Hunting for the Holy Grail With *"Emotionally Intelligent"* Virtual Actors

Clark Elliott

Institute for Applied Artificial Intelligence
School of Computer Science, Telecommunications and Information Systems
DePaul University
243 S. Wabash Ave.
Chicago, IL 60604
elliott@ils.nwu.edu
http://www.depaul.edu/~elliott

Introduction

In his keynote address to the Autonomous Agents 97 conference, Danny Hillis, vice president of research and development at Walt Disney Imagineering, listed four "holy grail" items with respect to entertainment agents:

1. A computable science of emotion,

2. Virtual actors,

3. Agent evolution, and

4. Computable storytelling.

By framing these items in the context of a broad, albeit shallow model of emotion, we make the case that significant progress has been made on three of these (items 1, 2 and 4) in the Affective Reasoning project.

Background

The Affective Reasoner (AR) is a broad platform for research on various aspects of computing emotions. The work is constrained to a descriptive model, based originally on the work of Ortony et al. [Ortony, Clore and Collins, 1988], wherein a broad comprehensive model of human emotion is used as a basis for describing and manipulating the social-emotional fabric of interaction between agents and their perceived world, agents and other agents, and agents and humans. A key element of the "emotionally intelligent" processing that agents perform is that they each have idiosyncratic, dispositional ways they construe the world around them, and manifest responses to internal states that arise. It is from this processing that their relatively rich personalities are formed. A second constraint is that agents do not actually experience emotions themselves since no attempt to model body processes has been made.

Despite the above constraints, AR agents have broad capabilities, some of which address three of the four areas of research mentioned by Danny Hillis in his talk. As a vehicle for presenting background on this work, we will discuss three ways the AR platform has been used: as a general test system for a real–time computable model of emotion [Elliott, 1994b]; as effecting a computable model of storytelling that uses a sophisticated representation of emotion interaction and personality to build a robust, dynamic model of stories [Elliott et al., 1998]; and as supporting theoretically rich, emotionally expressive virtual actors [Elliott, 1997; Elliott, 1994a].

A Computable Model of Emotion

The Affective Reasoner is a collection of AI and multimedia programs. AR agents listen through speech recognition software and respond, in real time, with morphing faces, music and text-to-speech. At the core of the AR is a set of 26 emotion categories, sketched in Figure 1 and based on the original work of Ortony, et al. [Ortony, Clore and Collins, 1988]. Situations arise in an AR agent's world, and are appraised by matching these against internal frames maintained by the agent. The dispositional way in which agents match these situations gives rise to interpretations, represented as sets of variable bindings [Elliott, 1992; Elliott, 1993]. Through a series of about 20 processing modules, these bindings are combined with states maintained internally by each agent, and eventually may themselves give rise to one or more emotion instances from the 26 categories [Elliott, 1992].

The processing in this appraisal stage accounts for agents' abilities to form, e.g., hypotheses about the ways in which other agents are presumed to appraise the world (necessary for fortunes-of-others emotions such as pity) [Elliott and Ortony, 1992]; matches against previous and presumed future world states (necessary for time-relevant emotions such as hope and relief); and compound emotions such as anger (involving thwarted goals, caused by the perceived intentional act of an agent). Processing in this stage includes, among other concepts, representations for the antecedents of emotion intensity (with some subset of about 20 variables relevant to each emotion category) [Elliott and Siegle, 1993]; for agents' moods (non-dispositional, temporary changes in the appraisal mechanism); for relationships between agents; for mixed and even conflicting emotions; and for heuristic classification of situation artifacts for abductive reasoning about the emotional states of others.

