# Reinforcement Learning and Animat Emotions

Ian Wright
CSRP-96-4
School of Computer Science
The University of Birmingham
Birmingham, B15 2TT, England
I.P.Wright@cs.bham.ac.uk
Phone: +44-21-414-3734

January 1996

**Abstract**

Emotional states, such as happiness or sadness, pose particular problems for information processing theories of mind. Hedonic components of states, unlike cognitive components, lack representational content. Research within Artificial Life, in particular the investigation of adaptive agent architectures, provides insights into the dynamic relationship between motivation, the ability of control sub-states to gain access to limited processing resources, and prototype emotional states. Holland's learning classifier system provides a concrete example of this relationship, demonstrating simple 'emotion-like' states, much as a thermostat demonstrates simple 'belief-like' and 'desire-like' states.

This leads to the conclusion that valency, a particular form of pleasure or displeasure, is a self-monitored process of credit-assignment. The importance of the movement of a domain-independent representation of utility within adaptive architectures is stressed. Existing information processing theories of emotion can be enriched by a 'circulation of value' design hypothesis. Implications for the development of emotional animats are considered.

## 1 Introduction

Motivation and emotion are often conspicuously absent from information processing theories of mind. For example, Allen Newell's description of the SOAR architecture (Newell 90, Laird et al. 87), the most advanced candidate for a unified theory of cognition, lists motivation and emotion as missing elements that need to be included in a more comprehensive theory.

Research into complete agent designs, such as 'animat' research within the field of Artificial Life, forces designers to attempt to integrate motivation, learning, sensing and acting within a single agent design to produce adaptive behaviour. If properly conceptualised, this work yields insights into the dynamic relationship between motivation, the ability of control sub-states to gain access to limited processing resources, and prototype emotional states. In particular, Holland's learning classifier system, a type of complex adaptive system (*cas*), often used in ALife research, provides a concrete, if simplified, example of this relationship.

A theoretical conclusion follows: the feeling component of some emotional states arises from the self-monitoring of a process of credit-assignment occurring within motivational subsystems. This conclusion enriches previous information processing theories of emotion and has implications for ALife research.

First, reasons are given why emotion may be thought 'difficult' for information processing theories.

# 2 Thinking refers but feelings just are

Cognitive representations can denote or refer to states-of-affairs that exist in an agent's domain. This is the physical symbol system hypothesis – the hypothesis that physical systems can implement symbols that contain information that denotes (Simon, 95). For example, an animat within a simulated domain may possess information about other agents in the environment, including their type, location, or speed. Such information sub-states of the animat are causally linked to their referents: the representation of the speed of agent A will alter if it is perceived that agent A has altered its speed. This is a simple example: referential links can be very indirect in more complex information processing systems. The principle, however, is conceptually clear and forms a basis of information processing theories of mind: *thinking refers.*

The emotions, however, differ from 'cold' cognition: they can be 'hot', often involving feelings of pleasure or displeasure with associated intensities. Unlike 'straightforward' representational thinking, an emotional state has *both* a representational content, e.g., a state of happiness *about* passing one's exams, and a hedonic, or *valenced* content, e.g., the particular form of intense pleasure one is experiencing. The hedonic component does not represent a state-of-affairs: *feelings just 'are'.*

Many information processing theories of emotion tend to avoid an explanation of the hedonic components of emotional states by concentrating on the semantics of representational components (e.g., Dyer's BORIS system (Dyer, 87), Frijda and Swagerman's ACRES system (Frijda & Swagerman, 87), and Pfiefer's FEELER system, reviewed in Pfiefer, 92). Alternatively, feelings are brushed under the physiological carpet by assuming that all valenced states arise from perceptions of bodily states. For example, Herbert Simon in his seminal paper on motivation and emotion (Simon, 67), outlines a view of 'feelings' that closely resembles William James' peripheric theory of the emotions: '... sudden intense stimuli often produce large effects on the autonomic nervous system, commonly of an "arousal" and "energy marshaling" nature. It is to these effects that the label "emotion" is generally attached'; and '... the feelings reported are produced, in turn, by internal stimuli resulting from the arousal of the autonomic system'. It is difficult to conceive how this view of valency could account for the mental pain associated with, for example, grief, which does not necessarily require bodily arousal or disturbance.

Therefore, there appear to be at least two reasons why explanations of hedonic states are generally absent from cognitive theories: first, valenced components appear not to conform to the representational model that supports cognition; and second, their possible functional role is unclear: what can such components possibly *do* if they do not represent? Why do such diverse and complex states such as happiness, sadness, glee, triumph, grief, despair, intense disappointment etc. have hedonic components?

## 2.1 A preliminary definition of valency

Such 'states' are phenomenologically highly variegated (compare your memories of being angry with being happy), with different causal antecedents (e.g., being slighted before your peers, or winning an Olympic gold medal) and different consequences (such as a desire for revenge or rest). The folk-psychological concepts that refer to these states play a communicative role between agents. The states themselves, however, are detected by sophisticated internal self-monitoring mechanisms, a kind of 'internal perception' (Wright, Sloman & Beaudoin, 96). This ability gives rise to the method of introspection or phenomenological analysis in psychology. Much theoretical work in the emotions draws on introspection, and this kind of knowledge places important constraints on possible theories. However, there is no reason to believe that the knowledge gained from our internal perception is any less fallible than that gained from 'external' perception. In the absence of good theories of the underlying mechanisms, care is required when employing phenomenological concepts.

