Behaviors that Emerge from Emotion and Cognition: Implementation and Evaluation of a Symbolic-Connectionist Architecture

Amy E. Henninger, PhD
Soar Technology, Inc.
3361 Rouse Road Suite 240
Orlando, FL 32817
407 207 2237
amy@soartech.com

Randolph M. Jones, PhD
Colby College & Soar Technology
5847 Mayflower Hill Drive
Waterville, ME 04901-8858
207-872-3831
rjones@soartech.com

Eric Chown, PhD
Bowdoin College
8650 College Station
Brunswick, ME 04011
207-725-3084
echown@bowdoin.edu

ABSTRACT

This paper describes the implementation and evaluation of a framework for modeling emotions in complex, decision-making agents. Sponsored by U.S. Army Research Institute (ARI), the objective of this research is to make the decision-making process of complex agents less predictable and more realistic, by incorporating emotional factors that affect humans. In tune with modern theories of emotions, we regard emotions essentially as subconscious signals and evaluations that inform, modify, and receive feedback from a variety of sources including higher cognitive processes and the sensorimotor system. Thus, our work explicitly distinguishes the subconscious processes (in a connectionist implementation) and the decision making that is subject to emotional influences (in a symbolic cognitive architecture).

It is our position that "emotional states" are emergent patterns of interaction between decision-making knowledge and these emotional signal systems. To this end, we have adopted an approach that promotes the emergence of behavior as a result of complex interactions between factors affecting emotions, integrated in the connectionist-style model, and factors affecting decision making, represented in the symbolic model.

This paper presents the implementation of emotions architecture and explains how we evaluated the system. This includes a description of the behaviors we used in our prototype, the design of our experiments, a representative set of behavior patterns that emerged as a result of exercising our model over the design space, and our project's lessons learned

Categories and Subject Descriptors

I.2.11 [**Distributed Artificial Intelligence**]: Intelligent Agents, Multiagent systems.

AAMAS'03, July 14–18, 2003, Melbourne, Australia.

General Terms

Design, Experimentation, Human Factors, Verification.

Keywords

Synthetic agents, emotions, cognition, hybrid architectures

1. INTRODUCTION

This paper describes a framework for modeling emotions in an interactive, decision-making agent. In tune with modern theories of emotions [2],[9] we regard emotions essentially as subconscious signals and evaluations that inform, modify, and receive feedback from higher cognitive processes. Thus, our work explicitly distinguishes the subconscious processes (in a connectionist implementation) and the decision making that is subject to emotional influences (in a symbolic cognitive architecture). We are integrating a connectionist model of emotions from [1] with Rosenbloom, Laird, and Newell's [11],[12] Soar architecture. Sponsored by the U.S. Army, the application area incorporates emotions and individual differences into the behavior models of synthetic Special Forces Agents in a battlefield simulation. This is an ideal test area for a model of emotions, because the intelligent agents must exercise a variety of reasoning capabilities, including situation assessment, planning, reacting to goal failures, and interacting with a team of agents. While the framework for this model is being developed within the military domain, we anticipate that the design is general enough to apply to other areas (e.g., animated characters) as well.

In our framework, symbolic assessments of a small set of "emotional attributes" reside in a working memory. These attributes are explained in the following section and include indicators of affect, arousal, and information degradation. Portions of working memory are accessible by the deliberate cognitive process, and portions are accessible by the emotion mechanisms, so memory serves as the interface between the two. These working memory elements combine with background knowledge to generate strategies, reasoning, and external behavior. At the same time, the cognitive model creates working interpretations of the environment and status of internal goals (situational awareness). Some of these interpretations and assessments feed into the connectionist model, which in turn continuously computes new values for each emotional attribute. Instead of considering cognition and emotion as opposing forces, this architecture supports the view that they evolved together to maintain effective responses to stimuli that influence the survival of the self and the species. At higher levels of cognition, these responses manifest themselves as decision-making that is constrained in a variety of ways. Because the theory underpinning our model assumes that these responses were ultimately provided by evolution, we assume that these constraints are, on *average*, beneficial to decision making. Clearly, not all emotional responses are always beneficial. Thus, we recognize the need to demonstrate such tradeoffs in our experiments.

