed rather than random) and adaptation to the environment.

Production rule-based systems have potential for both affective and effective learning, especially when internal reward is linked to satisfying identified needs. The ability to reason on emotional states, including combinations of multiple emotions, is essential for generating realistic behaviours. Furthermore, the ability to set emotions, change emotions and form chains of rules adds to the temporal, including episodic, learning capabilities.

Much work is required to build on positive results for abstraction and anticipatory learning, e.g. to enable transparent identification of affordances. Similarly, as the problem domains grow in complexity, memory and knowledge utilisation will require useful emotional traits for operation. Finally, instilling a need for predictive certainty coupled with emotional communication would lead to an agent approaching conscious behaviours when placing itself in an 'out-there' world.

It was noted that the control actions of the robot-emotions were similar to the actions of conventional controllers. A simple analogy is that some emotions reacted proportionally to the input signal, others built up over time (integrated) and others reacted to the rate of change of a signal (differential). Proportional, integral and derivative (PID) is a standard industrial control strategy. It is worth considering the links to other conventional control strategies, such as filter-based techniques (e.g. lead control), model based control (e.g. Smith predictor) and adaptive control to determine if natural emotions have similar actions that could be replicated artificially.

6. Conclusion

The topic of 'emotion' promotes debate due to the contributing interacting systems and varying definitions. However, inspiration that the concepts related to emotion is likely to significantly contribute to cognitive systems when approaching conscious

like behaviour. Specifically, emotions can set goals including balancing explore versus exploit, facilitate action in unknown domains and modify existing behaviours, which are explored in cognitive robotics experiments. A need for predictive certainty coupled with emotional communication is postulated to lead to an agent approaching conscious behaviours when placing itself in an 'out-there' world

References

Alexander I (2005) *The World in My Mind, My Mind in the World: Key Mechanisms of Consciousness in People, Animals and Machines*. Imprint Academic, UK

Alexander I and Morton H., (2007) Why axiomatic models of conscious? *Journal of Consciousness Studies*, 14, No 7, pp 15-27.

Arbib M. A. and Fellous J.-M., "Emotions: from brain to robot," *Trends in cognitive sciences*, vol. 8, pp. 554-561, 2004.

Arkin R. C., Fujita M., Takagi, T. and Hasegawa R., "An ethological and emotional basis for human-robot interaction," *Robotics and Autonomous Systems*, vol. 42, pp. 191-201, 2003.

Averill, J. R., Chon, K. K., & Haan, D. W. (2001). *Emotions and creativity*, East and West. *Asian Journal of Social Psychology*, 4, pp165-183.

Baars BJ (1997) *In the Theater of Consciousness: The Workspace of the Mind*. Oxford University Press

Baxter P. and Browne W. (2008), *Towards a developmental memory-based and embodied cognitive architecture*, *Epigenetic Robotics* 8, University of Sussex, July 30-31, Abstract, pp137-138

Bechara A., Damasio H., and Damasio A., "Emotion, Decision Making and the Orbitofrontal Cortex," *Cerebral Cortex*, vol. 10, pp. 295-307, 2000.

Breazeal C., *Function Meets Style: Insights From Emotion Theory Applied to HRI*, *IEEE Transactions On Systems, Man, And Cybernetics—part c: applications and reviews*, vol. 34, no. 2, May 2004

Brooks, R.A. "A robust layered control system for a mobile robot.", Technical report, Massachusetts Institute of Technology, Cambridge, MA, (1985).

Browne, W. N., Scott D. and Ioannides C.,

"Abstraction for Genetics-based Reinforcement Learning", in "Reinforcement Learning: Theory and Applications", editors Cornelius Weber, Mark Elshaw and Norbert Michael Mayer, Advanced Robotic Systems Publishing, Vienna, Austria, EU, January 2008.

Browne, W. N. and Tingley, C. (2006), Developing an Emotion-Based Architecture for Autonomous Agents. In Third International Conference on Autonomous Robots and Agents (ICARA 2006), pp 225-230, 12th-14th December 2006, Palmerston North, New Zealand.

Butz, M., and Wilson, S. W. An algorithmic description of XCS. In *Soft Computing: a fusion of foundations, methodologies and applications*, 6 (2002), 162-170.

Butz, M., Rule-based evolutionary online learning systems: learning bounds, classification and prediction. PhD thesis University of Illinois, Illinois, 2004.

Cabanac, M. (1971) Physiological role of pleasure. *Science*, 173, 1103-1107.

Cahill L, Babinsky R, Markowitsch HJ, McGaugh JL. The amygdala and emotional memory. *Nature*. 1995 Sep 28;377(6547):295-296.