Once emotions arise, agents have temperaments that control the ways in which these emotions are manifested in their world. These temperaments are represented as about 20 theoretically based channels of action specific to each emotion (but with overlap between related emotions), ranging from purely somatic responses (such as turning red) at one end of the spectrum, to highly intentional responses at the other end (such as activating a scheme for invoking a plan to get even; but note that any real planning is beyond the scope of this work). The resulting, approximately 440, expression channels are implemented as a rete-like network, and terminal nodes are realized as frames, constructed partially from the original appraisal bindings. A number of processing modules, such as those that choose compatible actions from competing expressions and those that take into account the current states of both the world and the agent, filter the path from emotion instance to emotion manifestation (e.g., one might shout in anger or might deny that there is anything wrong, but would not do both at the same time).

Using these and other devices, sophisticated personalities can be constructed: The appraisal mechanism gives them a rich disposition for construing the world, and the expression component gives them a unique temperament for expressing themselves. Disposition is constructed by encoding the goals (desires), principles (beliefs about right and wrong) and preferences (attractions) of the individual agents, and temperament is constructed by activating certain expression channels, thus allowing us to inspire them with qualities like impatience, talkativeness and shyness. Moods are effected by changing the thresholds for the variable bindings in the match process, and by altering the activation of the expression channels. For related approaches and discussion, see [Picard, 1997; Reilly, 1996; Colby, 1981; Bates, Loyall and Reilly, 1992; Frijda and Swagerman, 1987; Reeves, 1991; Sloman, 1987; Pfeifer and Nicholas, 1985; Scherer, 1993; Toda, 1982; Nass and Sundar, 1994; Nagao and Takeuchi, 1994; Simon, 1967].

A Computable Model of Storytelling

Over the past five years, one aspect of the Affective Reasoning project has been to test the representational coverage of the Ortony et al., and other theories to codify a comprehensive description mechanism for building computable systems. In service of this goal, we have analyzed about 600 emotion scenarios. The Ortony model has proved remarkably robust in this context, with only the addition of (admittedly less theoretically pure) specific categories for love (admiration plus liking), hate (reproach plus disliking), jealousy (resentment, with the goal being an exclusive resource also desired by the appraising agent), and envy (resentment when the agent desires a similar but non-exclusive goal). We felt these additions were required for adequate representation of the corpus of collected situations, at a suitable level of granularity.

One piece of fallout from this research has been the insight that many of the emotion scenarios reviewed make very good stories. In fact, the case can be made that every one of the sce-

GROUP	SPECIFICATION	EMOTION NAME AND TYPE
WELL-BEING	Appraisal of a situation as an event	**Joy:** pleased about an event **Distress:** displeased about an event
FORTUNES-OF-OTHERS	Presumed value of a situation as an event affecting another	**Happy-for:** pleased about an event desirable for another **Gloating:** pleased about an event undesirable for another **Resentment:** displeased about an event desirable for another ***Jealousy:** resentment over a desired mutually exclusive goal ***Envy:** resentment over a desired non-exclusive goal **Sorry-for:** displeased about an event undesirable for another
PROSPECT-BASED	Appraisal of a situation as a prospective event	**Hope:** pleased about a prospective desirable event **Fear:** displeased about a prospective undesirable event
CONFIRMATION	Appraisal of a situation as confirming or disconfirming an expectation	**Satisfaction:** pleased about a confirmed desirable event **Relief:** pleased about a disconfirmed undesirable event **Fears-confirmed:** displeased about a confirmed undesirable event **Disappointment:** displeased about a disconfirmed desirable event
ATTRIBUTION	Appraisal of a situation as an accountable act of some agent	**Pride:** approving of one's own act **Admiration:** approving of another's act **Shame:** disapproving of one's own act **Reproach:** disapproving of another's act
ATTRACTION	Appraisal of a situation as containing an attractive or unattractive object	**Liking:** finding an object appealing **Disliking:** finding an object unappealing
WELL-BEING/ ATTRIBUTION	Compound emotions	**Gratitude:** admiration + joy **Anger:** reproach + distress **Gratification:** pride + joy **Remorse:** shame + distress
ATTRACTION/ ATTRIBUTION	Compound emotion extensions	**Love:** admiration + liking **Hate:** reproach + disliking

* Requires additional information, but is necessary for some stories.