The following two sections present an overview of the symbolic component representing cognition and the connectionist component representing emotion, respectively. Detailed descriptions and computational representations of these components and their interactions may be found in [7]. After describing the model's components, we present a personality framework for exercising the model and a narrative of the scenario used to prototype the model.

2. COGNITIVE-SYMBOLIC MODEL

As illustrated in Figure 1, both the Decision Making and the Emotional Appraisal component of the emotions model occur within Soar, the cognitive model. Specifically, these two components reside in long-term memory where they are represented in the form of productions. Although decision -making is not a new component to Soar-based Intelligent Forces (IFORs), modeling the influence of emotions on decision-making is new. Thus, in the following two sub-sections, we explain both how the Emotional Appraisal system works as well as how the Decision Making process is influenced by the resulting emotions.

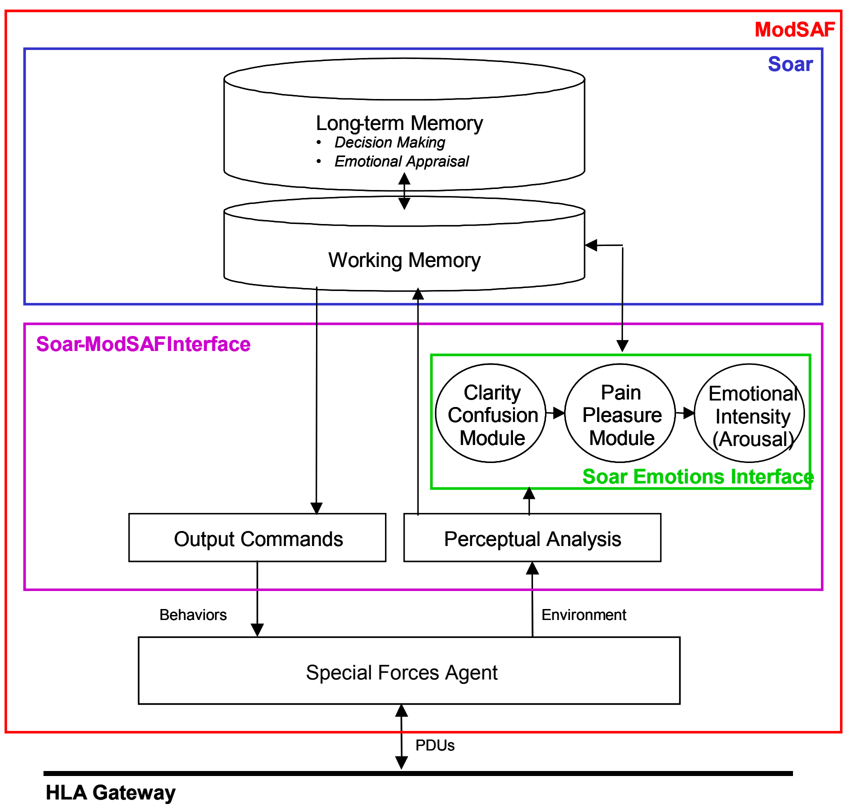

Figure 1. Block Diagram of Symbolic-Connectionist Framework

2.1 Emotional Appraisal

Following [5] appraisal in our system is based around goals. The most straightforward types of appraisal require monitoring whether goals have been achieved, have become likely or unlikely to be achieved, or have been deemed unachievable. The determination of the status of these goals comes from the cognitive system's assessment of situational awareness information, which is, in turn, provided by the Perceptual Analysis module in the Soar-ModSAF Interface, together with long-term situation-interpretation knowledge. Each of these types of appraisals results in signals to the "pleasure/pain" and "clarity/confusion" centers of the Emotions Interface

Again, since this form of appraisal is centered around the goals in the agent's current plan, only certain types of situational information are relevant for particular types of goals. The system will only be "concerned" about whether it is clear or confused about inputs when those inputs are germane to the current set of goals. For example, to satisfy the goal of destroying an enemy tank, the agent must be the goal of destroying an enemy tank, the agent must be able to detect the location of that tank. If the agent could not detect the location, it would experience an increase in confusion. On the other hand, if the agent's planner had not established the goal of destroying a particular tank, then lacking contact with that tank is of no concern to the planning system.

