Multiple Agents With Emotional Behaviour in The Blocks World.

S. Allen and T.J.M. Bench-Capon
Department of Computer Science, University of Liverpool,
PO Box 147, Liverpool, UK. Tel (051)794 3697.
Email:tbc@uk.ac.liverpool.cs

A general architecture for the computer implementation of a theory of emotions is presented. The model is based mainly on a cognitive orientated theory proposed by Frijda 1986 where emotions are described as the manifestations of a concern satisfaction system. The model applies the theory to multiple agents in a blocks world environment where actions may or may not be emotional. Related work is also discussed.

Introduction.

This paper will describe a system which will exhibit emotional behaviour. We want this to be shown not in the interaction between a human and a computer, where the emotional concerns of the computer could do no more than introduce inefficencies, but in the interactions between a ‘‘society’’ of computer agents. We have selected the blocks world as being an environment which is at once tractable and rich enough to contain the range of behaviour necessary to show emotion. Such a system could be used to explain emotional behaviour and to model groups of agents where emotions may influence their plans.

Naturally, any study begins with subject definitions. With emotion, there is no general consensus on the exact definition of emotion. For example, some scientists and laymen consider ‘‘hunger’’ an emotion, others do not (Frijda 1986). The difficulties inherent in defining emotions include:

- □ The lack of a single criterion, or even a group of criteria, for their identification.
- □ The complexity and lack of knowledge on anatomical and physiological mechanisms involved.
- □ Problems due to the lack of a unified terminology.

Delgado 1973 reports that psychologists in general consider emotions to have two aspects:

(i) The state of individual experience or feelings. This state may be analysed and reported through introspection. The experiential state requires the individual judgement of a situation which depends on sensory inputs, past experience, neurophysiological mechanisms, mental attitudes and other aspects.

(ii) The expressive or behavioural aspect of emotions. This affects mainly the motor system, e.g. gestures; the autonomic system, e.g. respiration; and the endocrine glands, e.g. adrenals. Other effects can be recorded for intracerebral changes such as thermal and electrical activity.

There are a great number of theories on emotion. One approach that offers the clearest path towards information processing theories that may have computational applicability is that of cognition. The central idea, whether explicit or implicit, in the cognitive theories is the appraisal of a situation or stimulus. Thus interpretation of a situation will determine the presence and quality of an emotion. It can be described as a process of cognitive evaluation.

The Intended System.

The system described here is largely based on Frijda’s 1986 emotion theory. The approach is a cognitive orientated model of emotions. It also emphasises the expressive aspect of emotions without dwelling on introspective reports. The reason for this is that current systems use natural language or a subset of that to express their emotions (see *Related Work* section.) The proposed system will take the environment of the blocks world (figure 1) where two agents compete and/or cooperate to position blocks in a set pattern. Each agent may have differing tasks or the same task. This will have the advantage of locating the emotional behaviour in a ‘‘social’’ situation.

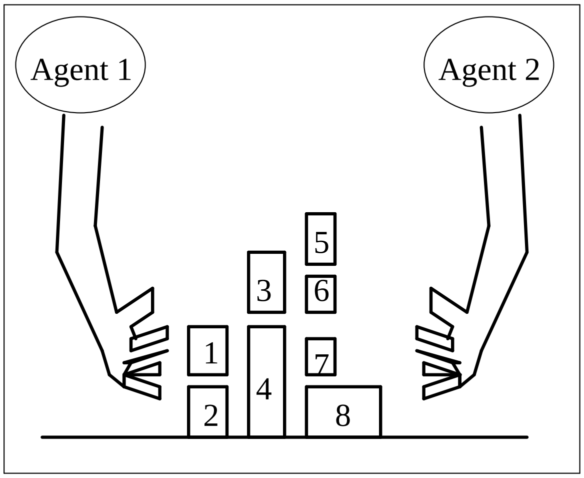

Figure 1. - The blocks world environment.