Campagne J C, Cardon A, Artificial emotions for robots using massive multi-agent systems, in SID2003, London, 2003

Canamero L., "Emotion understanding from the perspective of autonomous robots research," *Neural Networks*, vol. 18, pp. 445-455, 2005.

Canamero, D. "Issues in the design of emotional agents.", *Emotional and Intelligent: The Tangled Knot of Cognition: Papers from the 1998 Fall Symposium*, pp 23-27 (1998).

Clark A, Grush R (1999) Towards a cognitive robotics. *Adaptive Behavior*, 7(1): 5-16

Damasio A., "The somatic marker hypothesis and the possible functions of the prefrontal cortex," *Philosophical Transactions Of the Royal Society B*, vol. 351, pp. 1413-1420, 1996.

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

Darwin, C (1871). *The Descent of Man and selection in relation to sex*. John Murray, London. (Reprinted in 1981 by Princeton University Press)

Di Paolo E. and Iizuka, H. "How (not) to model autonomous behaviour," *BioSystems*, vol. 91, pp. 409-423, 2008.

Edelman GM (1987) *Neural Darwinism: The Theory of Neuronal Group Selection*, Basic Books, NY

Ekman P., "Basic emotions", Dalglish T. and Power T., *Handbook of Cognition and Emotion*, Sussex, John Wiley & Sons, pp 45-60 (1999).

Ekman, P. (1999) *Basic Emotions*, In T. Dalglish and T. Power (Eds.) *The Handbook of Cognition and Emotion* Pp. 45-60. Sussex, U.K.: John Wiley & Sons, Ltd.

Fellous J-M, From Human Emotions to Robotic Emotions, American Association for Artificial Intelligence – Spring symposium 3/2004, Stanford University, 2004.

Fragopanagos N., Korsten N. & Taylor J. G. (2006) "A neural model of the enhancement of perception caused by emotional cues," *Proceedings of the International Joint Conference on Neural Networks 2006*, Vancouver,

Franklin S. and Ramamurthy U., "Motivations, values and emotions: three sides of the same coin," presented at *Proceedings of 6th Int. Workshop on Epigenetic Robotics*, Paris, France, 2006.

Guazzelli A., Corbacho F. J., Bota M., and Arbib M. A., "Affordances, motivation, and the world graph theory," *Adaptive Behavior*, vol. 6, pp. 435-471, 1998.

Haikonen PO (2003) *The Cognitive Approach to Conscious Machines*. Imprint Academic, UK

Hamann S., "Cognitive and neural mechanisms of emotional memory," *Trends in cognitive sciences*, vol. 5, pp. 394-400, 2001.

Heraz, A., Razaki, R. & Frasson, C., Using machine learning to predict learner emotional state from brainwaves. 7th IEEE conference on Advanced Learning Technologies: ICALT 2007, Niigata, Japan,

Holland O (ed) (2003) *Machine Consciousness*. Imprint Academic, UK

Holland, J. H. adaptation in natural and artificial systems. Ann Arbor, MI: University of Michigan press, 1975.

James, W. "What is an Emotion?," Mind, vol. 9, pp. 188-205, 1884.

Kawamura et al. (2006) From Intelligent Control to Cognitive Control. In 11th International Symposium on Robotics and Applications (ISORA), Budapest, Hungary

Kawamura K, Dodd W, Ratanaswasd P, Gutierrez R (2005) Development of a robot with a sense of self. 6th IEEE Int'l Symposium on Computational Intelligence in Robotics and Automation, Espoo, Finland

Kawamura, K. and Browne, W. N., "Cognitive Robotics", Book Chapter in Encyclopedia of Complexity and System Science. Editor-in-Chief: Bob Meyers, Springer, to appear in 2008

Kieras DE, Meyer DE (1997) An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. Human-Computer Interaction

Krichmar JL, Edelman GM (2003) Brain-Based Devices: Intelligent systems based on principles of the nervous system. In IEEE/RSJ Int'l Conf. on Intelligent Robotics and Systems. Las Vegas, NV, pp.940-945

Krichmar JL, Reeke GN (2005) The Darwin Brain-Based Automata: Synthetic Neural Models and Real-World Devices. In Reeke GN, Poznanski RR, Lindsay KA, Rosenberg JR, Sporns O (ed) Modeling in the Neurosciences: From Biological Systems to Neuromimetic Robotics, Taylor & Francis F

Laird J, Newell A, Rosenbloom P (1987) Soar - An architecture for general intelligence. Artificial Intelligence 33:1-64

Lanzi, P-L., Learning classifier systems from a reinforcement learning perspective. In Soft Computing: a fusion of foundations, methodologies and applications, 6 (2002), 162-170.