2.2 Decision Making

The primary input from the emotions interface into the planning agent is a signal representing a level of arousal. One of the primary effects of a high level of arousal is to narrow the focus of attention. In the planning agent, we represent a narrowed focus of attention by restricting the knowledge that will be brought to bear on the plan monitoring, execution, and re-planning processes. One aspect of the narrowing of focus is that, when highly aroused, an agent will neglect to apply knowledge that is not well rehearsed. This will cause the agent to migrate its behavior toward its "core personality" or expertise during episodes of heightened arousal. For example, if an agent has a strong tendency toward risky behavior incorporated into its knowledge base, but it has been briefed with a low-risk mission, an increase in arousal will cause the agent to ignore the low-arousal knowledge (from the mission briefing) and revert to well-rehearsed, ingrained high-risk behavior. We accomplish this in the emotional planner by tagging agent rules with arousal thresholds. Only those rules with a threshold exceeding the current level of arousal are allowed to fire. The side effect of this approach is to allow much more thoughtful and deliberative reasoning under conditions of low arousal.

Other input from the emotions system includes current levels of pleasure and pain. These inputs may influence the preferences and evaluations that the agent uses when comparing alternative courses of action during re-planning or alternative interpretations during situation assessment.

3. EMOTION-CONNECTIONIST MODEL

As shown in Figure 2, the connectionist model consists of several interacting components: arousal, pleasure/pain and clarity/confusion.

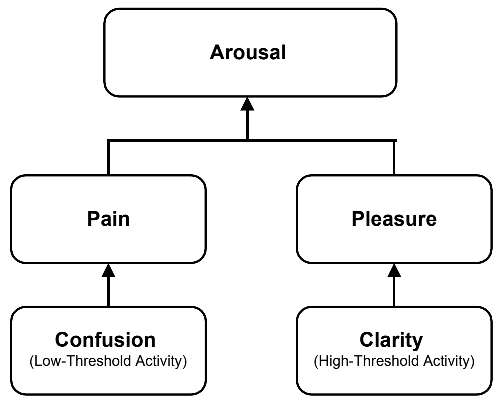

Figure 2. Block Diagram of Computational Arousal Mechanism

The clarity and confusion system, based on [8], represents important correlates of pleasure and pain in forms of higher intelligence. [8] suggests that members of each species possess particular characteristics (e.g., agility, camouflage, etc) that facilitate their survival. Since humans are not particularly fast, fierce, or camouflaged, like other species are, Kaplan asserts that we rely on our ability to organize, store, and use information to enhance our survival. Consequently, for humans, confusion is a potentially dangerous attribute and clarity is a desirable attribute

The pleasure/pain system interprets the level to which a stimulus represents a threat or enhancement to survival. In other words, stimuli that impede one's chances of survival would be tagged as painful and stimuli that would help one survive or reproduce would be tagged as pleasurable. This applies both to immediate sensations of physical pain as well as to deliberate predictions of situations and outcomes.

Whereas pleasure, pain, confusion and clarity all work to detect events of importance to an agent, the arousal system functions as a kind of interface between the emotional and higher cognitive systems. This relationship has been demonstrated by a number of researchers [3], [10] documenting the effects of arousal on a variety of cognitive factors such as learning, memory, and attention.

Clarity && Pleasure ➔ Joy
Clarity && Pain ➔ Anger
Confusion && Pleasure ➔ Surprise
Confusion && Pain ➔ Fear

Incorporating dimension of arousal, could yield:
Confusion && Pain && High Arousal ➔ Panic
Confusion && Pain && Low Arousal ➔ Anxiety

Or, incorporating dimension of time could yield:
Clarity && Pain && Past ➔ Regret

3. PERSONALITY–EXPERIMENTAL SPACE

Since different people have different reactions to the same situations (i.e., emotions and emotional responses are unique to individuals), we use the body of research in emotions and temperament to develop the bounds of an experimental design region for testing our model. Such differences can be thought of as an "emotional style" or temperament. Using this framework, we are able to account for individual differences in temperament by changing the connection strengths in the emotional subsystem. For example the psychological literature has long theorized that the critical factor that distinguishes introverts and extroverts is the relative susceptibility to becoming aroused [3], [4].