The interaction between the agents will allow scope for emotional phenomena. For instance, agent 1 may try to thwart agent 2's progress by dropping a block on it. Agent 2 will then react emotionally. In this environment, the two agents would not report their emotions verbally to one another but rather manifest them through physical behaviour. This will enable us to focus on a key aspect of emotional behaviour without encountering the well-known problems associated with ''faking'' natural language.

Characteristics of the Agents.

Besides the underlying emotion process the agent's interaction capabilities must be fully described. Two types of environmental interaction are defined: peripheral senses and physiological response.

Peripheral Senses.

Several aspects of awareness are outlined to provide an emotionally stimulating environment. Thus an agent should be aware of:

1) Whether or not it is grasping a block.
2) The position of individual blocks.
3) The position of agents including itself.
4) The speed of an agent including itself.
5) Whether or not an agent, including itself, is successful in grasping a block.
6) Whether or not an agent, including itself, is trembling.
7) Whether or not an agent, including itself, is crushed by a block.

Points 4) → 7) are important in the context of the next subsection.

Physiological Response.

Five groups of human physiological response can be distinguished:

- □ autonomic responses involving changes in functioning of the smooth muscles and other internal organs.
- □ changes in hormone secretions.
- □ changes in neural responses, e.g. such as those reflected in EEG readings.
- □ changes in the chemical composition of body fluids.
- □ muscle tension and overt movement.

It is the view of many theorists on emotion (see Schachter 1970; Arnold 1970; and Lazarus *et al.* 1970 for example) that physiological response is an essential part of any model. However, there is similarity between physiological response patterns in different emotions and in conditions such as hunger or cold (Cannon 1927). Similarly physiological actions may be unconnected with emotion such as increased heart rate after prolonged exercise.

It seems clear that some physiological responses may be emotional and some not. Thus a general action system is required in which an emotion system can be embedded.

Certain human physiological functions relating to emotions are difficult to map onto a computer agent's functions. Figure 2 is a brief mapping example, with the human functions taken from Delgado 1973.

Human	**Computer**
Heart rate	Change in the speed of response or action
Blood pressure and blood flow distribution	Change in energy for different parts of the agent
Respiration	Change in overall energy of agent.
Electrodermal activity and sweating	Change in friction attribute for movement or grasping of objects
Pupillary response	Change in ability to detect movement in the world.
Trembling	Trembling or shaking of various parts.
Temporal stability	Change in magnitude of response over time.
Response latency	Response latency.
Recovery rate	Recovery rate.
Muscle tension and tremor	Change in speed of response for agents parts.

Figure 2. - Mapping Examples For Physiological Response.

Actual motor movements such as a kick or punch are not included in the mappings because the computer equivalent of limb movement is obvious. Facial expressions are a combination of the above mappings along with jaw and eye movement. Facial expressions are assumed irrelevant to the computer agents.

It should be noted that Cannon 1927 regards such physiological response in emotion as being functional for preparation of active, energy-requiring response. Similarly Frijda 1986 suggests autonomic response 'is part of action readiness change at some stage of actual execution or preparation of actual overt (or cognitive) response, under conditions that such response is felt to need extra resources or adjustments.'

He concludes with 'physiological response, autonomic arousal included, is part of action readiness mode.'

Relating physiological response to the term non-goal directed is a misnomer. Physiological response is functional in the context of a person's intentional or goal directed actions in the world.

The Underlying Emotion Process.

Figure 3 represents an initial working definition of the emotion process. The emotion process being the sequence of steps leading from the stimulus to the phenomena.

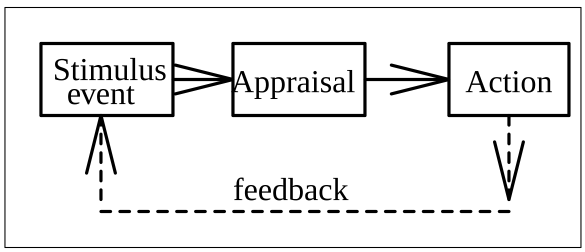

Figure 3. - Initial Definition of The Emotion Process.

Frijda's Model.

This model shares similar characteristics with other models. Namely

- Abelson 1983.
- Bower & Cohen 1981.
- Pfeifer & Nicolas 1981.
- Sloman & Croucher 1981.
- Toda 1982.
- Wegman 1985.

