A System for Modelling Agents having Emotion and Personality

Lin Padgham, Guy Taylor

Department of Computer Science
RMIT University, Melbourne, Australia
email: linpa@cs.rmit.edu.au, guy@yallara.cs.rmit.edu.au

1 Introduction

There is currently a widespread interest in agent-oriented systems, partly because of the ability of such systems to model and successfully capture complex functionality in a robust and flexible way, suitable for difficult applications. There has also been significant interest in going beyond the rational model of agents, such as that evidenced in the literature modelling agents as systems of *beliefs, desires* and *intentions* (BDI models, e.g. [IGR92, GI89]), to include aspects such as emotions and personality (e.g. [Bat94, BLR92b, Bat93, BLR92a, Slo81]).

Motivations for modelling "broader" agents than simply the rational logical agents of BDI architectures are several. Some consider it necessary to incorporate human aspects such as personality and emotion, in order to make agents more engaging and believable, so that they can better play a role in various interactive systems involving simulation. Entertainment is one obvious application area for such simulation systems, but another is education and training. A simulator that was able to realistically model emotional reactions of people could, for instance, be used in training programs for staff who need to be trained to deal with the public.

Some people also believe that emotions play a functional role in the behaviour of humans and animals, particularly behaviour as part of complex social systems (e.g. [Tod82]). Certainly the introduction of emotions, and their interaction with goals, at various levels, increases the complexity of the agents and social systems that can be modelled. Our belief is that a successful modelling of emotion will enable us to come closer to the goal of building software agents which approach humans in their flexibility and ability to be adaptable and survive in complex, changing and unpredictable environments. However significant work is needed before we can expect to understand the functional role of emotion sufficiently to successfully model it in our software agents.

In this paper we explore some of the ways that we believe emotions and personality interact with goal oriented behaviour, and we describe some of the simplifications we have made in order to build an initial interactive environment for experimentation with animated agents simulating personality and emotions as well as incorporating rational goal directed behaviour.

2 Interactions of emotions, behaviour, goals

In considering the behaviour of human agents, it is evident that emotions affect behaviour in a number of ways. One of the most straightforward ways, and one modelled in systems such as the woggles of oz-world [Bat94] and the agents of the virtual theatre project [HR95], is that emotion modifies the physical behaviour of agents. A happy agent moves faster, and more bouncily, while a sad agent is slower and flatter in its movements.

In addition, emotions can clearly affect an agent's goals, hence affecting their actions. An agent that is very scared is likely to drop a goal to explore its surroundings in favour of a goal to move to a secure position. Emotional effects on goals can be via reordering, or re-prioritising, existing goals, or by introducing completely new goals. Some goals may be introduced simply to manage extreme emotional states. An agent which is feeling intense anger may instantiate a goal to harm the subject of its anger, in order to alleviate the emotional pressure.

In addition to emotions affecting existence and prioritisation of goals, goals, and their success or failure, can affect emotional states. An agent which experiences a goal failure may feel unhappy, while one experiencing goal success may feel glad. Dyer [Dye87] develops a comprehensive lexicon of emotional states, based on goal success and failure. For example an agent which expects its goal to succeed will feel hopeful, an agent whose goal is achieved by the action of another agent will feel grateful, and an agent who expects its goal to be thwarted will feel apprehensive. Figure 1 shows examples of some emotions, indexed by contributing cause.

Frijda and Swagerman [FS87] postulate emotions as processes which safeguard long-term persistent goals or concerns of the agents, such as survival, a desire for stimulation, or a wish to avoid cold and damp. These long term concerns differ from the usual goals of rational agent systems in that they are not things which the agent acts to achieve. Rather they are states which are continually important, and if threatened, cause sub-goals to be instantiated. According to Frijda the process which instantiates appropriate sub-goals is emotion.