Figure 3 illustrates the personality framework used to exercise the emotions model. Fundamentally, we model the introversion/extraversion dimension of personality by incorporating a susceptibility to arousal parameter in the emotions model, the neuroticism/stability dimension with susceptibility to pain, and the explorer/preserver dimension with susceptibility to confusion. Thus, by adopting this mapping, we are able to model an individual's emotional style such that it can lead to distinct decision making profiles in a variety of emotionally charged scenarios. Again, more information on these mappings can be found in [7].

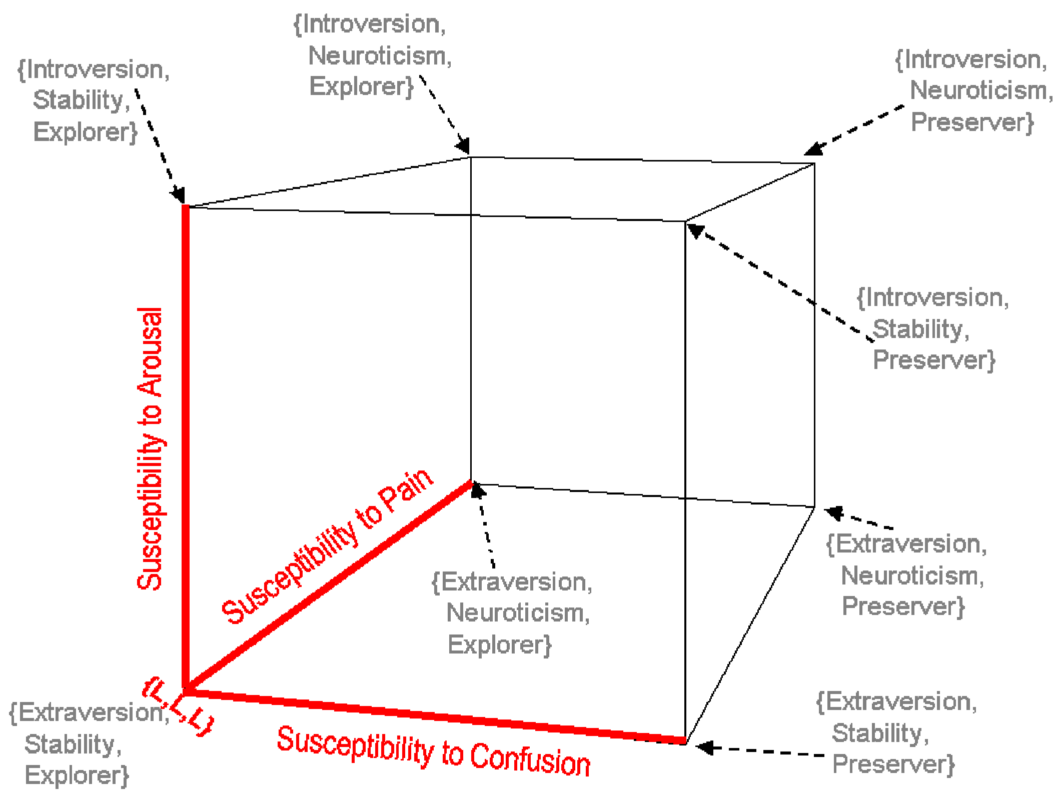

Figure 3. Personality Space used to Exercise Emotions Model

4. PROTOTYPE

The scenario used to prototype the emotions model was the Special Operations Forces (SOF)Soar Long Range Reconnaissance Mission. This task involves a 6-man team inserted deep within enemy territory. Once inserted, they travel anywhere from 20-50 km to the Objective Rally Point and split into three 2-man teams (i.e., two 2-man observation teams, and one 2-man radio team).

Figure 4 shows an example of how the physical mission parameters may be arranged. In this mission, there are five types of critical points (Drop, Rally, Observation, Transmit, and Pickup points) and one critical area, the Objective area. Currently, the prototype focuses on the behavior of a two-man team at an Observation point.