Figure 1: Emotion categories, after Ortony et al., 1988.

narios that fulfills the minimal requirements for the presence of emotion, as computed by our system, also meets the minimal requirements for "storyhood." For example, "The boy sits in the chair" is not a story, but "The boy sits in the chair, but knows that he should not" (containing the theoretical antecedent for shame) might very well have an essential element that does make it a story.

Extending this emotion representation exercise, we formally analyzed real stories for their emotion content according to our computable theory. AR agents then acted out the parts of the characters in the stories according to the structural descriptions of the emotions present. Subjects were able to understand the stories in this context, largely as commonly understood by those simply reading an account of the stories.

At the Autonomous Agents 98 conference [Elliott et al., 1998], we presented evidence that it is possible to automatically generate a great many stories from a single set of external events through a process we call "story-morphing," and that subjects considered the stories to be both distinct and quite plausible. Furthermore, using the multimedia AR agents as virtual actors to present the automatically fabricated stories, we found that on average subjects rated the stories as "makes sense"; more importantly, they showed no statistically significant evidence of ever rating the stories as "could never happen."

In a more recent study, we sought to show that while stories generated using the story-morphing techniques were considered plausible, stories generated using random emotion interpretations of the events in the story were not. In this work, we used the story-morphing techniques to assign appropriate but highly varied emotion interpretations to the externally observable events in a story, on behalf of two characters. Using these techniques, we generated two versions of the "bike race" story that appeared in the Autonomous Agents 98 presentation. Then, using random assignment of emotion as well as emotion intensity to create affective states on behalf of the two characters, we created two additional stories. We hoped to show that the stories generated using random emotion assignment would be seen as significantly less plausible than the stories generated using our story-morphing techniques.

To test this, two different sets of subjects (N = 23, N = 27) assessed two stories of each type. The narrative plot sequences of the stories were the same for all four stories, as was the dialog. For this study, we used printed text rather than agent actors. Descriptions of the characters' emotion states were kept simple and straightforward:

Sample Excerpt From Story One:

Sam and Rick race their bikes around the block. [...]

"Hey. I won the race again," says Rick.

Rick starts to gloat. Sam looks afraid.

Comparative Excerpt From Story Two:

Sam and Rick race their bikes around the block. [...]

"Hey. I won the race again," says Rick.

Sam is somewhat afraid. Rick looks really sorry for Sam.

Subjects were asked to give longhand appraisals of the stories, and at the end to give plausibility ratings for each. Plausibility ratings for the pairs of stories (story-morphed, random) were averaged for each subject. The overall average rating for story-morphed stories fell in the upper region of the interval labeled "seems plausible" whereas the overall average rating for the random emotion stories fell toward the center of the region labeled "does not make sense."

Tests for zero population mean differences between plausibility ratings were conducted using parametric t-tests as well as non-parametric signed rank tests. P-values for these tests were all highly significant (< 0.01) and so provided evidence that stories created using random assignment of emotion interpretation are less plausible on the average than stories created using story-morphing. These results are not sensitive to an assumption of normality. The normal plots were reasonably linear and the Shapiro-Wilk tests for normality were not significant, indicating that the normality assumption for the t-test was reasonable.

These results, combined with the earlier results of our exercises using AR multimedia agents to present story-morphed narratives [Elliott et al., 1998] as well as the study comparing the effectiveness of AR multimedia agents at conveying emotion states through nonverbal means (see "Emotionally Rich Virtual Actors" below), suggest that:

* We are able to use computational techniques to generate large numbers of distinct stories,
* These stories are generally rated as plausible,
* The AR multimedia agents are capable of conveying intended emotion states such as those arising for characters in stories, and
* The emotion states of characters in stories generated by the story-morphing techniques were seen as significantly more plausible than randomly generated emotion states in the otherwise identical stories.