The earlier work of Toda [Tod82] also postulates emotions as processes which affect the rational system of the agent, and which are based on basic urges. He groups these urges into emergency urges, biological urges, cognitive urges and social urges. These are similar in principle to the long-term concerns of Frijda and Swagerman. When activated, they focus rational activity and determine the extent to which various competing issues are addressed. Emotions are seen as varying in intensity where the intensity level is an important factor in determining the effect on the rational processing of the agent.

emotion(x)	status	to	goal-situation	by	mode
happy	pos		G(x) achieved		
sad	neg		G(x) thwarted		
grateful	pos	y	G(x) achieved	y	
hopeful	pos		G(x) achieved		expect
disappointed	neg		G(x) thwarted		expect achieved
guilty	neg	y	G(y) thwarted	x	

Fig. 1. Some emotions and their causes

Emotions can also be seen as processes by which agents manipulate each other in social situations, to enable achievement of their goals. By making my child feel grateful (by acting to help them achieve their goal) I increase the chance that the child will act to achieve my goals.

In the emotional model which we are developing it is necessary to capture in some way both the causes of emotion and the effects of emotion. At the most abstract level, we see emotions as being caused by a combination of the state of the world (including self and other agents) and the values of the agent. The emotional reaction caused in an agent by goal success or failure (an event in the world), will depend to some extent on the importance of that goal to the agent. An agent may have a number of motivational concerns, which reflect what is important to that agent. Threats and opportunities associated with these concerns will then generate emotions, as will successs and failure of goals associated with these concerns.

Emotions can affect behaviour directly, or indirectly, via the rational system. Direct effects on behaviour include such things as affecting facial expressions, stance and movement, or causing a relatively instantaneous action, as in the case of flight caused by extreme fear. Effects via the rational system include such things as re-prioritising goals, adding goals and deleting goals. An agent experiencing gratitude might delete (or prioritise lower) those goals which conflict with the goals of the agent to whom it is grateful. Or it may instantiate new goals in order to achieve directly the goals of the agent to whom gratitude is felt.

Clearly there are complex interactions between goals and emotions, which involve both effect of goals on emotions, and emotions on goals. In order to explore and increase our understanding of the dynamics of emotion, personality and goal directed behaviour we have implemented an interactive system for experimentation with various agent groupings.

3 Simplified emotional model

Our goal is to eventually develop an implemented version of a complex model of emotion and personality, incorporating all of the aspects discussed above. However in order to make a start, and to study via an implemented system, the interaction between the emotional subsystem, the cognitive subsystem, and the behavioural subsystem, we have initially developed a rather simple model of emotion.

In this model emotions are seen as existing in pairs of opposites[1] such as pride and shame, happiness and sadness, love and hate. They are represented as a guage, with a neutral point about which the emotion fluctuates in a positive or negative direction. Each agent has a specified threshold for the positive and the negative emotion. When this threshold is crossed, the agent is considered to have that emotion.

The fluctuation of the emotional guage is caused either by events which happen, or by the passage of time. The latter is relevant only when no events are actively influencing the emotion under consideration. This allows an activated emotional state to decay over time, returning eventually to the neutral state, if no events maintain the active state of the emotion.

The events having an influence on emotional state can be of several different types. The first are events which universally and directly affect the emotional state. In this case the event results directly in a modification of the emotional guage of the agent experiencing the event. Probably in a richer model there would be no need for such events, but at the current stage we have included them to allow the emotional effect of events which are not included in the other mechanisms. The second type of events are goal success and goal failure. Currently these events affect only the emotions of happiness, sadness and pride. Goal success (failure) of an agent affects the happiness (sadness) emotion of that agent. Success of the goal of another agent that the agent cares about, results in increase of the pride emotion.

The third, and perhaps most important source of events which affect an agent's emotions are events which recognise threats or opportunities with respect to the motivational concerns of an agent. Motivational concerns are associated with states which are recognised as threats (or opportunities) with respect to those concerns. The motivational concerns are also associated with emotional reactions to be triggered in an agent when the event indicating the relevant state occurs.

---

[1] Some emotions may not have an opposite, in which case it is left undefined.

A motivational concern of stimulation may have a threat associated with the agent being idle. When this state occurs it triggers an event which activates the stimulation concerns and results in an increase in the boredom emotion, which, if the idle state continues will eventually cause boredom to cross the threshold. This will then in its turn cause a bored event, which can trigger a goal to reduce the boredom.