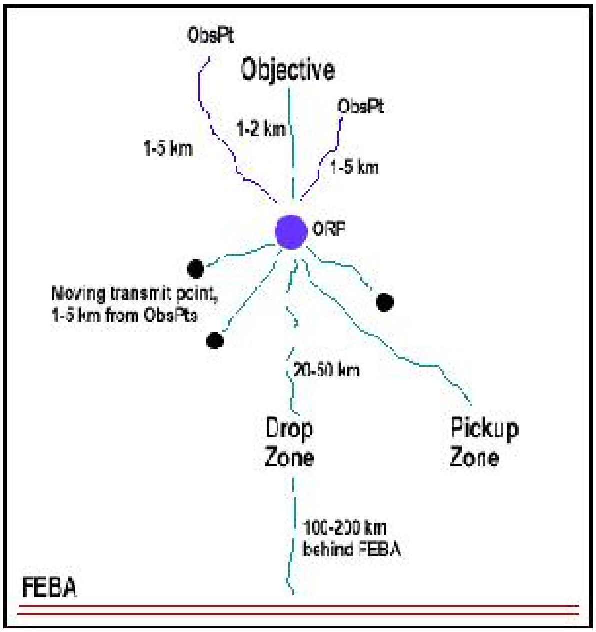

Figure 4. Example Map of Long Range Reconnaissance Mission Parameters

After splitting into 2-man teams, the observation teams seek cover and set up at the Observation point near the designated Objective area. Once an appropriate objective has been sighted, the team will report back to the radio team via wireline radios. At the conclusion of the mission, the teams will make their way to the Pickup Zone for exfiltration.

A run-time screen shot of our prototype may be seen in Figure 5, which shows the SOF Agents running in JSAF, an Agent's Soar Interface Panel, and an Agent's Emotions Interface Panel. The Soar Interface Panel enables operator control of an agent and communicates agent's decisions and actions. The Emotions Panel Interface is used to monitor the agent's emotional subsystems and responses. To evaluate our system, we enhanced the SOF Agents with emotional responses, given some range of triggers. For example, we could alter the scenario by allowing detections, engagements, injuries, etc. The specific cases used in this prototype are presented in the next section.

5. VALIDATION METHODOLOGY

As described in the previous sections, there are a number of important interactions in our system:

1. interaction amongst the emotional subsystems (i.e., arousal, pain/pleasure, clarity/confusion),
2. interactions between the emotional system and cognition, and
3. interactions between emotional entities.

Taken together, the whole creates a fairly complex feedback system, in which the resulting external behavior would be very difficult to predict analytically [15]. This justifies the approach of building these models within executable intelligent agents, so that the resulting behaviors and emotional effects can be characterized empirically.

The multi-level, embedded nature of these interactions forms the foundation of a model that exhibits emerging behavior. One form of an emerging behavior model that relates specifically to simulated agents is the individual-based model or IBM [14]. Because behaviors of IBMs are difficult to verify through conventional, analytical methods [13],[14] we adopt principles of the IBM community in validating the model. Essentially, this approach considers validation an issue of deciding whether the model output meets the required performance standards according to the model's purpose. [14] offers a number of strategies for verifying and validating the software accuracy of IBMs or other similar types of models based on principles of emerging behaviors. Central to these strategies is a clear experimental procedure, where testing is treated as a scientific enquiry with testers designing experiments, predicting the outcomes, and then running the code to compare the actual outcomes to the predicted outcomes. This can take the form of spot checks of key model subcomponents, pattern tests, and systematic tests against an independent implementation.

To test the model presented in this research, we adopt a dual approach, measuring the within-agent emerging behavior pattern (i.e., the interaction of the emotional model substructures and cognition for an individual agent, given a personality type) and the between-agent

