DRAFT: I picked up Catapia and other stories: A multimodal approach to expressivity for "emotionally intelligent" agents.*

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

July 8, 1996

Abstract

In our research we make use of "emotionally intelligent" agents as part of a collection of AI programs we refer to as the Affective Reasoner. These agents are able to interact with subjects using a multimodal approach which includes speech recognition, text-to-speech, real-time morphed schematic faces, and music. In a recent study we hoped to show that users could gather enough information from the agents' different communication modalities to correctly assign intended (social, emotional) meanings to ambiguous sentences, and specifically that this ability would compare favorably with a human actor's ability to convey such meanings.

In fact, 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% $\chi^2(1, N = 6507) = 748.55, P < .01$).

*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.

1 Introduction

In this short paper we present findings from a recent study on the use of multimedia presentations to communicate emotion content in otherwise ambiguous statements. In the broader scope of our research these presentations are used by "emotionally intelligent" autonomous agents as some of the ways of expressing their internal states to human users. The background context in which these agents operate is covered elsewhere and will only be minimally discussed here (c.f. [Elliott, 1992; Elliott, 1993; Elliott, 1994b; Marquis and Elliott, 1994; Elliott and Siegle, 1993; Elliott and Ortony, 1992; Elliott, 1994a; Elliott, 1994c]). These works also cover background on the area (see [Colby, 1981; Elliott, 1994b; Bates *et al.*, 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] for interesting approaches to related problems).

Very briefly, the Affective Reasoner is a collection of (mostly Common Lisp) programs embodying a theory of emotion proposed by Ortony, Clore, and Collins in their 1988 book, *The Cognitive Structure of Emotions* [Ortony *et al.*, 1988]. As manifested in the Affective Reasoner, this theory has twenty-four distinct emotion categories (from the original twenty-two) each of which contains many different actual emotions, a frame-based appraisal mechanism, an action-expression component, a set of intensity variables for each of the categories, and so forth. Agents operate in real time, perform appraisals of the world according to their own dispositional personality characteristics, respond to their internal states according to their temperaments, and perform rudimentary case-based, and logical, reasoning about other agents and the world around them.

To communicate with users the agents use music, text-to-speech, a set of approximately seventy faces (which are morphed in real time), and speech recognition. Children as young as two years old have been able to manipulate story-telling applications based on the system.

The study discussed here does not actually make use of the "intelligent" components of the agents per se. Nonetheless it does make a specific case for the usefulness of such agents as they develop, and lend some credence to the theory underlying the agents' development. That is, in this case we showed that static, pre-programmed social/emotion content can be effectively communicated by the presentations these agents have at their (real-time) disposal. Since our larger body of work establishes a relatively robust coverage of the emotion categories used in this study, and since these categories can be directly manipulated by our autonomous agents, the conclusion we hope will be drawn is that communication from "emotionally intelligent" computer agents (whatever form they ultimately take) to human users is both practical, and plausible.

2 The Study

There were 141 subjects that met for two sessions each, with approximately 14,000 responses analyzed. The subjects were urban undergraduate students of mixed racial and ethnic backgrounds, primarily upperclassmen. About half were evening students who tended to be over twenty-five years of age. Three different sets of subjects met. The studies were undertaken as part of the course of study, but students were first exposed to the material as participating subjects before any theoretical material was presented. The subjects were given tasks wherein they 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 twelve items. The presentations were approximately five seconds long with about twenty seconds between them (and approximately twelve seconds for second presentations). The presentations were of "talking-head" type (either computer or human) expressing facial emotion content with inflected speech (and in some of the computer cases, music).

For example, in one set, twelve presentations of the ambiguous sentence, "I picked up Catapia in Timbuktu," were shown to subjects. These had to be matched against twelve 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 archrival 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)].

Five minutes of instructions were given before the first session. These included verbal instructions, and a simple two-part practice session with a videotape talking-head computer presentations. Furthermore, written instructions were given at the top of each printed answer sheet, of the general form: "When the video begins, write the number of the video episode next to the sentence that best describes the emotion [Naomi] is expressing. (played twice)" The computer video display used an MS-Windows window with the name of the speaking character appearing in the title bar.

Confidence factors were additionally recorded for much of the material where subjects rated each of their responses from "1" (not confident) to "5" (highly confident).

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 face-only, face and inflection, and face-inflection-music sub-categories in the study.

The ratio of time invested between the human-actor version and the computer version was approximately 30:1.