By representing stories based on their emotion content, it is possible that we may have a great deal of flexibility in how we can alter the "story" portrayed by the external events, as long as we have a reasoning system that understands the relationships between the aspects of this representation. To effect this, we simply change the personalities of the agents in the story, and thus their subsequent internal responses to the events that arise. Since the AR tracks roughly 26 categories of emotion and up to 10 intensity variables for each, along with numerous aspects of mood, relationship and the like, a strong case can be made for using this as the basis for an interactive, dynamic storytelling system that has great flexibility in the stories it relates, yet still works under the constraint of maintaining "storyhood" in everything it produces. While this clearly fails to meet the larger goal of true story generation (which would require functional emotions on the part of the agents), it does open the door for significant progress in this area.

Emotionally Rich Virtual Actors

The third "holy grail" item is addressed by a formal study using the Affective Reasoner in virtual-actor mode. AR agents are able to interact with subjects, in real time, using a multimodal approach that includes speech recognition, text-to-speech, real-time morphed schematic faces and music (see Figure 2). In virtual-actor mode, the agents are given varying degrees of stage direction: from explicit instructions (for face, inflection, size, color, location, music selection, and MIDI and audio volume); to somewhat more general instructions such as those used in the study (given the emotion and the text to pick their own faces, music, color, inflection and size); to a degree of freedom (where they participate in picking the emotion based on their personality).

In one virtual-actor presentation, four agents participated in a dialog in various combinations. Two of the agents were "Chicago Bulls fans" and two were "New York Knicks fans." Without varying the text of the dialog, agents were able to make clear their positions as fans and get good agreement from viewers about their relative feelings on the events of the game they were discussing. This was true whether there were two Bulls fans talking, two Knicks fans, one of each, or all four together. An example of the spoken text used is: "I was really worried about the game tonight. I thought Michael Jordan started out really slowly. Then he just wiped the floor with the Knicks in the second half." Any sentence could be spoken by any agent since they were all simply statements of what happened. It was the agents' portrayal of their interpretations of the events described that conveyed the perspective of the character.

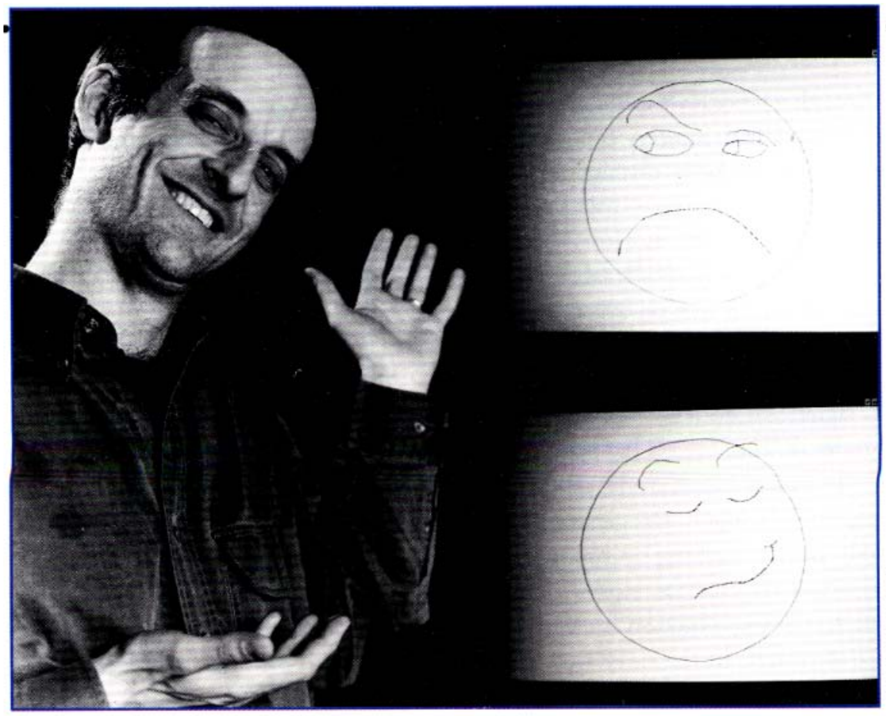
Photo: Robert Drea

Figure 2. Clark Elliott with AR agents.