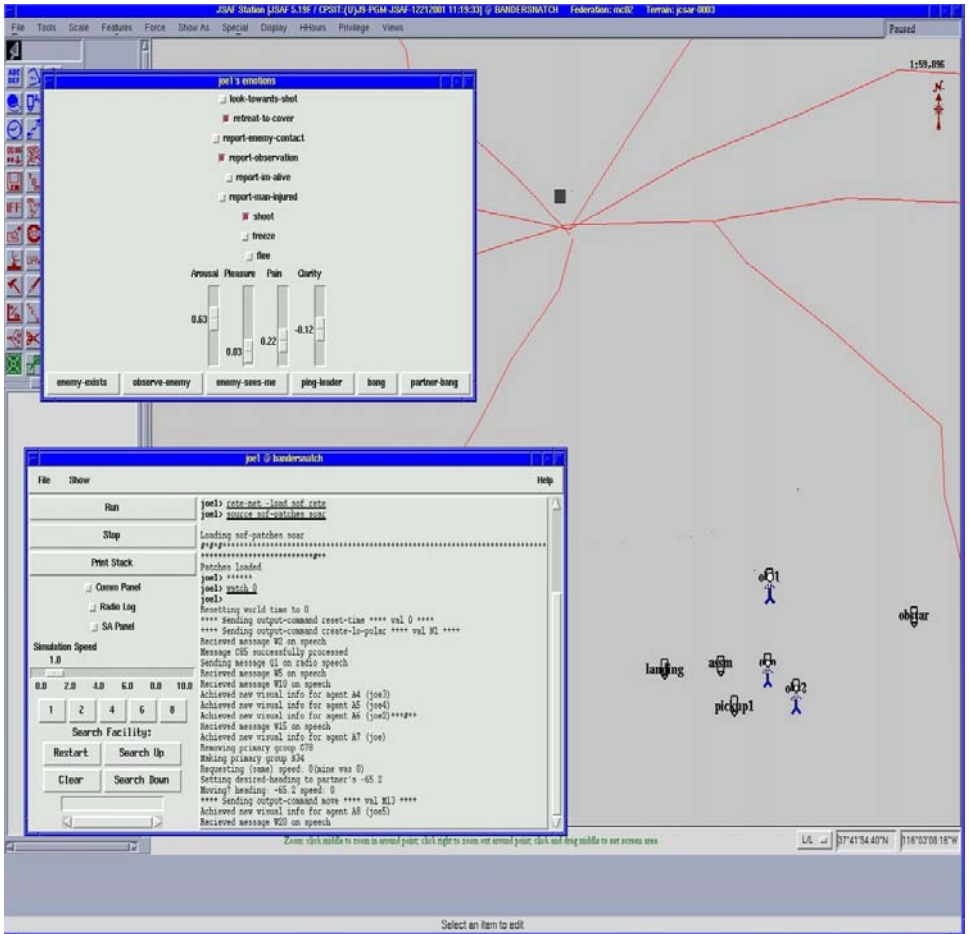

Figure 5. Runtime Screen Shot of Emotional SOF Soar Agent

emerging behavior pattern (i.e., the interaction between agents, both of whom can have unique emotional states and temperaments). These tests are performed over a number of increasingly complex test cases. Initially, as a means of verifying model code, we consider static test cases to measure the within-agent patterns. These measures include verifying the emotional system's numeric output according to personality types, and verifying that the correct response is selected, given some state vector. All of these measures are compared against a manual simulation of each test case to determine accuracy. Follow-on tests evaluate the more complex cases measuring the dynamic behavior emerging from between-agent interaction. As in the simple case, these tests also systematically increase in complexity. Thus, the general form to our testing procedure is to implement the simplest cases first and then allow the more complex cases to emerge as a result of the behaviors that have already been established and evaluated. For example, for the within-agent tests, each of the scenarios 1-5 (see Table 1) is used to statically evaluate the emotional system, in isolation. Then, the next more complicated round of tests evaluates the first-order case where emotion and cognition interact and is accomplished by verifying model results with independently implemented manual simulations.

Once these initial results are verified through static tests, dynamic cases starting with scenario 1 progressing to variations of scenarios 4/5 are executed to record patterns in between-agent behavior. Differences in behaviors over these scenarios will be due to differences in agent's arousal level and how that impacts cognition, where these components of the model were previously verified in within-agent tests.

6. RESULTS

As stated in the previous section, the primary focus of the validation effort is to develop a sense of the model's utility, given its purpose. The objective of this research was to make the decision-making process of complex agents less predictable and more realistic, by incorporating emotions. The

Table 1. Progression of Test Scenarios

Scenario	Description
1. At Observation Point.	Two-man Observation team is stationed at the Observation Point and there is no enemy in sight.
2. At Observation Point and Enemy Detected.	Two-man Observation team is stationed at the Observation Point and a high number of enemy have been sighted.
3. At Observation Point, Enemy Detected, and Shooting.	Two-man Observation team is stationed at the Observation Point, spotted enemy, and detected shooting.
4. At Observation Point, Enemy Detected, Shooting, and Teammate Hit.	Two-man Observation team is stationed at the Observation Point, detected enemy, and agent's teammate has been shot.
5. At Observation Point, Enemy Detected, Shooting, and I'm Hit.	Two-man Observation team is stationed at the Observation Point, detected enemy, and the agent has been shot.

following sub-sections present results pertaining to both of these cases. In the first sub-section, we demonstrate the model's utility by presenting scripted output of simple test scenario. Next, in the second sub-section, we report on means of determining how this system reduces the predictability of an agent's behavior.