Overall, 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% $\chi^2(1, N = 6507) = 748.55, P < .01$).

Among those participants matching computer-generated presentations to given emotions, there were no differences on correct matches between presentation types (face = 69%, face plus intonation = 71%, face plus intonation plus music = 70%). However, an overwhelming majority of these same 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 5 confidence levels shows that this is not the case. On a scale where "1" means low confidence and "5" means high confidence, these 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% of the time when it was "5".

Other results based partly on the coding of long-hand responses are not presented as part of this short paper.

3 Discussion

What the presentation studies tend to show is that (1) computers can be used to convey social information beyond that encoded in text and object representations, (2) that 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 studies was explicitly *not* that encoded in the text), (3) that this information can be encoded and delivered in real time, and (4) that the computer performs reasonably well on social communication tasks that are difficult for humans.[1]

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.

[1]While the computer did better in these studies than did the human actor, we prefer to use this simply as a guide to assessing the difficulty of the task rather than for making broad generalizations. See below.

Furthermore, the studies suggest the following: (1) That the underlying emotion theory is a plausible categorization system to the extent that subjects were able to discriminate the twenty-one different emotion categories used in the study. (2) That despite it being inexpensive, and commonly available, this is a viable platform for studying emotion interaction between humans and computers. (3) That the low-bandwidth model we have used (i.e., less than 14K bps), 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. (4) That potentially *useful* information can be conveyed about this complex, ubiquitous, and yet lightly studied component of human (and human-computer) interaction. (5) 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 our results showed that the computer actually did better at this restricted task than did the human actor, we are cautious about drawing general conclusions from this. Questions arise: (1) How good was the actor? (2) How does one measure "goodness" in an actor? (3) How appropriate was the actor for this medium, this audience, and this task? (4) How much does professional lighting, editing, and sound mixing of the human-actor presentations effect the identification task? It would be possible to control for these factors through, e.g., measuring the effectiveness of different actors for these specific tasks, seeking funding for [expensive!] professional studio time, and so forth. If this were to be pursued we might be able to make some claims about the computer being "better" at conveying social/emotion content in some situations than humans. However, this was not our goal. 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 seventy percent identification rate was admirable.

We can also address some of these questions from a common sense perspective: Presumably our professional actor, who has spent long years honing such skills, and who was specifically coached about how to discriminate the different emotion categories (e.g., was told to use technique "A" instead of "B" – both of which were valid – for expressing a specific interpretation because it would be less likely to be confused with another interpretation to be presented later) ...presumably this professional would be at least as good at these tasks as a "typical" person from the population. (Anecdotally, the actor was quite good, showing an impressive range of expressiveness and flexibility in addressing the task.)

It is important, also, to note that the sentences were entirely ambiguous: longhand ad hoc interpretations given by subjects before the presentations were given showed no patterns of interpretation whatsoever. A seventy percent correct interpretation rate, *with no content clues*, is rather high, considering that in practice such cues, divorced from content, 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 seventy percent of the time. People use additional cues to disambiguate situations, they ask questions that help them to 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 a very real possibility that subjects might well learn the specific emotion presentations used by our interactive computer agents, thus raising the identification rate significantly.

4 Miscellaneous notes

One issue we had to address in the study was the difference in reading and comprehension time between students. From preliminary trials it became clear that, on the one hand, if the presentations appeared too rapidly the identification task deteriorated into simply a reading task, with the component we were attempting to isolate driven largely by "rapid guessing." On the other hand, if we paused for too long a period between presentations, while this clearly helped some of the students, others soon became bored and inattentive (but strikingly less so when presentations included music – see below). It is our best guess that the compromise reached still caused confusion and pure guesswork for some responses in the slower-reading students (confusion which would not be present had we given them more time), and inattention in some of the faster students.

In an attempt to reduce the burden placed on students to recall, and manipulate, the different interpretations listed on the answer sheet, we found it expedient to use emotion-category labels. In trials this appeared to give us the best balance between, on the one hand, reduction in range of scenario identification and comprehension times between the fastest and slowest readers, and on the other hand truest matching of emotion *content* in each interpretation. Ideally we would have preferred to have left the labels out altogether, instead including the specific emotion category label in the text itself (as done in the Catapia examples).