A preliminary division can be made between *dispositional* and *occurrent* emotional states (Green 92, Ryle 49). A dispositional state is a latent state that may manifest in appropriate circumstances, such as the brittleness of a wine glass, whereas an occurrent state is a *running* state, such as the process of a wine glass breaking. For example, a man who has lost a parent may function normally at work (dispositional grief), only to break down in the evening (dispositional state manifests as an occurrent emotion).

There are two distinguishable components of an occurrent emotional state: *intentional* and *non-intentional*. The intentional component of an emotional state is what the state is about. A person is angry, disappointed, or ecstatic 'about' a perceived state of affairs[1]. This state of affairs may exist in the agent's environment, or entirely within cognition (as in the case of the mathematician irritated with himself for being unable to solve an equation). The intentional component has representational content.

The non-intentional component of an emotional state is often referred to as its 'hedonic tone', *feeling*, or *valency*. 'Feeling' is an ill-defined word, for it can cover such diverse sensations as one's cheeks burning with embarrassment, an itch on the left ear, or the mental happiness associated with triumph. 'Hedonic tone' is similarly semantically overloaded: it can be used to refer to the enjoyable sensation of a full stomach after a large and hearty meal. The word 'valency', if given a suitable definition, can avoid such confusion. Before giving such a definition we require some more distinctions.

A division can be made between *physiological* forms of pleasure and displeasure, and *cognitive* valency. For example, the (self-monitored) 'itchiness' on my left ear is a form of displeasure linked to information concerning bodily location. In contrast, the (self-monitored) mental pain of intense grief is a form of displeasure linked to information about a loved

[1]Moods can be considered an exception. (Sloman, 92) defines moods as global, semantic free control states

one's death. There can be no 'pain receptors' for this kind of displeasure, unlike the nerves that detect a pin pricking one's finger.

To illustrate: an athlete may be experiencing the occurrent emotional state of triumph while standing on the winner's podium. The intentional component of her state includes thoughts pertaining to her achieved goals; the non-intentional component *includes* the feeling of increased heart-rate or arousal, the warm sun beating on her brow, *and* a valenced state of cognitive pleasure not located on or in the body.

'Happiness' and 'sadness' provide the clearest examples of valency. The discussion, therefore, will restrict itself to these emotional states. Particular examples of happiness are triumph, glee, joy, ecstasy, gladness, love; and of sadness, despair, disappointment, grief, and sorrow. In a similar way to object-oriented programming languages, a preliminary taxonomy can be constructed in which 'happiness' and 'sadness' define classes of emotions with *valence* and *intention* slots. Particular instances of these generic emotions have additional slots that define finer-grained attributes. In this way, an 'isa' hierarchy of sub-types of the emotions 'happiness' and 'sadness' is formed (see Ortony, Clore & Collins, 88 for a similar treatment). Consequently, the terms 'happiness' and 'sadness', in this paper, refer to generic, abstract definitions. For the sake of brevity, it will be stated that a person is 'happy' when a desired goal is believed to have been achieved, and 'sad' when it is believed that failure has occured in achieving a desired goal. This is an oversimplification: concrete emotional states are rarely this straightforward or as simple.

Unlike the intentional component of an emotional state, valency does *not* represent a state of affairs. It can differ qualitatively in only a very restricted sense, i.e. it can be *either* pleasurable or displeasurable, and allow quantitative degrees of *intensity*: the valency can be very displeasurable or only mildly so. From a phenomenological perspective, valency is a 'brute fact'[2] of one's present state, unlike beliefs that can be true or false, or goals that have been achieved or not.

A preliminary definition of valency can now be provided: *Valency* is a form of pleasure or displeasure not located in the body, and is *a* non-intentional component of occurrent emotional states of happiness or sadness.

Our problem is to provide a theoretical account of valency.

# 3 A design-based answer: thermostats and classifiers

There are almost as many theories of emotion as emotion theorists. The field, as has often been remarked, is characterised by terminological confusion (Kagan 78, Read & Sloman 93) and riven by differing 'schools of thought' (Pfeifer, 92).

Approaches to the study of emotions can be very broadly categorised as *semantics*-based, *phenomena*-based and *design*-based (Sloman, 92). Semantics-based theories analyse the use of language to uncover implicit assumptions underlying emotion words (e.g., Wierzbicka, 92). Phenomena-based theories assume that emotions are a well-specified category and attempt to correlate contemporaneous and measurable phenomena with the occurrence of an emotion, such as physiological changes (an early example is William James' theory – see

[2]This term borrowed from (Chalmers, 96).

Calhoun & Solomon, 84; for a comprehensive review of many phenomena-based theories, see Strongman, 87).

The design-based approach, in contrast, takes the stance of an *engineer* attempting to build a system that exhibits the phenomena to be explained, and is a 'rational reconstruction' of the practice of Artificial Intelligence, considered as the general science of intelligent systems, both natural and artificial (Sloman, 93b). The methodology involves exploring an abstract space of possible requirements for functioning agents (*niche-space*) and the space of possible designs for such agents (*design-space*) and the mappings between them (Sloman, 95). Research strategies vary: they may be top-down, bottom-up or middle-out. All are potentially useful.