6.1 Model Output

Output behavior of one example of a simple dynamic case is scripted in Table 2. This test case contrasts the behavior of Agent assigned emotional style of <Extravert, Stability, Explorer> with an Agent assigned emotional style of <Introvert, Neurotic, Preserver>. In this scenario, the SOF Agents are at the Observation Point and detect enemy, the objective on which Agents should report. Up to this point, even though the Agents exhibit differences in the values of their emotional parameters, the Agents propose and select the same reactions based on the same world events. The next event, "Enemy-Sees-Me" causes Agent2 to propose one more action ("flee"). However, both Agents choose to "Retreat-to-Cover". During the next event, "Partner-Shot", behavior of the two Agents starts to diverge (i.e., Agent1 chooses to report the injury, whereas Agent2 continues to seek cover). Lastly, as the "Shooting" continues, Agent1 remains active in seeking cover, whereas Agent2 "freezes", rendering him useless in the rest of the scenario.

6.2 Reducing Agent Predictability

To measure the first objective we compared the range of the agent's response space using a classic, deterministic state-transition approach with the range of the agent's response space using our emotional model, which is also deterministic. Thus in a classical state-transition construct based on change in world state, as seen in equation 1, there is some fixed number of outputs, given a current world state and an input.

$$\lambda_i = \lambda(q, E_i) \qquad (1)$$

where λ_i = set of outputs,

for any external input, E_i and any state, q

On the other hand, our approach, still viewed from perspective of state-transition construct, also selects output as a function of the world state and an input. In this case, however, that input is augmented by another state variable internal to the agent (e.g., arousal).

$$\lambda_j = \lambda(q, E_j, I_j) \qquad (2)$$

where I_j = is an input internally generated by Agent

To compare the two methods, we manually calculated the number of behaviors possible for each state in the prototype system for the classic approach and the emotions approach. Comparison of these numbers reveals an increased size in response space by average of 3.1, as shown in equation 3.

$$3.1(\lambda_i) \cong \lambda_j \qquad (3)$$

What makes this approach useful for generating less predictable behavior is the fact that the additional input is internal to the Agent and thus, not detectable by humans interacting with the scenario. So, for example, while it might be easy for a human participant to learn that "when X happens in the world, the agent will do Z", it is more difficult to learn that "when X happens in the world and the agent's emotional state is Y, the agent will do Z". Primarily, this is difficult to predict is because "Y", the agent's emotional state, is not obvious to the human participant.

7. FUTURE WORK AND DISUCSSION

From a research perspective, future work includes continued experimentation to assist in understanding and improving the theory underlying this model. As we have attempted to make clear throughout this paper, our research is the implementation of one theory of emotions. Many theories exist. Also, other implementations of other theories exist [5],[6]. We offer high-level comparisons of our approach with other approaches in [7].

Ultimately, in the training community for which this was designed, the worth of these models must be measured in terms of improved training. However, we know of no studies that have investigated whether military training is of indeed improved by the use IFORs with these capabilities.

Table 2. Example of Simple Dynamic Test

Event	Agent1(ESE)	Agent2(INP)
At ObsPt	Arousal = .54 Pleasure = .2 Pain = .0 Clar/Conf = .15 **Proposed action(s):** no change **Selected action:** no change	Arousal = .77 Pleasure = .59 Pain = .0 Clar/Conf = .35 **Proposed action(s):** no change **Selected action:** :no change
Observe Enemy	Arousal = .54 Pleasure = .15 Pain = .06 Clar/Conf = .09 **Proposed action(s):** report-observation **Selected action:** report-observation	Arousal = .77 Pleasure = .43 Pain = .14 Clar/Conf = .21 **Proposed action(s):** report-observation **Selected action:** report-observation
Enemy-Sees-Me	Arousal = .54 Pleasure = .03 Pain = .19 Clar/Conf = -.03 **Proposed action(s):** retreat-to-cover, report-observation, shoot **Selected action:** retreat-to-cover	Arousal = .76 Pleasure = .07 Pain = .47 Clar/Conf = -.07 **Proposed action(s):** retreat-to-cover, report-observation, shoot, flee **Selected action:** retreat-to-cover
Partner-Shot	Arousal = .62 Pleasure = .00 Pain = .22 Clar/Conf = -.12 **Proposed action(s):** retreat-to-cover,shoot,report-man-injured **Selected action:** report-man-injured	Arousal = .93 Pleasure = .00 Pain = .62 Clar/Conf = -.28 **Proposed action(s):** retreat-to-cover, shoot **Selected action**: retreat-to-cover
Shooting	Arousal = .87 Pleasure = .0 Pain = .27 Clar/Conf = -.39 **Proposed action(s):** flee, retreat-to-cover **Selected action:** retreat-to-cover	Arousal = .98 Pleasure = .0 Pain = .63 Clar/Conf = -1.0 **Proposed action(s):** flee, freeze **Selected action:** freeze