LeDoux J., The emotional brain, New York, 1996

Macas, M., Ventura, R., Custodio, L., and Pinto-Ferreira, C. 2001b. Experiments with an emotion-based agent using the DARE architecture. In

Proceedings of the Symposium on Emotion, Cognition, and Affective Computing, AISB'01 Convention.

McMahon A., Scott D., Baxter P., Browne W. (2006), "An Autonomous Explore/Exploit Strategy", Proceedings of AISB'06, Bristol, UK, vol 2, pp192-201

Michaud F., Robichaud E, and Audet J., Using motives and artificial emotions for long-term activity of an autonomous robot, Proceedings of the fifth international conference on Autonomous agents, Montreal, Quebec, Canada Pages: 188 – 189, 2001

Minsky M., *The Society of Mind*, Simon & Schuster, New York (1988).

Minsky, M. 'The Emotional Machine' Simon & Schuster, November 7, (2006)

Moren, J. Balkenius, C. "A Computational Model of Emotional Learning, in The Amygdala," From animals to animals 6: Proceedings of the 6th International conference on the simulation of adaptive behavior, Cambridge, Mass., 2000. The MIT Press.

Oudeyer P-Y., and Kaplan F. How can we define intrinsic motivation? 8th Int. Conf. on Epigenetic Robotics, 2008, pp93-98

Parslow P, Emergent Consciousness – a discussion piece, COGRIC (2006): <http://www.cogric.reading.uk/>, Cognitive Robotics, Intelligence and Control, August 16-18

Pessoa L., On the relationship between emotion and cognition, Nature Reviews: Neuroscience, vol. 9, pp. 148-158, 2008.

Pfeifer R, Bongard J (2007) How the Body Shapes the Way We Think: A New View of Intelligence. MIT Press, MA

Pfeifer R, and Scheier C (1999) Understanding Intelligence. MIT Press, MA

Phillips JL, Noelle DC (2005) A biologically inspired working memory framework for robots. IEEE International Workshop on Robots and Human Interactive Communication, 599-604, Nashville, TN

Plutchik, R. (1991). The emotions. Lanham, MD: University Press of America.

Rolls, E. T 1999. 'The brain and emotion'. Oxford University Press.

Savage T., "The grounding of motivation in artificial animals: Indices of motivational behaviour," *Cognitive Systems Research*, vol. 4, pp. 23-55, 2003.

Schachter, S (1964) The interaction of cognitive and physiological determinants of emotional state. In *Advances in Experimental Social Psychology*, ed. L. Berkowitz, pp. 49-79. New York: Academic Press.

Scherer, K. R. (1994). Emotion serves to decouple stimulus and response. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 127-130). New York/Oxford: Oxford University Press.

Scherer, K.R. "Criteria for emotion-antecedent appraisal: a review.", *V. Hamilton, G. Bower, and N. Frijda, Cognitive perspective on Emotion and Motivation*, Kluwer Academic, pp 89-126 (1988).

Scheutz, M. "Useful roles of emotions in artificial agents: A case study from artificial life.", *American Association for Artificial Intelligence AAAI 2004*, pp 42-48 (2004).

Shanahan MP (2006) A cognitive architecture that combines internal simulation with a global workspace. *Consciousness and Cognition* 15: 433-449

Sloman A. and Chrisley R., "More things than are dreamt of in your biology: Information Processing in Biologically-Inspired Robots," *Cognitive Systems Research*, vol. 6, pp. 145-174, 2005.

Sloman A. and Logan B., "Cognition and affect: architectures and tools.", *Proceedings of the 2nd International Conference on Autonomous Agents 1998*, New York, pp 471-472 (September 1998).

Smith L B, and Thelan E., *Development As a Dynamic System*, *Trends in Cognitive Sciences*, Vol 7 No 8, August 2003

Sutton, R. S., and Barto, A. G. *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press, 1998.

Takeo J, et al. (2005) Experiments and examination of mirror image cognition using a small robot. In

IEEE International Symposium on Computational Intelligence in Robotics and Automation, Espo, Finland. 493-498

Wiener N (1948) *Cybernetics, or Control and Communication in the Animal and the Machine*. MIT Press, MA.

Wise, R. A. & Bozarth, M. A. A psychomotor stimulant theory of addiction. *Psychol. Rev.* 94, 469-492 (1987).

Zatuchna Z. V., *AgentP: a Learning Classifier System with Associative Perception in Maze Environments*, Thesis, School of Computing Sciences, University of East Anglia, 2005

Ziemke T., "On the role of emotion in biological and robotic autonomy," *BioSystems*, vol. 91, pp. 401-408, 2008.