Frijda 1987 believes emotions are part of a system for realising concerns. Concerns are dispositions that the individual brings to the process. Concerns develop through experience. They are connected with the action readiness mode: that is, the level of alertness, the interestedness and eagerness for interaction. Concerns involve other individuals, specific environments, e.g. home, specific life goals, society values and objects. They are also tied up with behaviour systems of individuals, performance functions, monitoring of success of planned action etc.

Thus concern realisation tries to satisfy conditions under which the system can continue to function in the manner it was 'designed' for.

At first this may appear an incomplete definition of emotion. However, in matching a stimulus event to the individual's concerns, the process of appraisal is started. This process evaluates events as desirable or as undesirable; as an opportunity or as a threat; in terms of ability to cope; as a detailed plan of action etc.

Necessary Components of the Emotion Process.

The stimulus event should be thought of as a transaction. It is not a true event since it is continuous over time; is elicited by the subject involved; imposed upon him by others; and is affected by other stimuli. Lazarus & Folkman 1984 also explain that which actually elicits a given emotion may have been sought as well as encountered. Further it also develops in such a way that it becomes relevant and loses its relevance, varies in urgency and in quality. Thus, the situational meaning structure of the emotion stimulus must be continuously updated. This implies both feedback and moment to moment scanning of the environment even while other information is appraised.

Having established that dispositions, or concerns in Frijda's terminology, are an essential part of the emotion process there must be some method of detecting objects or events that interact with these concerns. This concern match/mismatch evaluation should be a continuous operation since the stimulus event is changing from moment to moment.

Given that concern interaction occurs there must be a process for providing action. This includes physiological behaviour as well as overt behaviour, i.e. movement. An action repertoire that encompasses social signals and fast operating emergency actions is required. There must also be a plan construction capability for novel situations.

The appraisal and action processes both involve feedback and provide information that affects all aspects of the emotion process. For example, if the appraisal determines that a particular action should be performed but the action process cannot perform that action then feedback will occur and a new action proposed by appraisal.

All this requires a general, overall monitoring of the emotion process, of its elicitation, its course, and its results. Monitoring needs central integration of all relevant information coming from different sources and moments. Such monitoring by centrally integrated information is provided by the 'blackboard control structure' (Erman *et al.* 1980)

The Blackboard Control Structure.

A blackboard architecture is suitable for interpretations that develop as multiple knowledge sources contribute their advice opportunistically and asynchronously so that interpretations gain support from the various sources. The partial interpretations produced by knowledge sources are called hypothesis elements. Typically, the input to a knowledge source is the output of another knowledge source. For example, a stimulus event might have to be identified cognitively by one expert before its emotional significance can be appraised by other experts. Several competing hypotheses may be under consideration simultaneously. However, attention eventually focuses on that hypothesis for which the evidence is greatest; the alternative hypotheses slowly fade from the blackboard.

The Intended Emotion Process.

Figure 4 represents the proposed system as a general architecture. The processes are derived from Frijda's 1986 misleading linear diagram (ref p. 454.)

Events from the blocks world or generated in thought are both affected by and affecting the *analyser*. Elements from the events that are deemed relevant to the system are continuously posted on the blackboard. The *analyser* will be notified if further information is required from the stimulus events.

The *comparator* then checks each of the concerns against the elements posted by the *analyser*. In theory, Frijda suggests this checking is parallel for each concern. In practice however, the concerns will be listed in priority order and checked in order. Swagerman 1987 cites Schank & Abelson's 1977 themas for defining concerns. The themas represent preconditions or triggers for action and should be regarded as dormant demons. General concepts are required for the themas that allow fuzzy matching (Rosch 1978, Kosko 1988 for example). Once a degree of match/mismatch is obtained the systems intention for action is stated. This is achieved through the coupling of themas to goals. It should be noted that one stimulus event may trigger the posting of several goals.

The *diagnoser* decides whether the goals can be realised. However knowledge of available actions must be known before this is possible. It does not infer a detailed plan of action rather a degree of similarity between known plans of action and the current goal (see Rich 1988 Sec. 8.1 for example).