Research within the field of Artificial Life is an example of the design-based approach, and is characterised by the investigation of complete agents that integrate many capabilities, such as sensing, action selection, acting and (particularly) adaptation to a simulated or real niche. A consequence of such a methodology is that the analytic isolation of emotions from other cognitive phenomena, often characteristic of semantic and phenomena-based approaches, can be directly avoided by the investigation of complete systems. This methodological course is likely to bear the most fruit and have the most relevance for unified theories of cognition.

The development of negative feedback control systems demonstrated that simple, materially embodied systems can have sub-states with different functional roles, in particular 'belief-like' and 'desire-like' control sub-states (Sloman 93a, Powers 88, Braitenburg 84). For example, the belief-like sub-state of a thermostat is the curvature of its bi-metallic strip, which alters in accordance with the ambient temperature of a room; the desire-like sub-state is the setting of the control knob. Negative feedback ensures that the temperature of the room stabilises around the control knob setting, i.e. the thermostat 'acts' in the world to achieve its 'desire'. Of course, a thermostat does not have sufficient architectural complexity required to support human beliefs and desires, but it is an illustrative 'limiting case' (Sloman, 93a).

The simple thermostat implements a function that maps an input temperature to an output signal, which controls a heater. It does not learn. We can move through design-space adding architectural complexity to the negative feedback loop, such as varying the kinds of sub-states, the number and variety of sub-states, the functional differentiation of sub-states, and the kinds of causal influences on sub-states, such as whether the machine can change its own desire-like states, and so on (see Sloman, 93a for an extended discussion).

Holland's classifier system (Holland 95, Holland 75, Holland et al. 86, Riolo 88) is one of a class of relatively well-understood machine learning algorithms. Unlike the thermostat, it is sufficiently complex to exhibit prototypes of 'emotion-like' states, much as the thermostat exhibits prototypes of 'belief-like' and 'desire-like' sub-states. An analysis of its functioning reveals an important role for non-intentional representations. First, a brief description of a classifier system.

## 3.1 The learning classifier system

A classifier system consists of a *performance system*, and *credit-assignment* and *rule discovery* algorithms.[3] The performance system consists of a *classifier list* that consists of a set of condition-action rules called *classifiers*, a *message list* that holds current messages (like the working memory of production systems), an *input interface* that provides the classifier system with information about its environment in the required form, and an *output interface* that translates action messages into world events. The *basic cycle* of the performance system matches messages in the message list (including sensory messages) with classifiers, which then post their actions 'back' to the message list. Many classifiers may become active and fire in parallel. Any current action messages are sent to the output interface. The performance system is computationally complete; that is, any computable function can be implemented as a collection of classifiers.

The performance system alone cannot learn. The credit-assignment algorithm in the classifier system is a *bucket-brigade* (*bb*) algorithm. This algorithm introduces competition between classifiers based on a quantitative 'strength'. Each classifier that has its condition activated by a message *bids* to post its action part to the message list. Only the highest bidders are allowed to post their actions. The bid of a classifier depends on its strength[4], which is a measure of the classifier's 'usefulness' to the system. The higher the strength of a classifier the more likely it will win the competitive bidding round and post a message. The behaviour of the classifier system, therefore, can be modified by changing the strengths associated with individual classifiers. If the strength of the classifiers that tend to lead to 'useful' behaviour can be increased, and the strength of the classifiers that tend to lead to 'useless' behaviour can be decreased, the system will learn to produce more useful behaviour. The *bb* is designed to bring about these types of changes in strength.

The basis for the *bb* is information from *reinforcement mechanisms* about whether the classifier system as a whole is behaving correctly. This is achieved via *rewards*, i.e. the system will receive positive reward when it behaves correctly and negative reward when it behaves incorrectly. More often than not neither positive or negative reward will be received. This is a type of *reinforcement learning* (for a survey, see Kaelbling et al., 95), which derives its name from behaviourist theories of animal learning (e.g., Mackintosh, 83). The phrase 'credit assignment' will be preferred over 'reinforcement learning': the former emphasises internal mechanisms, whereas the latter is often associated with external operations.

When a reward is received the *bb* adds the reward value to the strength of all classifiers currently active, thereby changing the strength of classifiers *directly* associated in time with useful behaviour. Also, when a classifier is activated it 'pays' the amount it bid to the antecedent classifier that produced the message it matched. The strength of the active classifier is decreased by its bid amount. In this way, the *bb* acts to increase the strength of classifiers *indirectly* involved in the production of useful behaviour. This is a partial solution to the *temporal credit assignment problem*, which is the problem of determining

[3] This specification abridged from (Riolo, 88). Readers unfamiliar with classifier systems should consult (Holland, 95).

[4] This is a simplification for the sake of brevity. The bid of a classifier can depend on at least three factors: strength, the 'specificity' of the classifier, which is a measure of its relevance to a particular set of messages, and 'support', which allows internal messages to have differential importance.

which antecedent rules to strengthen given that there can be long delays between antecedent classifiers and the resultant rewarding act. The *bb* allows reward to 'circulate back' through the system (in a similar way to the 'back-propagation' of an error signal in artificial neural networks). Chains of high strength classifiers, performing useful computation, can emerge from such a scheme.

An alternative name for the quantitative measure associated with each individual classifier is *value*. This switch in terminology can be better understood by recalling that the *bb* was originally inspired by an economic metaphor (Holland et al., 86), in which classifiers are agents (consuming and producing messages) who possess a certain amount of money ('strength' or value) which they exchange for commodities at the market (the message list or blackboard). Much as money mirrors the flow of commodities in a simple commodity economy, *value mirrors the flow of messages* in the learning classifier system.