In another application, children as young as 2 years old, using a speech-driven interface, were able to manipulate storytelling applications using virtual actors to deliver children's stories.

Figure 3. Sample Text From the Formal Study

Here are some sample scenarios:

(Wanda discusses) Butler in the news

Spoken text: "Butler is in the news again today."

Vehicle: Two parts: four positive then eight negative choices, played twice through.

PART A

Gloating: Wanda is gloating because her adversary Butler is again being embarrassed in the news.

Joy: Wanda is joyful because Butler, the congressman she works for, is in the news again.

Happy-for: Wanda is happy for her friend Butler, who is in the news again.

Love: Wanda is in love with Butler, her idol, and she sees him in the news again.

PART B

Hate: Wanda hates Butler, the Nazi party candidate, and she sees him in the news again today.

Anger: Wanda is angry because Butler, one of her subordinates, is again saying damaging things about her in the news.

Fear: Wanda is fearful because Butler, the district attorney who is prosecuting her, is in the news again today.

Reproach: Wanda is reproachful of Butler because he is foolishly talking to reporters, and it is certain just to do him more harm than good.

The Formal Study

The formal study was designed to show that subjects could gather enough information from the AR agents' multimedia communication modalities to correctly assign intended, complex social/emotional meanings to ambiguous sentences, and specifically that this ability would compare favorably with a professional human actor's ability to convey such meanings. In fact, subjects were better able to correctly match videotapes of computer-generated virtual actors with the intended emotion scenarios than they were videotapes of a professional human actor attempting to convey the same scenarios.

Consistent with virtual-actor mode, the formal study does not actually make use of the "emotionally intelligent" components of the agents per se. Nonetheless, it does make a specific case for the usefulness of such agents as they develop, and lends some credence to the theory underlying the agents' development. That is, in this case we show that static, pre-programmed social/emotion content can be effectively communicated by the presentations these agents have at their real-time disposal.

In the study, 141 subjects met for two sessions each, with approximately 14,000 responses analyzed. The subjects were urban undergraduate students of mixed racial and ethnic backgrounds, and primarily upperclassmen. About half were evening students who tended to be over 25 years of age. Three different sets of subjects met. The sessions were undertaken as part of the students' course of study, but they were first exposed to the material as participating subjects before any theoretical material was presented.

The subjects were instructed to match a list of emotion scenarios with a set of videotape presentations in one-to-one correspondence. The lists ranged in length from four to 12 items. The presentations were of "talking heads" (either computer or human) expressing facial emotion content with inflected speech (and, in some of the computer cases, music). The presentations were approximately five seconds long with about 20 seconds between them (and approximately 12 seconds between them for second presentations).

For example, in one set, 12 presentations of the ambiguous sentence, "I picked up Catapia in Timbuktu" were shown to subjects. These had to be matched against 12 scenario descriptions such as: a) Jack is proud of the Catapia he got in Timbuktu because it is quite a collector's prize; b) Jack is gloating because his horse, Catapia, just won the Kentucky Derby and his arch-rival Archie could have bought Catapia himself last year in Timbuktu; and c) Jack hopes that the Catapia stock he picked up in Timbuktu is going to be worth a fortune when the news about the oil fields hits [etc., (d)—(l)]. Other examples of text from the study are shown in Figure 3.

The subjects in the formal study were given seven scenario/interpretation sets. The order of the video presentations of the different interpretations was chosen randomly, but once

Disliking: Wanda sees Butler in the news again, and she really dislikes him.

Sorry-for: Wanda feels really sorry for Butler when she sees him in the news again.

Resentment: Wanda resents the fact that Butler, her opponent, gets coverage in the news again instead of her.