We can represent the set of emotional guages as E, and a particular emotional guage as e, where $v_0(n)$ represents the neutral value on the guage, e^+ represents the name of the positive emotion associated with this guage, and e^- the name of the corresponding negative emotion.

The value of the emotional guage e for a particular agent at any time point is then determined by the previous value of that guage, incremented (or decremented) according to what events the agent is aware of which affect the emotion, or the decay rate[2] of the emotion for that agent. We let the decay rate be a function over the emotions for each agent, written $D_A(e)$. Thus we can say that, with respect to a particular agent, A:

$$V_{A_{t+\delta}}(e) = V_{A_t}(e) + F(\{events\}_t^{t+\delta}, D_A(e, \delta))$$

The effect of an emotion on the behavioural and cognitive subsystems of the agent comes into play when an emotion crosses the threshold for that agent. This then results in the belief that the agent experiences the emotion being asserted (or retracted) in the cognitive subsystem. The behavioural subsystem may also be notified. This in its turn can lead to goals being asserted, deleted, or reprioritised, in the cognitive subsystem, and to behaviour being modified directly in the behavioural susbsystem. For example, if the belief that the emotion sadness is being experienced is asserted, it results in modifying the agent's stance and movement in the behavioural system, to reflect sadness. An assertion of boredom[3] results in instantiating a goal to relieve the boredom. Emotional thresholds are represented individually for each agent and emotion. We let P and N be functions giving the positive and negative thresholds for all emotions for a given agent. Thus $P_A(e)$ represents the threshold at which emotion e^+ is asserted, and $N_A(e)$ represents the threshold at which e^- is asserted, for agent A. Thus we can say that for a given agent, A:

$$(V_{A_t}(e) \geq P_A(e)) \wedge \neg believe_A(e^+) \Rightarrow assert\ believe_A(e^+) \text{ which results in behavioural and cognitive effects.}$$

and

$$V_{A_t}(e) \leq N_A(e) \wedge \neg believe_A(e^-) \Rightarrow assert\ believe_A(e^-) \text{ which results in behavioural and cognitive effects.}$$

---

[2] A decay rate may also be negative, allowing an emotion to increase over time, if not affected by any events. This is used particularly for modelling physical urges, such as hunger, which in many ways are similar to emotions with respect to their effect on the cognitive system.

[3] For ease of expression we will sometimes speak about asserting an emotion. What is strictly meant is that we assert the belief that the agents is experiencing the emotion.

Also when values cross back into the neutral zone emotions are retracted in a similar way.

4 Personality

Personality is a concept which is ill-defined, but which clearly is related to some aspects of our simple emotional model of agents. In particular, the emotional model described provides three different mechanisms which can be used for modelling agent personality. Thes are the motivational concerns, the emotion thresholds, and the rate of decay for an emotion.

One aspect of personality is a notion of what things are important to that person - a person who is always concerned about money and financial matters could be represented as a person having a motivational concern for financial well-being. This person will respond emotionally to events affecting this aspect of life, and will thus appear to have a quite different personality than an agent who does not have this motivational concern, and therefore does not respond emotionally to the particular events associated with threats and opportunities associated with this concern. Some motivational concerns will be more or less universal - such as that for stimulation or for survival. However others will be quite individual, and will be a significant aspect of the agent's personality.

The threshold at which an emotion is asserted is also an important aspect of personality. The individual who experiences many anger increasing events, before becoming "angry" has a different personality to the agent who becomes angry easily - perhaps after one or two such events.

Finally the decay rate for each emotion is an important aspect of personality. Two agents with the same threshold for anger, and the same motivational concerns, may still exhibit differing personalities (with respect to this emotion), based on their decay rate for anger. An agent whose anger wears off very slowly has a different personality to the agent whose anger dissipates almost directly. The former personality trait could be described as "the sort of person who bears a grudge".