The *bb* can lead to improved system behaviour through the selection of some classifiers over others; however, it cannot create entirely new classifiers. A genetic algorithm (*ga*) implements rule discovery. Periodically, classifiers of high strength are selected as parents. The genetic operators of *crossover* and *mutation* are applied to classifiers considered as chromosome strings. The resultant offspring are placed in the classifier list. Normally, the *ga* is applied less frequently than the *bb* otherwise new classifiers will not have had sufficient time to be evaluated.

To summarise: within a classifier system *selection on value* occurs twice: in bidding rounds of classifiers competing for messages, and in rule discovery where high value classifiers generate offspring and low value classifiers are eventually removed from the system altogether. The combined effects of *circulation of value* via the *bb* and a double selection on value by competition and rule discovery allows the classifier system to reallocate classifier rules from unrewarding to rewarding processing. This ability makes it an adaptive system.

The classifier system, when embedded in a simulated or real environment, can construct classifiers that produce satisficing behaviour given the constraints and guidance of reinforcement mechanisms. It has been extensively used in the ALife community (Steels, 94), for example in developing control programs for robots through supervised learning (Dorigo & Colombetti, 93).

## 3.2 The intentional component of classifier states

The internal state of an implemented and *running* classifier system is continually changing. We will denote the classifier system's internal state at time step $t$ as $C_t$, and define it as the joint operation of the basic cycle and *bb* considered as one indivisible moment. $C_t$ has two distinguishable components: *intentional* and *non-intentional*.

The intentional component of $C_t$ is the message list containing messages with representational content. The simplest example of the representational content of messages can be found in (Holland, 95). We can imagine an artificial frog embedded in an environment of real or simulated flies. The control program for *simfrog* is a classifier system with an adapted set of classifiers. The *simfrog* has an eye sensor, which forms part of the classifier system's input interface. The eye can detect a number of attributes of any fly within range. Attributes could include whether the fly is moving, what colour it is, its size and proximity.

If a fly is detected the eye sensor posts a message to the message list that encodes this information. This sensory message is the result of a *mapping* between a state of the environment and a sub-state of *simfrog*. It is in this sense that messages have representational content. Internal messages, less directly linked to sensing or acting, will have more complex representational roles within the system. The *semantics* of messages depends on the dynamic relationship between message and environment. For example, the sensory message may match a classifier that posts an action message that results in *simfrog* throwing its sticky tongue in the direction of the detected fly. The meaning of the message, therefore, would be an impoverished version of 'eat that fly!'.

## 3.3 The non-intentional component of classifier states

The non-intentional component of $C_t$ is the circulation of value between antecedent classifiers, messages and matching classifiers. Value is exchanged for messages, i.e. matching classifiers pay the 'owners' of messages an amount of value, via the *bb*. Unlike messages, the value that circulates has no representational content. The values associated with classifiers specify a probabilistic partial ordering on the classifier list. The ordering is partial because only some classifiers will bid for the same message. The ordering is probabilistic because the classifier selection mechanism is stochastic.

For example, both classifiers $\mathrm{c}_i$ and $\mathrm{c}_j$ match message m; $\mathrm{c}_i$ has value $\mathrm{v}_i$, and $\mathrm{c}_j$ has value $\mathrm{v}_j$, with $\mathrm{v}_i > \mathrm{v}_j$. In competitive bidding for message m, $\mathrm{c}_i$ will out bid $\mathrm{c}_j$ more often than not, i.e. there is a probabilistic total ordering on the set $A = \{c_i, c_j\}$. Other classifiers in the classifier list, such as $\mathrm{c}_k$, *never* compete against any $c \in A$; consequently, no ordering holds between them[5]. Classifier values, therefore, specify a partial and probabilistic relation of utility over the classifier list. Value is internal economy alone. It does not represent anything within or external to the classifier system; rather, it specifies a *relation* between classifiers. It is this property of value that helps to make the classifier system a *domain-independent* learning algorithm: the representation of utility does not alter from domain to domain.

The classifier system tightly integrates impoverished conceptions of cognition, conation and affect.

The cognitive engine is the performance system consisting of condition-action rules and a global blackboard (cf. the production system and working memory of Newell's SOAR architecture).

Conation, or 'motivational force', is also represented in the classifier system. The value of a classifier is its dispositional and relative ability to fire and post a message. Whether the implementation of the classifier system is truly parallel (with perhaps separate processors for each classifier), or only simulates parallelism, *the value of a classifier is an ability to buy processing power.* A high value classifier will be more likely to win bidding rounds, be processed, and post its action part. An internal sensory message, for example a message that *simfrog*'s energy is below a danger threshold, may match the first in a chain of high value classifiers that instigate a search for flies in the environment. The high value of such

[5] In classifier systems with maximum size message lists the situation can become more complex, as classifiers compete for space in addition to competing for messages.

a processing chain will make it unlikely that other rules will out bid and switch processing to other ends. In this impoverished sense, value is motivational force.

Finally, the operation of the bucket-brigade, which alters the 'buying power' of sets of classifiers, involves losses or gains of value that are ultimately derived from reinforcement mechanisms, i.e. information regarding the 'goodness' or 'badness' of agent-environment situations. Positive and negative reinforcement are often linked to pleasure and pain.