The task of the *evaluator* is to decide whether or not to interrupt ongoing action/appraisal to process the new goal. This is achieved on a number of levels. If the *diagnoser* has not contributed evidence to the blackboard about the goal and the *comparator* indicates the goal is of a sufficiently higher-level than the ongoing goals then interruption will occur. This is the case for emergency situations, e.g. a mortal threat. However, if the situation is not an emergency then the *evaluator* will wait for information from the *diagnoser*.

The *action proposer* will develop a plan of action depending on the situation. For example, if an emergency action is required, information will be sent part-parcel to the blackboard as it is generated.

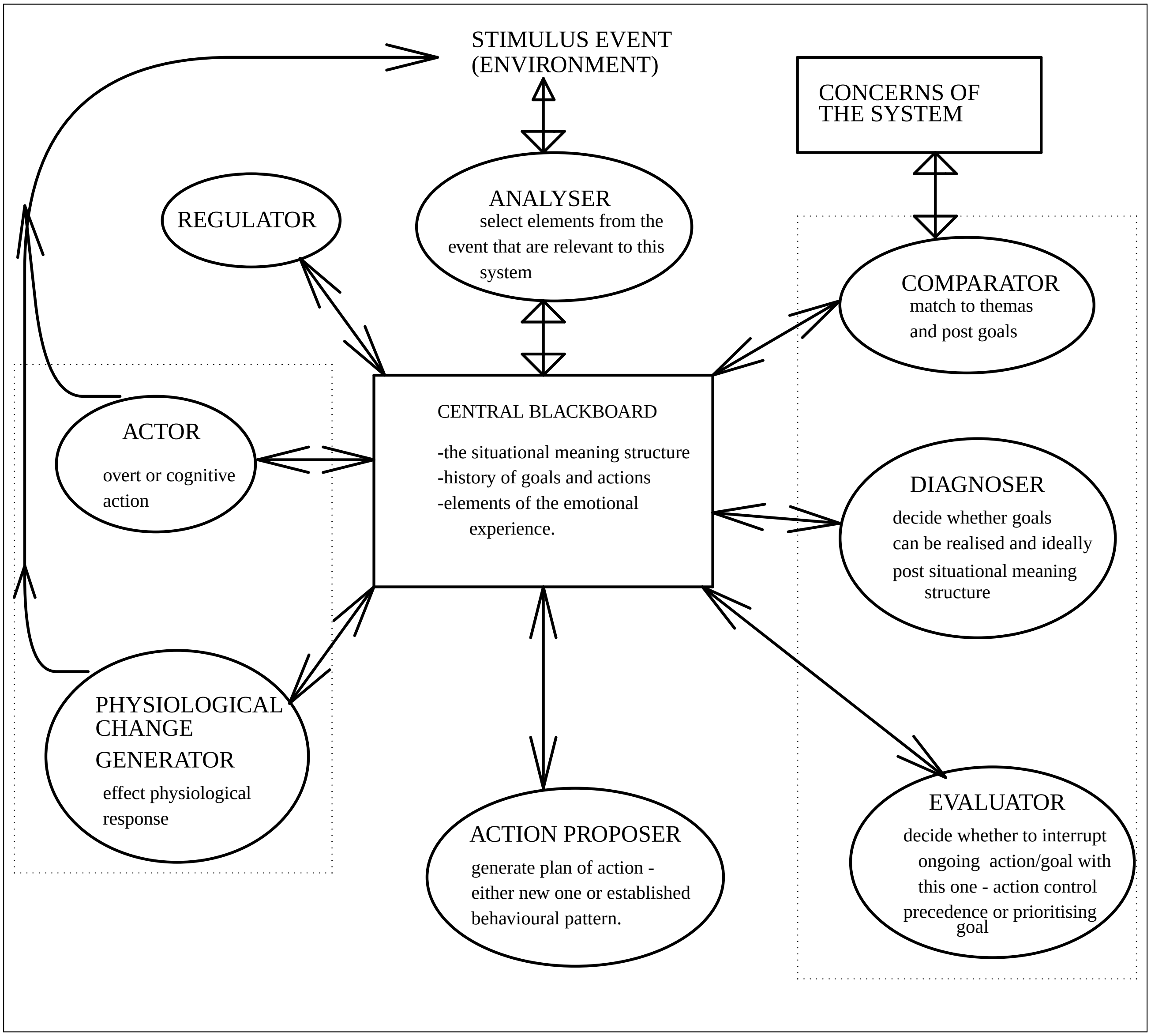

Figure 4. - Emotion Process Architecture.