In one session with the one-part Catapia scenario (see below), we sought to show differences in comprehension with music when the presentations were presented rapidly, thus putting the majority of the students under duress. The hypothesis was that music might allow them to rapidly make an improved guess at the emotional content when snap judgments were required. We did not show any significant results. Our assessment is that this was because the task was simply too difficult and that such an exercise would have to be carefully controlled for reading speed, and ethnic/age differences (regarding the music selections) or else designed differently.

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 relatively 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, 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.

A follow-up study measuring the effects of music on (1) learning emotion cues of the emotion presentations, and (2) postponing fatigue when interacting with such agents might well show results.

5 A low-bandwidth approach suitable for the World Wide Web

We are currently integrating our work with the world-wide-web. All aspects of the presentations (midi music, morphing faces, text-to-speech) have been tested as applications which run (transparently to the calling modules) as either local or remote applications, where remote applications are established through the Web. Licensing agreements have been considered so that text-to-speech is reduced to Realaudio format before it is transmitted. Higher-quality, lower-bandwidth reproduction is available if the client has an AT&T text-to-speech license. Combined transmission of the real-time signal is under 14k bps.

While not central to the theoretical component of our work, we feel that the fact that our emotion reasoning, and presentation, mechanisms can be integrated into a web-based environment allows for significant data collection possibilities, and opens up additional applications. Over the years we have consistently operated under the constraints imposed by using a low-bandwidth approach, supported by inexpensive hardware. Because of this we are able to speculate on the very real possibility of constructing real-time, truly multimodal, interactive internet applications that operate at a social level.

Various methods have been used, varying from client-resident Lisp interpreters, to small multi-port routing modules called from Web-clients, to Java applications. The delivery mechanism is less important than the ratio of usable social information to number of bits, one which we have shown to be effective over a 14.4 modem.

We have additionally run trials using Realaudio-encoded signals as input to the speech-recognition package and believe this to be a viable mechanism for additionally running the speech recognition components of our research over the web.

6 The text of the study

Subjects were given the following scenario/interpretation sets. The order of the video presentations of the different interpretations was chosen randomly, but once 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 three presentation modes (face only; face, and inflection; face, inflection, and music) were repeated through the entire set of scenarios (e.g., music appeared once every three presentations).

6.1 Naomi can't take any more

Spoken text: "I can't take any more."
Vehicle: One part, seven choices, played twice through.

LIKING: Naomi really likes food and gourmet desserts, and is just finishing a huge meal at a top restaurant in the city.

SORRY-FOR: Naomi feels sorry for her son. She has to stand by and watch him suffer as he is getting a cast put on a broken arm, and there is nothing she can do.

FEAR: Naomi lives in constant fear for her life and just received another death threat.

HATE: Naomi hates the despicable drug addicts who have overrun her neighborhood. She is going to fight back.

ANGER: Naomi is angry at her teenage daughter's behavior, and confronts her when she comes home two hours late from a date, again.

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.

DISLIKING: Naomi has been cleaning up a flooded basement, and she really dislikes the smell.

6.2 The boy is sitting in the chair.

Spoken text: "I am again sitting in the chair."
Vehicle: One part, nine choices, played twice through.

NEUTRAL: The boy was sitting in the chair again, but did not care one way or the other about it.

FEAR: The boy found himself in the chair again by accident, and is afraid of what will happen when it is discovered by the authorities.

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.

JOY: The boy finds that he is sitting in the chair again and is really happy about it.

GLOATING: The boy is gloating over the girl because he gets to sit in the chair again, and she does not.

PRIDE: The boy is proud of himself for again being chosen to sit in the honor chair.

SATISFACTION: The boy really wanted to sit in the chair, and is satisfied now because he gets to.

LOVE: The boy is sitting in the chair next to someone he loves.

LIKING: The boy is sitting in the chair and he really likes to.

6.3 Karen sees people like that

Spoken text: "I see people like that all the time"
Vehicle: Two parts, four positive, then eight negative choices, played twice through.

Part A

LOVE: Karen the lifeguard meets another very attractive young man at the beach and momentarily feels love for him.

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.

LIKING: Karen the optimist just met yet another person that she liked very much.

GLOATING: Karen gloats over an opponent she just trounced in a chess match — after the opponent had foolishly bragged about how he was going to beat Karen.

Part B

RESENTMENT: Karen is resentful because she just saw another woman smiling after getting the job that Karen wanted.

FEARS-CONFIRMED: Karen had been worried that the strange man was going to rob her, and once again her fears were confirmed.

DISLIKING: Karen tells a friend how much she dislikes the smell of a panhandler she passes every day on her way to work.