With these three aspects of the emotional model at our disposal, the personality of an agent can then be said to depend on the motivational concerns of that agent plus the thresholds and decay rates for each emotion, for that agent. If we let M_A be the set of motivational concerns for agent A, the personality of A can be described by:

$$\Phi_A = < M_A, N_A, P_A, D_A >$$

These relatively simple mechanisms, while clearly not capturing all aspects of personality, nevertheless do give us a reasonably rich mechanism for beginning to represent varying personalities of agents.

5 System description

The system is implemented on an SGI and a Sun workstation and uses dMars[4] a descendant of PRS [IGR92, GI89], and Open Inventor^{TM} a set of library routines for high level graphics coding.

The system developed is similar to that reported in [PTQ96], but generalised to an interactive, menu-based system which allows users to develop arbitrary scenarios. Users can choose agents and objects from menus, in order to populate the world. Agent plans are then defined and edited via the dMars plan-editor, available via a system menu. Personality variables will also be definable via the menu interface, although they currently are read in from files.

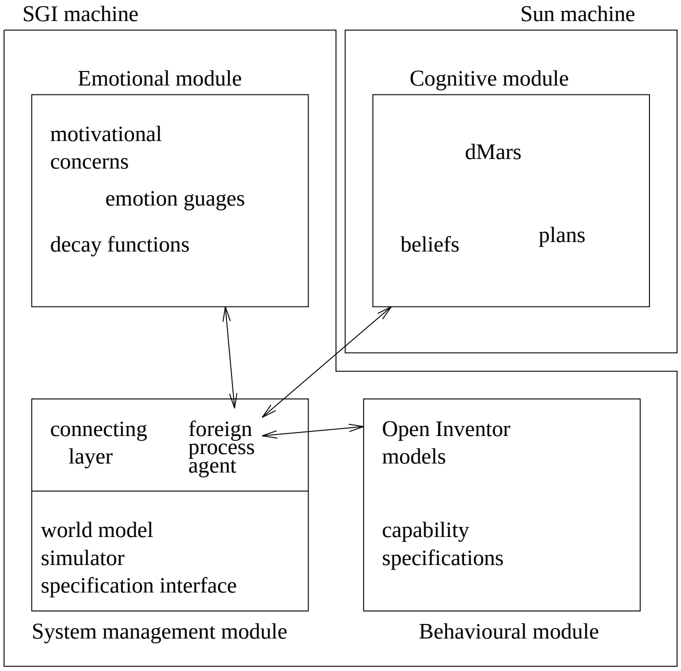

Fig. 2. Architectural overview.

There are four main modules in the system, as shown in figure 2. The cogni-

---

[4] Distributed Multi Agent Reasoning System, from the Australian Artificial Intelligence Institute.

tive module (which is exactly dMars) manages the cognitive processing of all agents - their beliefs, desires, intentions and the plans available to them for rational action. The behavioural module, which uses Open Inventor, manages the graphical models for agents and objects in the world, and their animation. The emotional subsystem manages the emotional processing for each agent while the system management module manages the interface of the interactive system, the simulation and world model, and the connecting layer which controls the communication between the cognitive, physical and emotional subsystems of agents. A part of the connecting layer is a "Foreign Process Agent" which is a module supplied by dMars for managing communication between dMars and other parts of an application. The foreign process agent supports the use of sockets, facilitatingthe use of different machines for the different subsystems. All communication between dMars and the emotional or physical subsystems goes through the foreign process agent. Each individual agent consists of a logical thread within dMars, the emotional subsystem, and the behavioural subsystem, with the integration of these subsystems managed by the connecting layer.

Because of the difference in the level of granularity which is needed by the graphical subsystem, and that which makes sense for the cognitive (or emotional) subsystem, there is some significant management of behaviours in the connecting layer. For example, the cognitive subsystem may request the action to 'go across the room'; the connecting layer will then manage the sequence of steps required to execute this action. The connecting layer also manages the perceptions from the environment, such as sight and smell, passing these up to the cognitive and emotional subsystems whenever there is a change. While percepts are received continually, we consider it worthy of attention by the cognitive or emotional subsystem, only when these change.