## 3.4 Self-monitoring of credit-assignment

We now add a simple 'self-monitoring' mechanism to the classifier system. The mechanism is required to monitor the circulation of value and send its output to a suitable device. For current purposes self-monitoring need play no functional role within the classifier system, so the device can be a computer screen that displays results to an ALife engineer.

$C_t$ involves a *net* exchange of value, denoted $V_t$, from matching classifiers to antecedent classifiers. The self-monitoring mechanism records each $V_t$ over a specified time period, say $t = 1 \dots n$, and displays the *change* in value, denoted $\delta V_t$, which is exchanged from one time step to the next, where $\delta V_t = V_{t+1} - V_t$. $\delta V_t$ can be either:

- Positive, implying (a) a net increase in the utility of antecedent classifiers, and (b) currently active classifiers are likely to lead to positively rewarding consequences;
- Negative, implying (a) a net decrease in the utility of antecedent classifiers, and (b) currently active classifiers are likely to lead to negatively rewarding consequences; or
- Zero, implying no net change in the utility of antecedent classifiers.

Therefore, the self-monitoring of the non-intentional component of $C_1 \dots C_n$ will display a rate of change of value with both sign and magnitude. We now connect the output of self-monitoring to *simfrog*'s skin, which can change colour. If $\delta V_t$ is zero *simfrog* remains *green*, if $\delta V_t$ is positive he displays *yellow* with an intensity $|\delta V_t|$, and if $\delta V_t$ is negative he displays *blue* with intensity $|\delta V_t|$. When *simfrog* catches and eats a fly he will blush bright yellow as innate reinforcement mechanisms strongly positively reward antecedent classifiers. If *simfrog* possessed more sophisticated reflective capabilities he might wonder why he has beliefs that refer *and* an odd quantitative intensity that is either positive or negative but doesn't seem to be 'about' anything or serve any apparent purpose. Depending on philosophical prejudice, one might be tempted to say that *simfrog feels* happy, sad or indifferent depending on circumstance.

Some caveats are in order. A classifier system is limited in many ways. It does not have an explicit memory store. It tends to be an entirely reactive system with no representation of goals. It does not anticipate, or perform prior search within a world model before acting. In real-world applications it can be difficult for a classifier system to learn appropriate behaviours (Wilson & Goldberg, 89). Also, an improved domain-independent learning algorithm may be discovered tomorrow: the classifier system is unlikely to the be the last word on the subject. However, by abstracting from implementation details we can examine certain *design principles* embodied in the classifier system. The following section examines the implications of such design principles for information processing theories of human emotions.

# 4 Circulation of value

There are two kinds of finite computational[6] resource in the classifier system: (a) the total *information* capacity of the performance system, which consists of the set of 'if-then' rules, a (usually fixed) number of classifiers that perform simple computations, and (b) *processing* limits, which is the amount of parallel computation allowed per time step; for example, a maximum of ten classifiers may be allowed to fire during each basic cycle. For current purposes, information limits will be placed to one side.

## 4.1 Processing limits: emotions as interrupts

Existing information processing theories of emotion agree on the importance of *processing* resource limits in accounting for emotional states. For example, Simon's 'interrupt theory' associates emotional states with the interruption of a resource-bound, high level, attentive system due to new, perhaps urgent motives: 'The theory explains how a basically serial information processor endowed with multiple needs behaves adaptively and survives in an environment that presents unpredictable threats and opportunities. The explanation is built on two central mechanisms: 1. A goal-terminating mechanism [goal executor] ... [and] 2. An interruption mechanism, that is, emotion allows the processor to respond to urgent needs in real time.' (Simon 67, Simon 82). An interrupting stimulus, such as the presence of a predator, disrupts ongoing goal processing in the serial processor and substitutes new goals to deal with the new situation producing, amongst other things, emotional behaviour, e.g. the flight-fight-fright response.

Aaron Sloman's attention filter penetration theory (Sloman & Croucher 81, Sloman 87, Sloman 92) extends the interrupt theory by introducing new, architectural detail implicit in Simon's paper. Two types of processing are distinguished: pre-attentive, highly parallel and automatic motive generation processes, and attentive, resource-bound 'motive management' processes that exhibit a limited degree of parallelism. The concept of *insistence* is introduced, which is the dispositional ability of a motivator to 'surface' through a variable threshold filter and disrupt attentive processing. The *intensity* of a motivator is its ability to 'keep hold' of attention once surfaced. A common characteristic of many emotional states is the phenomenon of *perturbance* (Beaudoin, 94), which occurs when a motive has been postponed or rejected but nevertheless keeps resurfacing and disrupting ongoing, motive processing. The concept of perturbance has been extensively used to provide an architectural account of grief or 'loss' (Wright, Sloman & Beaudoin, 96). As in Simon's theory, 'emotional' interrupt mechanisms are required to overcome finite processing resources.

Both interrupt theories rely (implicitly in Simon's case) on a distinction between *control* and *semantic* signals in information processing architectures. Semantic signalling is the propagation of information that has representational content, whereas control signalling does not refer or have semantic content but performs a control function, such as changing the control flow of the system, or putting it into a distinct kind of processing state[7]. It is

[6] A classifier system animat embedded in an environment will also have 'physical' resource constraints, such as the number and kind of effectors and detectors.