Distress: Wanda is distressed because Butler, another reporter, is in the news again. If she keeps missing the big stories, she knows she will lose her job.

OTHER SCENARIOS

"I can't take any more." Sample, Resentment: Naomi is resentful about watching men in her department get promoted ahead of her even though she does a better job than they do.

"I am again sitting in the chair." Sample, Remorse: The boy is once again outside the principal's office. He is remorseful because he knows he should not have done what he did.

"I see people like that all the time." Sample, Satisfaction: Karen the teacher experiences satisfaction when she is stopped on the street by a former student who wanted to thank her for all he learned in her class.

"I didn't plan for any of this." Sample, Fears-confirmed: Al had had great plans for his life. They all came to a halt when his test results at the hospital confirmed his worst fears.

"I am going to give you the midterm now, but I already have an idea of how well this class is going to do." Sample, Pride: The teacher is quite proud of the job she did preparing the class.

chosen remained constant throughout the study. The ordering was the same for both the computer presentations and the human-actor presentations. The presentation of each interpretation was numbered, and subjects were instructed to write down that number next to the "best" interpretation. The number of presentations was the same as the number of interpretations, resulting in a one-to-one mapping.

The order in which the scenarios were presented to each group of subjects varied only slightly. For the computer presentations, cycles of the three presentation modes (face-only; face and inflection; face-inflection-music) were repeated through the entire set of scenarios (e.g., music appeared once every three presentations).

Confidence factors were additionally recorded for much of the material; subjects rated each of their responses from 1 (low confidence) to 5 (high confidence).

The human actor was coached on the subtleties of the different emotion categories and on what would help to distinguish them. Three to eight takes were made of each interpretation for each scenario. The most expressive take was chosen during editing and a final tape compiled.

The computer was simply given the emotion category and the text, and it automatically selected the face, music and spoken inflection appropriate to that category. Face morphing, speech generation, and music retrieval and synthesis were all done in real time. Actual music selection was up to the program, based on pre-existing categories. The computer presentations were further broken down into the face-only, face and inflection, and face-inflection-music subcategories.

The ratio of time invested between the human-actor version and the computer version was approximately 30:1. Despite this lopsided investment in time, subjects did significantly better at correctly matching videotapes of computer-generated presentations with the intended emotion scenarios (70%) than they did with videotapes of a human actor attempting to convey the same scenarios (53% χ^2 (1, N = 6507) = 748.55; $P < .01$).

Among those participants matching computer-generated presentations to given emotions, there were no significant differences on correct matches between presentation types (face-only = 69%, face and inflection = 71%, face-inflection-music =70%). However, an overwhelming majority of these participants felt that music was very helpful in making a correct match (75%), and another 8% felt that it was somewhat helpful. Less than 3% felt the music was unhelpful or distracting.

One group was asked to rate their confidence after each match. An analysis of their confidence ratings indicated that participants were significantly more confident of matches with displays including music (F (2, 1638) = 19.37, $P < .001$). This could be problematic if music inspired confidence but, in fact, impaired matching ability. A simple look at the proportion of correct matches across the five confidence levels (again, 1 = low confidence; 5 = high confidence) shows that this is not the case. Participants correctly matched 41% of the time when their confidence was "1," 56% of the time when it was "2," 58% of the time when it was "3," 64% of the time when it was "4" and 76% at level 5.

Inflection has not been stressed in either the study or analysis, because the techniques we can support in this area are not very sophisticated. Our best guess is that rudimentary emotion inflection in generated speech enhances the believability of characters.

Discussing the Study

The virtual-actor study tends to support the following points: Computers can be used to convey social information beyond that encoded in text and object representations; this information can be delivered in ways that do not take up bandwidth in the traditional text communication channel (that is, the content measured in the study was explicitly not that encoded in the text); this information can be encoded and delivered in real time; and the computer performs reasonably well on social communication tasks that are difficult for humans.