Having briefly described the overall architecture, we now describe the architecture of the connecting layer in more detail.

5.1 Connecting Layer

The connecting layer includes two major subsystems, the action subsystem and the sensing subsystem. The action subsystem controls how the bodies of the agents move within the physical environment, and are synchronised with the behavioural control emanating from the cognitive subsystem. The sensing subsystem manages how the agents percieve the environment, and communicates these perceptions to the emotional and cognitive subsystems.

The action subsystem enables high level actions, appropriate to the cognitive and emotional subsystems to be transformed to the primitive actions needed for the animation system.

The simplest actions are atomic actions which are conceptually instantaneous and have a negligible duration. As far as the cognitive system is concerned they can either succeed or fail. These actions are at the same level of granularity as atomic actions in the animation system and can be mapped to a single command in the animation system. An example is a turn action, which would be mapped to a rotation.

Other kinds of actions have duration. They may either have a known end condition such as an action to move to a particular location, or they may be continuous, with no fixed end point such as the action to wag a tail.

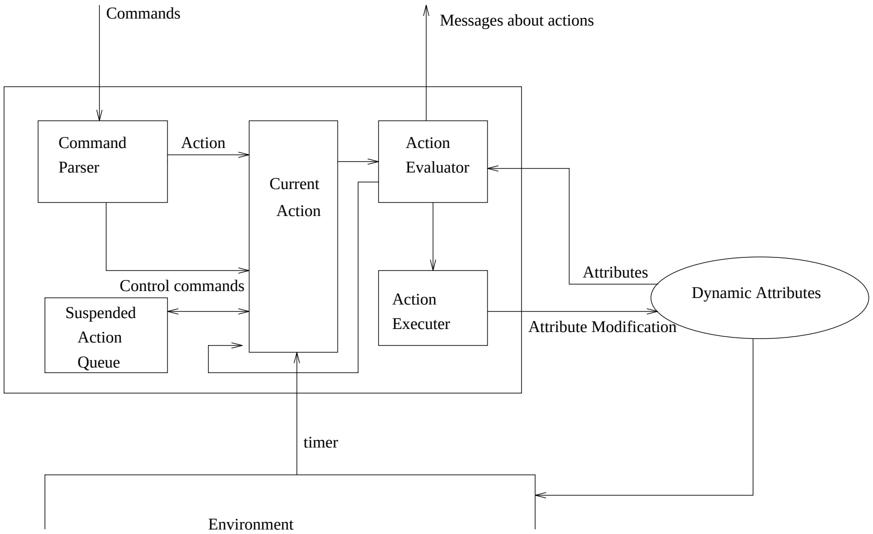

Fig. 3. Action section of Connecting Layer.

A figure of the architecture for the action subsection of the connecting layer is shown in figure 3.

The Command parser looks at commands received from the agent layer, of which there are two types, control commands and action commands.

Action commands are placed in the 'current action' processor. This will calculate what is needed for the next frame and check to see if the action is complete. For example, with the action 'move to object X', at each frame the agent will take a step closer to the object. It will see if it is at the desired location, and if so, stop the current action. If not, this sequence will be repeated at the next frame.

Control commands are used for controlling the queue in ways other than the normal execution. In order to allow the system to be fully reactive, new goals/intentions must be able to take priority over current goals at any point in time. This requires that executing actions must be able to be aborted and/or suspended, at the animation level when requested by the cognitive or emotional subsystems.

The action evaluator checks the validity of the next action. An action may be started, but at some time during its execution a condition could arise which means the action is no longer able to finish executing. Collisions with other objets are an example. In this case the action evaluator sends a message to the agent's cognitive and emotional subsystems and aborts the current activity. Thus the action evaluator checks the agent's capabilty with respect to the action, at each step.

Finally, the action token executer sets the values of the graphical models in the world at each timer event by reading the head of the event queue. Each action token has associated with it a motion path, allowing Open Inventor to create a smooth animation for that atomic action. For example, an action like 'take a step' can have the motion path for leg movement included in the model file, allowing smooth walking to be easily realised. The action executor also considers input from the emotional subsystem regarding the current physical ramifications of the emotional state.