[7] The following analogy may help capture the distinction. Imagine trains travelling on a complex network of tracks. Postal trains contain mail (semantic content) with destination addresses on the envelopes. These trains travel to the destinations and deposit the mail (the information). However, a different kind of train,

the non-representational nature of control signalling that is held to account for the non-intentional, 'feeling' component of emotional states. However, a theoretical explanation of valency is absent: reasons why some control signals are pleasurable or displeasurable, i.e. possess a qualitative dimension of valency that can be either positive or negative, and also vary in quantitative intensity, are not provided.

Keith Oatley and Philip Johnson-Laird's (Oatley 92, Oatley & Johnson-Laird 85, Johnson-Laird 88) communicative theory addresses this question by introducing basic, irreducible and phylogenetically older architectural control signals. 'Each goal and plan has a monitoring mechanism that evaluates events relevant to it. When a substantial change of probability occurs of achieving an important goal or subgoal, the monitoring mechanism broadcasts to the whole cognitive system a signal that can set it into readiness to respond to this change. Humans experience these signals and the states of readiness they induce as emotions' (Oatley, 92). Control signals, of which 'happiness' and 'sadness' are examples, communicate significant junctures of plans to other cognitive subsystems. Emotional states, therefore, are viewed as a design solution to certain problems of the transition between plans in systems with multiple goals. The functional role of valenced signals is to enforce state transitions by interrupting a central processing system; for example, the 'sadness' control signal, broadcast when a major plan fails, causes a state transition to search for a new plan. On this view, control signals differ in valency because they differ in their functional roles.

All three theories lack a consideration of *adaptive* architectures, i.e. architectures able to modify themselves to improve their behaviour. Explicitly considering adaptation provides a new functional role for control signalling.

## 4.2 Adaptation: changes in the ability to buy processing power

The phrase *circulation of value* denotes a general design principle, and refers to any mechanism that (i) alters the dispositional ability of sub-agents of a complex system to gain access to limited processing resources, via (ii) exchanges of a quantitative, domain-independent representation of utility that mirrors the flow of agent products. Circulation of value is a type of credit-assignment. A particular example of circulation of value is the bucket-brigade algorithm, where a 'sub-agent' is a single classifier, 'complex system' is the set of competing and cooperating classifiers, and 'agent products' are messages.

The circulation of value is a pattern of flow of control signals. Such signals have no semantic content and propagate around the system altering control flow. Additionally, classifier bids attempt to 'grab' processing resources and may 'interrupt' current processing causing internal or external behaviour to take different routes. A quantitative representation of utility, therefore, need not contradict the interrupt function identified by Simon and Sloman, nor the control signal function identified by Oatley and Johnson-Laird. The local exchange of value between classifiers can generate the negative and positive control signals of the communicative theory: instead of two signals, we now have one (see section 3.4). This is a more parsimonious state of affairs. But, additionally, there is a new, previously

a 'control signal' train, can travel through the network altering the points of the tracks. This has the effect of changing the topology of the network, i.e. trains will continue to deposit their mail but will use different routes.

unidentified functional role for the control signal: the circulation of value implements a type of *adaptation*. This is not inductive learning of new hypotheses about a domain, but an ordering and reordering of the utility of control sub-states to dispositionally determine behaviour. In a classifier system, the circulation of value adaptively changes the ability of classifiers to buy processing power.

The design principle of circulation of value opens up the possibility of architectures that generate valenced states far removed from physiology. High level cognitive processes may be saturated with value, allowing the production of semantic messages coupled with losses or gains in quantitative value. For example, negatively valenced states, such as grief, may, amongst other things, involve the gradual loss of the accumulated value of a structure of attachment (Bowlby 79, Bowlby 88, Wright et al. 96) towards a loved one. Negative valency may be a necessary consequence of adaptive change: the structure of attachment, no longer useful for motive generation, loses its ability to buy processing power, grab attentive resources and dispositionally determine behaviour. This process is self-monitored as displeasurable.

These considerations lead to a *design hypothesis* of a quantitative representation of utility that circulates within a subset of cognition. The next step is the construction of a circulation of value theory of emotions that builds on previous work. This will require another iteration of the design-based approach. The design-space of autonomous agent architectures that integrate circulation of value learning mechanisms with more complex forms of motive management needs to be explored. The capabilities, properties and explanatory power of such designs can then be examined. Further comparisons can then be made between designs and psychological phenomena. Preliminary efforts in this direction are reported in (Wright, 95). In addition, it will be necessary to investigate the results and theories of neuropsychology in order to map the postulated mechanisms onto the neural substrate. If some form of circulation of value is found to occur in human brains this would add credence to such a theory. However, if the further exploration of adaptive architectures finds no use for circulation of value mechanisms then the hypothesis will need to be modified or replaced: ultimately, it is an empirical question.

However, even at this early stage it is possible to make some theoretical claims. Given the above considerations, and our preliminary definition of valency, we can define a corresponding architectural process that gives rise to it: *valency is a self-monitored process of credit-assignment*. Circulation of value mechanisms are necessary preconditions for the hedonic, or 'feeling' component of many emotional states. Self-monitoring mechanisms, of varying sophistication, begin to constitute sufficient conditions for such internal states. (A type of) 'feeling' is (a type of) adaptation, and 'hot' cognition – forms of pleasure or displeasure involved in short term and long term control – need pose no insurmountable problems for information processing theories of emotion.