The preliminary work with music tends to show that music is rated by subjects as having a significant effect on guiding their social perception, but that this effect is not well understood (or, possibly, the musical triggers for this effect are not well understood). We feel that there is strong potential in this area.

Furthermore, the study suggests that the following general principles hold:

- ✖ The underlying-emotion theory is a plausible categorization system to the extent that subjects were able to discriminate the 21 (subset of 26) emotion categories used in the study,
- ✖ The PC platform, despite being inexpensive and commonly available, is viable for studying emotion interaction between humans and computers,
- ✖ The low-bandwidth model we have used (i.e., less than 14 kbps), which shows great promise as a Web-based data collection and delivery mechanism, nonetheless provides sufficiently rich channels for real-time multimodal communication conveying social/emotion content,
- ✖ Potentially useful information can be conveyed about this complex, ubiquitous and yet lightly studied component of human (and human-computer) interaction, and
- ✖ Highly significant reductions in time investment can be achieved for selected, pre-programmed emotion content in "social" scenarios when using multimedia,

multimodal computer presentations in place of human actors in a real-time environment without reduction of the effective content.

While the study showed that the computer actually did better at this restricted task than did the human actor, we are cautious about drawing general conclusions from these results. We used the human actor simply to illustrate that, as designed for the study, correct identification of the broad range of interpretations was a difficult task, and that a 70% identification rate (for the computer-generated presentations) was admirable. Presumably, the professional actor, who was quite skilled, would be at least as good, if not better, at these presentation tasks than a "typical" person from the population.

It is also important to note that the sentences were entirely ambiguous; longhand interpretations given by subjects before the presentations showed no patterns of interpretation. A 70% correct interpretation rate, with no content clues, is thus rather high, considering that in practice the communication of such content completely divorced from cues will be rare.

Additionally, we suggest that, in general, one-time real-life emotion assessment of the sort required here might well be correct less than 70% of the time. People use additional cues to disambiguate situations, they ask questions that help them clarify their interpretations, they observe emotion in a continuous social context (and thus make continual revisions in previous interpretations), and they simply get it wrong much of the time.

Lastly, we specifically made no attempt to give any feedback about the correctness of interpretations during the course of the study. There is the very real possibility that subjects might well learn the specific emotion presentations used by our interactive computer agents, thus raising the identification rate significantly.

Miscellaneous Notes

One issue we had to address in the study was the difference in reading and comprehension time between subjects. We found it necessary to carefully balance forced guesswork on the part of the slower-reading subjects and inattention on the part of the faster-reading subjects. This compromise was made easier when we included music with the presentations because overall the task appeared to be more "fun."

The different numbers of interpretations for the various scenarios arose because certain ambiguous sentences had a greater number of plausible interpretations than others. Additionally, scenarios that had more than four each of positive and negative interpretations were segregated into positive and negative content because trials showed that valence could be easily discriminated by the subjects. The smaller, more similar groupings were preferred because these created an optimal balance between the burden placed on the subjects to read, and comprehend, the different interpretations in the limited amount of time (a burden we sought to reduce), and the difficulty of discriminating subtle differences between similar emotion categories (a difficulty we sought to increase).

While it does not appear in the statistics, one striking anecdotal feature of the study was the change in the testing atmosphere when music was used as part of the presentations. Without the music, subjects tended to be quiet, reserved and studious. With music, the subjects became animated, laughed, made surreptitious comments (although not in ways deemed damaging to the study), and generally responded with vigor to the displays, as though they were more personal.

Conclusion

In the Affective Reasoner, we have made significant progress toward three of the "holy grail" items mentioned by Danny Hillis. Underlying our research premise is the idea that people commonly traffic in social communication, and that much of the human experience revolves around our relationship to our goals, principles and preferences — each of which is central to our emotional lives. We suggest that emotion, as expressed and perceived through various modalities, is computable, and should be accounted for in general-purpose architectures for autonomous agents that interact with users.