The other subsystem of the connecting layer is the sensing subsystem. Two types of sensing are modelled - goal-directed sensing and automatic sensing. Goal directed sensing allows an agent to explicitly ask for a piece of sensory information, receiving a response from the connecting layer. An automatic sense is triggered whenever a certain condition occurs. For example smell may be modelled as an automatic sense. As soon as an object is within range of the agent's smelling capability, a message would be triggered, alerting the agent to the newly found smell.

An automatic sense is set up by indicating what conditions would trigger the sense, for the given agent. For example smell may be triggered by any object with a smell attribute, within a certain radius of the agent. The connecting layer then monitors for the occurrence of this situation. At each timer event, the connecting layer checks to see if something is sensed, based on the agents capabilities and the current environment. If the sensory information is different from that of the last timer event, then a message is generated and sent to the agent's emotional and cognitive subsystems. This mechanism avoids redundant messages when sensory information remains unchanged.

6 Discussion

We have currently developed three scenarios within this system - "Dog World" containing dogs which explore a world containing food and obstacles, and engage in actions such as playing, barking, exploring and eating (see figure 4), a world containing mice, a wheel and a house where in addition to the dogworld actions the mice play, hide, and get frustrated, and a world based on the Luxo Jr animation with a parent and a child lamp, and some balls.

For our initial scenario we had two dogs, Max and Fido, both with the same motivational concerns of hunger and stimulation. Max had $N_{Max}(e_1)$ relatively close to the neutral point for the emotional guage e_1 where e_1^- is the emotion anger. $D_{Max}(e_1)$ was low, giving a gradual decay rate for the anger emotion. This led to Max behaving as if he had a relatively aggressive personality. For the emotional guage e_2 measuring fear, $N_{Max}(e_2)$ was relatively far from the neutral point and $D_{Max}(e_2)$ was high. Fido, on the other hand became easily fearful with $N_{Fido}(e_2)$ close to the neutral point, but also lost his fear quite quickly with $D_{Fido}(e_2)$ being quite high. $N_{Fido}(e_1)$ and $D_{Fido}(e_1)$ were both at intermediate levels.

In both the dog and the lamp scenarios the agents successfully displayed the features of different personalities, based on their emotional profiles. Despite the simplicity of the emotional model, it was possible to model agents whose differing personalities resulted in differing behaviour, given similar external situations.

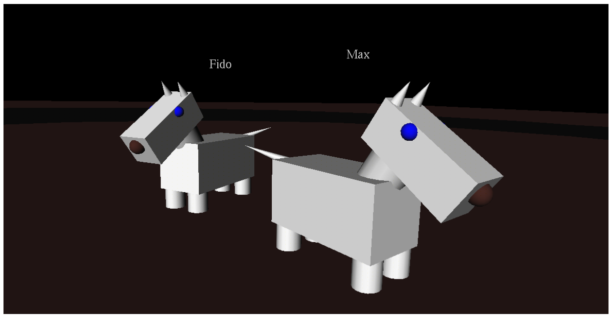

Fig. 4. Max and Fido in Dog World

An example of interaction in the Max and Fido system is when Max barks at Fido (because Fido has approached the food while Max is eating), causing Fido's fear guage to quickly reach the negative threshold, triggering the fear reaction and causing Fido to run away. Because $D_{Fido}(e_2)$ (Fido's decay rate for fear) is quite high his fear disappears quickly once Max stops barking. Thus the following happens for Fido:

$e_2^- = fear$;
$N_{Fido}(e_2) = -3$;
$D_{Fido}(e_2) = -1$;
$V_{Fido_{t1}}(e_2) = 0$;
$F(\{events\}_{t1}^{t3}, D_A(e_2, \delta)) = -3$ (based on a bark from Max at t_1, t_2 and t_3)
Thus, given that $\neg believe_{Fido}(fear)$ at t_3
$V_{Fido_{t3}}(e_2) = -3 \leq N_{Fido}(e_2) \Rightarrow assert\ believe_{Fido}(fear)$ with the effect that Fido runs away;