Alternatively, if the concern is of low-level importance a whole plan may be developed before notification is made. The stock of action plans should be sufficient to cope with emergency situation and useful for developing new plans of action. It is advantageous for the system to record successes and failures of action plans - a learning system.

The *actor and physiological change generator* simply follow the prescribed plan and notify any problems in execution to the blackboard.

The *regulator* monitors information and interaction between concerns on the blackboard. This may involve the thwarting of certain plans of action as with inhibition in humans. The *regulator* may be considered the house-keeper of the blackboard.

Further Features.

Although the diagram shows the *concerns of the system* as indirectly communicating with the blackboard, this is not accurate. The concerns' activation points, i.e. preconditions, are changed dynamically according to the social and physical environment. To do this the *concerns of the system* would have to be connected to the blackboard.

Sloman and Croucher 1981 suggest motives - comparable to concerns - are dynamic in the following ways.

- Stimulus events constantly stimulate motives that are lying dormant into activity.
- Active motives are made dormant when their goal is achieved.
- Motives create subgoals if the situation to be achieved must be realised in stages.
- Motives must be adapted to the social and physical environment.

In the environment of the blocks world, interpretation and anticipation of an agent's moves are invaluable. The *analyser* should therefore be able to perceive threats or opportunities in the context of prediction. This will require both simple planning and extrapolation: simple because time is important.

Emotional Experience.

Many Authors on cognitive orientated theories of emotion view emotional experience as a reflection of the processing or evaluation of a stimulus (e.g. Leventhal 1974, Scherer 1982.) Elements from the evaluation are brought together to characterise a wide range of emotions. This attitude is termed an appraisal dimension approach. Frijda 1986 believes emotional experience consists of

> 'awareness of situational meaning structure, awareness of autonomic arousal or dearousal, and awareness of state of action readiness ... In addition, it consists of pleasure or pain ... awareness of control precedence, or its complement, the sense of being overwhelmed.'

Further Frijda uses twelve dimensions to characterise his set of thirty one emotion concepts. Concept examples are joy, fear and indignation. Dimension examples are: whether the situation was attractive and whether the course of events could be modified by the emotional agent.

Similarly, Dalkvist and Rollenhagen 1989 suggest nineteen dimensions. In both models values for the dimensions are evaluated from different parts of the theory. The values coalesce into the emotional experience which can then be labelled.

Testing the System.

Anticipation of testing requires the system to report its internal reasoning. By labelling specific behaviour, an operator may agree or disagree. Conversely, the operator may find it difficult to label behaviour with a specific emotional name - try it with a friend's behaviour.

Another aspect of testing will be the interaction between concerns. To this end, the system will be designed to allow concerns to be changed.

Related Work.

Colby 1981 implemented a simulation model called PARRY to explain the paranoid mode of behaviour. The program, however, does not have emotions it simply has conceptual representations: it simulates a patient, it is not a patient. The program communicates with an interviewer through a natural language interface.

Mueller & Dyer 1985 implemented a theory of daydreaming called DAYDREAMER. Like the proposed system it uses concerns that are triggered by stimulus events, and it uses goals attached to those concerns. It has aspects of control precedence; states of action readiness; and plans to achieve goals. It could be described as a concern realisation program.

It does not label emotional experiences but rather prints out a daydreaming stream of thoughts.