## 5 Towards emotional animats

The emotions encompass a broad range of human experience, whereas valency is a component of only some emotional states. To approach a real understanding of human emotions will require an investigation of more sophisticated architectures satisfying more complex requirements. This is long-term research. Yet simple animats already exhibit simple 'emotion-like' states, as long as we take care over definitions and avoid hyperbole. 'Broad but shallow' agent designs can be illuminating. For example, (Balkenius, 95) provides a clear discussion on relations between simple motivations and simple emotions based on experiments with animats. Also, the simulation of societies of competing and cooperating adaptive agents will impose new requirements on architectures and give rise to new kinds of internal state. (Aube & Senteni, 95) view (more complex) emotional states as 'commitment operators' that manage resources in multi-agent systems.

An important difference between such design-based (e.g., see Beaudoin & Sloman 93, Beaudoin 94, and Sloman, Beaudoin & Wright 94) approaches and previous computer simulations of emotions is that they move from requirements for *complete agents* to possible designs, and do not directly 'program in' correlates of emotions. In this way, design features are linked to niche features providing explanations of *why* emotional states are present in nature. A small example is provided in this paper: a requirement for adaptability entails, at some level, credit-assignment mechanisms that use a domain-independent representation of utility or value, a kind of internal 'common currency'. Such a representation does not refer but is relational, and can be gained or lost depending on whether actions are successful or unsuccessful in leading to rewarding consequences. These kinds of processes are self-monitored as non-intentional states, or 'feelings', which are quantitative in nature and either pleasurable or displeasurable.

Those philosophically inclined may doubt that animats constructed in this way 'really' experience their non-intentional states. Is it *anything to be like* a simulated animat? This kind of question is experimentally undecidable, and has no engineering consequences whatever. It is therefore scientifically uninteresting.

# 6 Conclusion

A subset of emotional states was examined and the concept of valency – the non-intentional component of occurrent emotional states of happiness or sadness – was introduced. Valency is a preliminary definition of a subset of the states colloquially referred to as 'feelings'. The internal state of a complex adaptive system, the learning classifier system, able to function as a complete agent within a niche, was examined. Just as the simple thermostat exhibits prototypes of 'belief-like' and 'desire-like' sub-states, the classifier system was found to exhibit simple examples of valenced states.

The design principle of circulation of value was abstracted from the classifier system and employed to overcome existing inadequacies in information processing theories of emotion. The circulation of value mechanism is more parsimonious and more general than Oatley and Johnson-Laird's control signalling, and, more importantly, it introduces a basic form of adaptation. These new considerations build upon and do not contradict previous theories. A theoretical conclusion is that, to a first approximation, valency is a self-monitored process of credit-assignment. 'Feeling' is the self-monitoring of adaptation; that is, the non-intentional component of generic 'happiness' and 'sadness' states includes a movement of internal value, which functions to alter the dispositional ability of control sub-states to buy processing power and determine behaviour. Such a process has a 'brute' feeling component because value does not refer, unlike beliefs that can be true or false, or goals that can be achieved or not.

To approach the complexity of concrete, human emotional states will require many more iterations of the design-based approach.


# 7 Acknowledgements

My thanks to Aaron Sloman, Chris Complin, and members of the Cognition and Affect[8] and EEBIC[9] (Evolutionary and Emergent Behaviour Intelligence and Computation) research groups at Birmingham.


# 8 References


Aube, Michel & Senteni, Alain (1996). Emotions as commitments operators: a foundation for control structure in multi-agents systems. In *Proceedings of the 7th European Workshop on Modelling Autonomous Agents in a Multi-Agents World, MAAMAW'96*, Lecture Notes on Artificial Intelligence, No. 1038, Springer-Verlag.

Balkenius, C. (1995). *Natural intelligence in artificial creatures.* Lund University Cognitive Studies 37.

Bates, J., Loyall, A. B., & Reilly, W. S. (1991). Broad agents. Paper presented at the *AAAI spring symposium on integrated intelligent architectures.* Stanford, CA: (Available in SIGART BULLETIN, 2(4), Aug. 1991, 38-40).

Beaudoin, L. P. (1994). *Goal Processing in Autonomous Agents.* PhD Thesis, School of Computer Science, University of Birmingham.

Beaudoin, L. P. & Sloman, A. (1993). A study of motive processing and attention. In *Proceedings of AISB93*, A. Sloman, D. Hogg, G. Humphreys, A. Ramsay & D. Partridge (Eds), 229-238, Oxford: IOS Press.

Bowlby, J. (1979). *The making and breaking of affectional bonds.* Tavistock Publications Ltd.

Bowlby, J, (1988). *A secure base.* Routledge.

Braitenburg, V. (1984). *Vehicles – experiments in synthetic psychology.* The MIT press.

Calhoun, C., & Solomon, R. C. (Editors) (1984). *What is an Emotion? Classical Readings in Philosophical Psychology.* Oxford University Press.

Chalmers, D. (1996). *The Conscious Mind.* Oxford University Press.

Dorigo, M. & Colombetti, M. (1993). Robot shaping: developing situated agents through learning. Technical report TR-92-040 (revised), International Computer Science Institute, Berkeley, CA.

Dyer, M. G. (1987). Emotions and their computations: Three computer models. *Cognition and Emotion* 1(3), 323-347.


[8]Information at ftp: //ftp.cs.bham.ac.uk /pub /groups /cog_affect/ -WWW-INTRO.html.
[9]Information at http: //www.cs.bham.ac.uk/ rmp/ EEBIC.html.