Subsequently
$F(\{events\}_{t3}^{t4}, D_A(e_2, \delta)) = -1$ (a further bark from Max);
$F(\{events\}_{t4}^{t5}, D_A(e_2, \delta)) = 1$ (no relevant events, decay of 1);
$F(\{events\}_{t5}^{t6}, D_A(e_2, \delta)) = 1$ (no relevant events, decay of 1);
Thus, noting that $believe_{Fido}(fear)$ at t_6
$V_{Fido_{t6}}(e_2) = -2 \not\leq N_{Fido}(e_2) \Rightarrow$ *retract* $believe_{Fido}(fear)$ with the effect that Fido stops running away;

One noticeable shortcoming of the current emotional model is that it does not support aspects of emotion requiring more knowledge about the objects of emotion. For example the pride emotion is generated whenever a person in a close relationship to the agent achieves their goal. A more accurate representation of the pride emotion would however be *proud of*, where the emotion has an object. This is true of many emotions, for example *angry at*, *love towards*, *fear of*. This suggests the need to extend the representation of emotions to allow for this information.

Currently the emotional update is purely a function of events, the decay rate for the emotion, and time. Probably in a richer model we would also want to include beliefs and goals in this emotional update equation. My emotional response to the failure of a particular goal if I believe that another agent caused my goal failure (leading to anger), is different when I believe that the failure is my own fault (perhaps leading to frustration). These aspects are being explored in ongoing work.

Acknowledgements

The authors thank the Australian Artificial Intelligence Institute for use of the dMars software, and for useful discussions and technical support, particularly from Ralph Rönnquist, Anand Rao and Andrew Hodgson. The authors also acknowledge Ghassan Al Qaimari's co-supervision of the honours project that preceded this work.

References

[Bat93] J. Bates. The nature of characters in interactive worlds and the oz project. *Virtual Reality: Anthology of Industry and Culture*, 1993.

[Bat94] J. Bates. The role of emotion in believable agents. *Communications of the ACM (Special Issue on Agents)*, 1994.

[BLR92a] J. Bates, A.B. Loyall, and W. S. Reilly. An architecture for action, emotion and social behavior. *Proceedings of the Fourth European Workshop on Modeling Autonomous Agents in a Multi-Agent World*, 1992.

[BLR92b] J. Bates, B. Loyall, and W. S. Reilly. Broad agents. *Proceedings of the AAAI Spring Symposium on Integrated Architectures*, 2(4), August 1992.

[Dye87] Michael G. Dyer. Emotions and their computations: Three computer models. *Cognition and Emotion*, 1(3), 1987.

[FS87] Nico H. Frijda and Jaap Swagerman. Can computers feel? theory and design of an emotional system. *Cognition and Emotion*, 1(3), 1987.

[GI89] M. Georgeff and F. Ingrand. Decision-making in an embedded reasoning system. In *Proceedings of the International Joint Conference on Artificial Intelligence - IJCAI '89*, pages 972–978, Detroit, USA, Aug 1989.

[HR95] Barbara Hayes-Roth. Agents on stage: Advancing the art of ai. In *IJCAI'95*, Montreal, 1995.

[IGR92] M. F. Ingrand, M. P. Georgeff, and A. S. Rao. An architecture for real-time reasoning and system control. *IEEE Expert*, pages 34–44, Dec 1992.

[PTQ96] L. Padgham, G. Taylor, and G. Al Qaimari. An intelligent believable agent environment. Technical Report TR-96-31, Royal Melbourne Institute of Technology, Melbourne 3001, Australia, August 1996. Presented at AAAI-96 Workshop on Entertainment and AI/A Life, Portland, Oregon.

[Slo81] Aaron Sloman. Why robots will have emotions. In *IJCAI'81*, Vancouver, 1981.

[Tod82] Masanao Toda. *Man, Robot and Society.* Martinus Nijhoff Publishing, Boston, 1982.

This article was processed using the LaTeX macro package with LLNCS style