Swagerman 1987 implemented a concern realisation program called ACRES. This was based on Frijda's 1986 theory and involved a dialogue with an operator via a terminal. A simple language interface allows the operator to ask for or offer information. ACRES takes the role of an emotional expert in the domain of emotion knowledge. The system uses 6 concerns:

1) Waiting Times Concern.
2) Correct Input Concern.
3) Variety in Input Concern.
4) Safety Concern.
5) Not Being Killed Concern.
6) Vicarious Learning Concern.

References

Abelson, R.B., 1983, Whatever became of consistency theory?, *Person. Soc. Psychol. Bull.,* 9, 37-54.

Arnold, M.B., 1970, *Feelings and Emotions: The Loyola Symposium.* Academic Press, New York, London.

Bower, G.H., & Cohen, P.R., 1982, Emotional Influences in memory and thinking: data and theory. *Affect and Cognition.,* Hillsdale:Erlbaum.

Cannon, W.B., 1927, The James-Lange theory of emotion: a critical examination and an alternative theory, *Am. J. Psychol., 39.*

Colby, K.M., 1981, Modelling a paranoid mind. *The behavioural and brain sciences*, 4, 515-560.

Dalkvist, J., & Rollenhagen, C., 1989, On the Cognitive Aspect of Emotions: A Review and a Model. *Reports from the Department of Psychology,* University of Stckholm, No.703

Delgado, J.M.R., 1973, Emotions, *Intro. to General Psychology, 2nd ed.,* WM. C. Brown Company Publishers.

Erman, L.D., Hayes-Roth, F., Lesser, V.R., & Reddy, D.R., 1980, The Hearsay-II speech-understanding system: Integrated knowledge to resolve uncertainty. *Computing Surveys 12, 213-253.*

Frijda, N.H., 1986, *The Emotions, Studies in Emotion & Social Interaction,* Cambridge University Press.

Frijda, N.H., & Swagerman, 1987, Can Computers Feel? Theory and Design of an Emotional System., *Cognition and Emotion Vol 1 Iss 3.*

Kosko, B., 1988, Hidden Patterns in Combined and Adaptive Knowledge Networks, *International Journal of Approximate Reasoning, 2, 377-393.*

Lazarus, R.S., Averill, J.R., and Opton, E.M.Jr., 1970, Towards a cognitive theory of emotion, *Feelings and Emotions: The Loyola Symposium.* Academic Press, New York and London, pp207-32.

Leventhal, H., 1974, Emotions: A basic problem for social psychology, *Social Psychology: Classic and Contemporary Integrations*, Rand-McNally, Chicago, pp.1-51.

Mueller, E.T., & Dyer, M.G., 1985, Towards a computational theory of human daydreaming. *Proceedings of the seventh annual conference of the cognitive science society.* Irvine, California.

Pfeifer, R., & Nicholas, D.W., 1981,Towards computational models of emotion., *Proc. 7th Int. Conf. Artif. Intell. 2369-71.*

Rich, E, 1988, *Artificial Intelligence,* McGraw Hill.

Rosch, E., 1978, Principles of categorisation. *Cognition and categorisation. Hillsdale, N.J.: Lawrence Erlbaum Associates Inc.*

Schachter, S., 1970, The assumption of identity and peripheralist - centralist controversies in motivation and emotion, M. B. Arnold (Ed.) *Feelings and Emotion: The Loyola Symposium.,* Academic Press, New York and London.

Schank, R., & Abelson, R., 1977, *Scripts plans goals and understanding. An inquiry into human knowledge structures.* Hillsdale, N.J.:Lawrence Erlbaum Associates.

Scherer, K.R., 1982, Emotion as a process: Function, origin and regulation, *Soc. Sci. Inf., 21, 4/5, 55-70.*

Sloman, A., & Croucher, M., 1981, Why robots will have emotions, *Proc. 7th Int. Conf. Artif. Intell. 197-202.*

Swagerman, J., 1987, *The Artificial Concern REalization System ACRES: A computer model that has emotions and can name them.* Ph.D. dissertation, Department of Psychology,

Toda, M., 1982, *Man, Robot and Society.* The Hague: Nijhoff.

Wegman, C., 1985.,*Psychoanalysis and cognitive psychology.* London: Academic Press.