Acknowledgement

Preparation of this article was supported in part by Northwestern University's School of Education and Social Policy, and by Andersen Consulting through Northwestern's Institute for the Learning Sciences.

References

Bates, J.; Loyall, A. Bryan; and Reilly, W. S. 1992. Integrating reactivity, goals, and emotion in a broad agent. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society.* Bloomington, IN: Cognitive Science Society.

Colby, K. M. 1981. Modeling a paranoid mind. *The Behavioral and Brain Sciences* 4(4):515–560.

Elliott, C., and Ortony, A. 1992. Point of view: Reasoning about the concerns of others. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society,* 809–814. Bloomington, IN: Cognitive Science Society.

Elliott, C., and Siegle, G. 1993. Variables influencing the intensity of simulated affective states. In *AAAI Technical Report SS-93-05 for the Spring Symposium on Reasoning About Mental States: Formal Theories and Applications,* 58–67. American Association for Artificial Intelligence. Stanford University, March 23-25, Palo Alto, CA.

Elliott, C.; Brzezinski, J.; Sheth, S.; and Salvatoriello, R. 1998. Story-morphing in the affective reasoning paradigm: generating stories semi-automatically for use with "emotionally intelligent" multimedia agents. In *Proceedings of the Second International Conference on Autonomous Agents*, 181–188. Minneapolis, MN: ACM Press.

Elliott, C. 1992. *The Affective Reasoner: A Process Model of Emotions in a Multi-agent System.* Ph.D. Dissertation, Northwestern University. The Institute for the Learning Sciences, Technical Report No. 32.

Elliott, C. 1993. Using the affective reasoner to support social simulations. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, 194–200. Chambery, France: Morgan Kaufmann.

Elliott, C. 1994a. Multi-media communication with emotion-driven "believable agents." In *AAAI Technical Report for the Spring Symposium on Believable Agents*, 16–20. Stanford University: AAAI.

Elliott, C. 1994b. Research problems in the use of a shallow artificial intelligence model of personality and emotion. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, 9–15. Seattle, WA: AAAI.

Elliott, C. 1997. I picked up catapia and other stories: A multimodal approach to expressivity for "emotionally intelligent" agents. In *Proceedings of the First International Conference on Autonomous Agents*, 451–457.

Frijda, N., and Swagerman, J. 1987. Can computers feel? Theory and design of an emotional system. *Cognition & Emotion* 1(3):235–257.

Nagao, K., and Takeuchi, A. 1994. Social interaction: Multimodal conversation with social agents. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, 9–15. Seattle, WA: AAAI.

Nass, C., and Sundar, S. S. 1994. Is human-computer interaction social or parasocial? Stanford University. Submitted to *Human Communication Research.*

Ortony, A.; Clore, G. L.; and Collins, A. 1988. *The Cognitive Structure of Emotions.* Cambridge University Press.

Pfeifer, R., and Nicholas, D. W. 1985. Toward computational models of emotion. In Steels, L., and Campbell, J. A., eds., *Progress in Artificial Intelligence.* Ellis Horwood, Chichester, U.K. 184–192.

Picard, R. 1997. *Affective Computing.* Cambridge, MA: MIT Press.

Reeves, J. F. 1991. Computational morality: A process model of belief conflict and resolution for story understanding. *Technical Report UCLA-AI-91-05*, UCLA Artificial Intelligence Laboratory.

Reilly, W. S. 1996. *Believable Social and Emotional Agents.* Ph.D. Dissertation, CMU.

Scherer, K. 1993. Studying the emotion-antecedent appraisal process: An expert system approach. *Cognition & Emotion* 7(3):325–356.

Simon, H. A. 1967. Motivational and emotional controls of cognition. *Psychological Review* 74:29–39.

Sloman, A. 1987. Motives, mechanisms and emotions. *Cognition & Emotion* 1(3):217–234.

Toda, M. 1982. *Man, Robot and Society.* Boston, MA: Martinus Nijhoff Publishing. ■