ANGER: Karen tells a friend how angry she is at a panhandler because he bothers her every day on her way to work instead of getting a job.

DISTRESS: Karen is distressed because her favorite uncle died.

SORRY-FOR: Karen tells a friend how sorry she is for a panhandler she sees every day on her way to work.

REPROACH: Karen is a department store clerk. She is reproachful of a customer who broke a toy and then returned it saying it was defective.

HATE: Karen tells a friend that she hates the bigot standing on the street corner yelling filth at people who pass by.

6.4 Al didn't plan

Spoken text: "I didn't plan for any of this"
Vehicle: Two parts, four positive, then seven negative choices, played twice through.

Part A

RELIEF: Al had resigned himself to loosing his job, but it turned out his boss got fired instead.

HOPE: After the letter he got from Rita, Al was hopeful that he would get her back.

LOVE: Al thought Rita would just be a good friend, but he really fell in love with her.

LIKING: It turned out that Al really liked the climate on the French Riviera after all.

Part B

REMORSE: Al started a garbage can fire as a practical joke, and a friend of his ended up getting hurt.

FEAR: Al took Rita out for a night on the town, but they were afraid because they ended up in a scary neighborhood with a flat tire.

ANGER: Al was angry at his wife because his carefully planned dinner party was ruined when his wife invited three extra couples, and moved the starting time up an hour.

REPROACH: Al was reproachful because as Sarah's academic advisor he had laid out a careful plan for her studies. She had disregarded his advice, however, and now needed his help to graduate.

DISTRESS: Al had a real string of bad luck and his life was a mess.

RESENTMENT: Al set Rita up with Joe because he figured Joe was such a pig that Rita would gladly come back to him after that. But it backfired, and Joe and Rita told him they were going to get married.

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.

6.5 (Wanda discusses) Butler in the news

Spoken text: "Butler is in the news again today."
Vehicle: Two parts, four positive, then four negative choices, played twice through.

Part A

GLOATING: Wanda is gloating because her adversary Butler is 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 to just do him more harm than good.

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.

6.6 Teacher gives the midterm

Spoken text: "I am going to give you the midterm now, but I already have an idea of how well this class is going to do."
Vehicle: Two parts, four positive, then five negative choices, played twice through.

Part A

REPROACH: The teacher is reproachful of the class because they did not study.

ANGER: The teacher is angry because she knows how badly it will reflect on her when the dean sees their final exam scores.

FEAR: The teacher is really afraid many students will fail.

DISTRESS: The teacher is very distressed about how poorly the class is prepared.

Part A

HOPE: The teacher is hopeful the class will do well.

ADMIRATION: The teacher admires how well the class has prepared for the exam.

PRIDE: The teacher is quite proud of the job she did preparing the class.

GLOATING: The teacher thinks the class has been very lazy preparing for the exam, and now they are going to pay for it.

SATISFACTION: The teacher is satisfied at the way the class has prepared for the exam.

6.7 Catapia – two parts

Spoken text: "I picked up Catapia in Timbuktu"
Vehicle: Two parts, six positive, then six negative choices, played twice through.

Part A

- Jack is proud of the Catapia he got in Timbuktu because it is quite a collector's prize.
- Jack picked up his fiancee, Catapia, in Timbuktu, and is in love with her.
- Jack is gloating because his horse, Catapia, just won the Kentucky Derby and his archrival Archie could have bought Catapia himself last year in Timbuktu.
- Jack is really joyful about picking up the Catapia, because it has worked out great.
- Jack and his friend Sue are listening to a Catapia recording. They really like it. He picked it up in Timbuktu.
- 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.

Part B

- Jack is really angry that he had to go all the way to Timbuktu to pick up his daughter Catapia.
- Jack's worst fears were confirmed when he realized it was catapia he picked up in Timbuktu.
- Jack picked up his friend Catapia in Timbuktu. She has malaria. Jack feels sorry for her.
- Jack resents Bill, because Bill got gold in Timbuktu, but Jack only got catapia.
- Jack picked up an embarrassing disease, Catapia, in Timbuktu, and is ashamed.
- Jack is afraid that the Catapia he picked up might prove to be really serious.

6.8 Catapia – one part

Spoken text: "I picked up Catapia in Timbuktu"

Vehicle: One part, twelve choices, played twice through.