Kagan, J. (1978). On emotion and its development: a working paper. In *The Development of Affect* edited by M. Lewis and L. A. Rosenblum. Plenum Press, New York and London.

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

Green, O. H. (1992). *The Emotions*, a philosophical theory. Kluwer Academic Publishers.

Holland, J. H. (1975). *Adaption in natural and artificial systems.* MIT Press.

Holland, J. H. (1995). *Hidden Order, How Adaptation Builds Complexity.* Helix Books.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction: processes of inference, learning and discovery.* MIT Press.

Johnson-Laird, P. N. (1988). *The Computer and the Mind: An Introduction to Cognitive Science*, Fontana.

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1995). Reinforcement learning: a survey. In *Practice and Future of Autonomous Agents*, Vol. 1. Centro Stefano Franscini, Monte Verita, Ticino, Switzerland.

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). SOAR: an architecture for general intelligence. *Artificial Intelligence* 33, 1-64.

Mackintosh, N. J. (1983). *Conditioning and associative learning.* Oxford Psychology Series No. 3, Oxford University Press.

Minsky, M. (1987). *The society of mind.* Picador.

Newell, A. (1990). *Unified theories of cognition.* Harvard University Press, Cambridge, MA, 1990.

Oatley, K. (1992). *Best laid schemes, the psychology of emotions.* Studies in Emotion and Social Interaction, Cambridge University Press.

Oatley, K. & Johnson-Laird, P. N. (1985). Sketch for a cognitive theory of emotions, Cognitive Science Research Papers CSRP 045, University of Sussex.

Ortony, A., Clore, G., & Collins, A. (1988). *The cognitive structure of emotions.* Cambridge University Press.

Pfeifer, R. (1992). The new age of the fungus eater: Comments on artificial intelligence and emotion. Extended and revised version of an invited talk at the *AISB-91* conference in Leeds, U.K.

Powers, W. T. (1988). *Living Control Systems* – selected papers of William T. Powers. The Control Systems Group, Kentucky.

Read, T., & Sloman, A. (1993). The terminological pitfalls of studying emotion. Paper presented at the *Workshop on Architectures Underlying Motivation and Emotion - WAUME 93*, Birmingham.

Riolo, R. L. (1988). CFS-C: A package of domain independent subroutines for implementing classifier systems in arbitrary, user-defined environments. Logic of Computers Group, Division of Computer Science and Engineering, University of Michigan.

Ryle, G. (1949). *The Concept of Mind*, Hutchinson.

Simon, H. A. (1967). 'Motivational and Emotional Controls of Cognition', reprinted in *Models of Thought*, Yale University Press, (1979) 29–38.

Simon, H. A. (1982). Comments, In M. S. Clark & S. T. Fiske (Eds.), *Affect and cognition.* Hillsdale, NJ: Lawrence Erlbaum Associates, 1982.

Simon, H. A. (1995). Aritificial Intelligence: an empirical science. *Artificial Intelligence* 77 (1995) 95-127.

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

Sloman, A. (1992). Prolegomena to a theory of communication and affect, *Communication from an Artificial Intelligence Perspective: Theoretical and Applied Issues*, A. Ortony, J. Slack, and O. Stock, eds. Springer: Heidelberg, 229 – 260.

Sloman, A. (1993a). The mind as a control system. In *Proceedings of the 1992 Royal Institute of Philosophy Conference 'Philosophy and the Cognitive Sciences'*, (eds) C. Hookway and D. Peterson, Cambridge University Press.

Sloman, A. (1993b). Prospects for AI as the general science of intelligence. In *Prospects for Artificial Intelligence – Proceedings of AISB93*, Oxford IOS Press, Editors: A. Sloman, D. Hogg, G. Humphreys & D. Partridge.

Sloman, A. (1995). Exploring design-space and niche-space. In *Proceedings 5th Scandinavian Conf. on AI*, Trondheim May 1995, IOS Press, Amsterdam.

Sloman, A., Beaudoin, L.P., & Wright, I. P. (1994) Computational modeling of motive-management processes, *Proceedings of the Conference of the International Society for Research in Emotions*, Cambridge, July 1994. (ed) N.Frijda, ISRE Publications. 344-348.

Sloman, A., & Croucher, M. (1981). Why robots will have emotions. In *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*, (197-202). Vancouver.

Steels, L. (1994). The artificial life roots of artificial intelligence. *Artificial Life Journal*, Vol 1, 1. MIT Press, Cambridge.

Strongman, K. T. (1987). *The Psychology of Emotion.* Third Edition, John Wiley & Sons.

Wierzbicka, A. (1992). Defining emotion concepts. *Cognitive Science* 16, 539-581.

Wilson, S. W. & Goldberg, D. E. (1989). A criticial review of classifier systems. In *Proceedings of the Third International Conference on Genetic Algorithms* (pp. 244-255), Los Altos, California: Morgan Kaufmann (1989).

Wright, I. P. (1995). *Cognition and Currency Flow*, notes toward a circulation of value theory of emotions. Unpublished research document, Cognitive Science Research Centre, University of Birmingham. (Available on request from the author.)

Wright, I. P., Sloman, A., & Beaudoin, L. P. (1996). Towards a design-based analysis of emotional episodes. To appear, with commentaries, in *Philosophy Psychiatry and Psychology.*