Based on our research, we believe that the incorporation of an emotions model can make the behavior of less predictable. But, we have not demonstrated that emotional IFOR behavior is more realistic, nor have we demonstrated that the use of emotional IFORs will improve training. It is our opinion and recommendation that somewhere along this vein of research, funding agencies and sponsors of behavior moderator research formally investigate the assumed benefits of incorporating these models into IFOR system.

8. ACKNOWLEDGEMENTS

This work was sponsored by the U.S. Army Research Institute's Emotional Synthetics Forces STTR – Ph II, contract number: DASW01-99-C-0037. That support is gratefully acknowledged.

9. REFERENCES

[1] Chown, E., The Adaptive Power of Affect: Learning in the SESAME Architecture. In J.A. Meyer, H.L. Roitblat, and S. Wilson (eds.), *From Animals to Animats 2: Proceedings of the Second International Conference on Simulation Adaptive Behavior.* Cambridge, MA: MIT Press, 1993.

[2] Damasio, A., Descartes' error: Emotion, reason, and the human brain. New York, NY: Avon, 1995.

[3] D'Ydewalle, G., Ferson, R., and Swerts, A., Expectancy, Arousal, and Individual Differences in Free Recall. *Journal of Memory and Language*. pp. 519-25, 1985.

[4] Eysenck, H.J. and Eysenck, M., *Personality and Individual Differences: A Natural Science Approach.* New York, NY: Plenum Press, 1985.

[5] Gratch, J., Why You Should Buy an Emotional Planner. Proceedings of the Agents '99 Workshop on Emotion-based Agent Architectures (EBAA'99) and ISI Research Report ISI/RR-99-465.

[6] Gratch, J., and Marsella, S., Modeling Emotions in the Mission Rehearsal Exercise. *Proceedings of the 10th Conference on Computer Generated Forces and Behavioral Representation*. May 15-17, 2001, Norfolk, VA. pp. 457-65.

[7] Jones, R. M., Henninger, A.E., and Chown, E., Interfacing Emotional Behavior Moderators with Intelligent Synthetic Forces. *Proceedings of the 11th Conference on Computer Generated Forces and Behavioral Representation*. May, 2002. Orlando, FL.

[8] Kaplan, S., Beyond rationality: Clarity-based decision making. In T. Garling and G. Evans (eds.), *Environment, Cognition, and Action: An Integrative Multidisciplinary Approach*. pp. 171-90. New York, NY: Oxford University Press, 1991.

[9] LeDoux, J.E., Brain Mechanisms of Emotion and Emotional Learning. *Current Opinions in Neurobiology*. Vol(2), pp. 191-7, 1992.

[10] Milner, P., Brain Stimulation Reward: A Review. Canadian Journal of Psychology, Vol(45), pp. 1-36, 1991.

[11] Newell, A. *Unified Theories of Cognition.* Cambridge, MA: Harvard University Press, 1990.

[12] Rosenbloom, P., Laird, J., and Newell, A. *The Soar Papers: Research on Integrated Intelligence.* Cambridge, MA: MIT Press, 1993.

[13] Ropella, G.E.P., Railsback, S.F., and Jackson, S. K., Software Engineering Considerations for Individual-based Models. *Natural Resource Modeling.* Vol(15), no(1), 2002.

[14] Rykiel, E., Testing ecological models: the meaning of validation. *Ecological Modeling.* Vol(90), pp. 229–44, 1995.