- Jack is really angry that he had to go all the way to Timbuktu to pick up his daughter Catapia.
- Jack's worst fears were confirmed when he realized it was catapia he picked up in Timbuktu.
- Jack is proud of the Catapia he got in Timbuktu because it is quite a collector's prize.
- Jack picked up his fiancee, Catapia, in Timbuktu, and is in love with her.
- Jack is gloating because his horse, Catapia, just won the Kentucky Derby and his archrival Archie could have bought Catapia himself last year in Timbuktu.
- Jack picked up his friend Catapia in Timbuktu. She has malaria. Jack feels sorry for her.
- Jack is really joyful about picking up the Catapia, because it has worked out great.
- Jack resents Bill, because Bill got gold in Timbuktu, but Jack only got catapia.
- Jack and his friend Sue are listening to a Catapia recording. They really like it. He picked it up in Timbuktu.
- 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.
- Jack picked up an embarrassing disease, Catapia, in Timbuktu, and is ashamed.
- Jack is afraid that the Catapia he picked up might prove to be really serious.

7 Closing

People commonly traffic in social communication. Much of the human experience revolves around our relationship to our goals, our principles, and our preferences – all of which are antecedents of emotions. This study illustrates that many possibilities exist for including emotion content in communications between computer agents and humans. We suggest that such content, expressed, and perceived through various modalities, should be one of the goals in an ideal, yet plausible, architecture for a general-purpose autonomous agent.

References

[Bates *et al.*, 1992] Joseph Bates, A. Bryan Loyall, and W. Scott Reilly. Integrating reactivity, goals, and emotion in a broad agent. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, Bloomington, IN, August 1992. Cognitive Science Society.

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

[Elliott and Ortony, 1992] Clark Elliott and Andrew Ortony. Point of view: Reasoning about the concerns of others. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, pages 809–814, Bloomington, IN, August 1992. Cognitive Science Society.

[Elliott and Siegle, 1993] Clark Elliott and Greg Siegle. Variables influencing the intensity of simulated affective states. In *AAAI technical report for the Spring Symposium on Reasoning about Mental States: Formal Theories and Applications*, pages 58–67. American Association for Artificial Intelligence, 1993. Stanford University, March 23-25, Palo Alto, CA.

[Elliott, 1992] Clark Elliott. *The Affective Reasoner: A Process Model of Emotions in a Multi-agent System.* PhD thesis, Northwestern University, May 1992. The Institute for the Learning Sciences, Technical Report No. 32.

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

[Elliott, 1994a] Clark Elliott. Multi-media communication with emotion-driven 'believable agents'. In *AAAI Technical Report for the Spring Symposium on Believable Agents*, pages 16–20, Stanford University, March 1994. AAAI, American Association for Artificial Intelligence.

[Elliott, 1994b] Clark Elliott. 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*, pages 9–15, Seattle, WA, August 1994. AAAI, American Association for Artificial Intelligence.

[Elliott, 1994c] Clark Elliott. Two-way communication between humans and computers using multi-media extensions to the IBM PC, and a broad, shallow, model of emotion. Draft of Technical Report in preparation, 1994.

[Frijda and Swagerman, 1987] Nico Frijda and Jaap Swagerman. Can computers feel? theory and design of an emotional system. *Cognition & Emotion*, 1(3):235–257, 1987.

[Marquis and Elliott, 1994] Stuart Marquis and Clark Elliott. Emotionally responsive poker playing agents. In *Notes for the Twelfth National Conference on Artificial Intelligence (AAAI-94) Workshop on Artificial Intelligence, Artificial Life, and Entertainment*, pages 11–15. American Association for Artificial Intelligence, August 1994.

[Nagao and Takeuchi, 1994] Katashi Nagao and Akikazu Takeuchi. Social interaction: Multimodal conversation with social agents. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, pages 9–15, Seattle, WA, August 1994. AAAI, American Association for Artificial Intelligence.

[Nass and Sundar, 1994] Clifford Nass and S. Shyam Sundar. Is human-computer interaction social or parasocial? Stanford University. Submitted to Human Communication Research, 1994.

[Ortony *et al.*, 1988] Andrew Ortony, Gerald L. Clore, and Allan Collins. *The Cognitive Structure of Emotions.* Cambridge University Press, 1988.

[Pfeifer and Nicholas, 1985] Rolf Pfeifer and D. W. Nicholas. Toward computational models of emotion. In L. Steels and J. A. Campbell, editors, *Progress in Artificial Intelligence*, pages 184–192. Ellis Horwood, Chichester, UK, 1985.

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

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

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

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

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