Agents with Faces:
A Study on the Effects of Personification of Software Agents

Tomoko Koda

B.A. in Human Engineering
Osaka University, Japan
March 1985

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of
Master of Science in Media Arts and Sciences
at the Massachusetts Institute of Technology

September 1996

Signature of Author _____
Program in Media Arts and Sciences
August 9, 1996

Certified by _____
Pattie Maes
Associate Professor of Media Technology
Program in Media Arts and Sciences

Accepted by _____
Stephen A. Benton
Chairperson
Departmental Committee on Graduate Students
Program in Media Arts and Sciences

Agents with Faces:
A Study on the Effects of Personification of Software Agents

Tomoko Koda

B.A. in Human Engineering
Osaka University, Japan
March 1985

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
on August 9, 1996
in partial fulfillment of the requirements for the degree of
Master of Science in Media Arts and Sciences
at the Massachusetts Institute of Technology

Abstract

It is still an open question whether software agents should be personified in the interface. In order to study the effects of faces and facial expressions in the interface, a series of experiments was conducted to compare subjects' responses to and evaluation of different faces and facial expressions.
The experimental results obtained demonstrate that: 1) personified interfaces help users engage in a task, and are well suited for an entertainment domain. 2) people's impressions of a face in a task are different from ones of the face in isolation. Perceived intelligence of a face is determined not by the agent's appearance but by its competence. 3) there is a dichotomy between user groups which have opposite opinions about personification.
Thus, agent-based interfaces should be flexible to support the diversity of users' preferences and nature of tasks.

Thesis Supervisor: Pattie Maes
Title: Associate Professor of Media Technology

This research was supported by Apple computer and British Telecom.

Agents with Faces:
A Study on the Effects of Personification of Software Agents

Tomoko Koda

The following people served as readers for this thesis:

Reader _____

Mitchel Resnick
Associate Professor of Media Arts and Sciences
Program in Media Arts and Sciences

Reader _____

Ronald MacNeil
Principal Research Associate
MIT Media Laboratory

Acknowledgments

I would like to thank my advisor, Pattie Maes, for her professional advice, strong encouragement, and being my role model in all respects, as well as being one of the poker playing characters.

I would like to thank my readers, Mitchel Resnick for sharp insight for designing the experiments and encouragement in all respects, and Ron MacNeil for helping me broaden my point of view.

Thanks to Richard Bolt for giving me invaluable support in the beginning stages.

Thanks to my office mate, Anthony Chavez for helping me with debugging, keeping me sane, and enduring my continuous interruptions.

Thanks to the members of the Autonomous Agents Group for participating in the pilot experiment and giving me many useful comments on how to improve the user interface.

Thanks to my UROP crew, Adrian Soviani who helped me with Java programming, Kuo-Hsiung "Hanson" Wong, who helped me collect the data and edit this thesis.

Thanks to all subjects from all over the world who gave me lots of useful comments.

Thanks to Liz Manicatide for drawing such artistic, expressive and fun characters. Without her help, this whole study could not happen.

Thanks to Linda Peterson and Santina Tonelli for supporting me in all respects. Without their sincere support, I had not been able to complete my study at the Media Lab.

Thanks to Roz Picard for opening up my eyes toward affective computing.

Thanks to Minoru Kobayashi and Akira Kotani for giving me suggestions for forming my research topic and helping me not forget Japanese.

Thanks to Kris Thorisson for helping me form a basic idea of my research and cheering me up.

Suguru Ishizaki provided professional comments early on in the proposal process.

Ellen Hoffman gave me lots of advice as an experienced poker player.

Jocelyn Riseberg helped me battle with statistics.

Thanks to Janet Cahn and Sara Elo Dean for their friendship, helping me with editing, giving lots of encouragement, feeding me, and most of all, sharing precious and memorable days at the Media Lab.

FInally, I would like to thank to my father, my sister, and my mother-in-the-sky, who have always given me priceless support.

Table of Contents

Acknowledgments **7**

CHAPTER 1 *Introduction* **17**

Motivation for This Research -A Need for Agent-based Interfaces **17**
Software agents **17**
Agent-based Interface **18**
Contributions of this Research **21**
Overview of This Document **21**

CHAPTER 2 *Related Work in Personified Interfaces* **23**

Metaphors with Character **23**
Apple Knowledge Navigator **24**
Apple Guides **24**
Research Related to "Social Interface" **25**
Microsoft Persona Project **26**
Microsoft "Bob" **26**
Research on Believable and Lifelike Agents **27**
MIT Media Lab's "ALIVE" System **27**
Bates' "Oz project" **27**
Elliott's "The Affective Reasoning Project" **28**
Personifying Software Agents **29**
MIT Media Lab's "Newt" **29**
MIT Media Lab's "Maxims" and the Calendar Agent **30**
Research on Users' Impression of Personified Interfaces **31**
Walker et al.'s Study of Using a Human Face in a Questionnaire Survey **31**
Takeuchi et al.'s "Situated Agent" **32**
Brennan et al.'s Study on Effects of Message Style **32**
King et al.'s Study on Appraisal of Agents **33**
Isbister's Study on Perceived Intelligence **33**
Need for User Interface Studies in Actual Applications **34**

CHAPTER 3 *Poker Game Design* 35

Why a Poker Game? 35
Designing Characters 36
Designing Facial Expressions 38
Selecting Emotion Types -the OCC model- 39
Applying the Emotion Model to the Poker Game Environment 43
Different Modes of Expressiveness 44
Designing the Interface 45
System Design 47
Overall Architecture 47
Poker Playing Strategy 48

CHAPTER 4 *Design of Experiments* 51

Lessons Learned from the Test Experiment 51
Design of Web-based Experiments 53
Purpose 53
Subjects 54
Procedure 54
Conditions for the Web-based Experiments 56
Questionnaire 63
Pilot Experiment 64

CHAPTER 5 *Results and Discussion* 67

Impressions Based on Visual Appearance (Pilot Experiment) 68
Results of GENDER_LOOK 68
Results of HUMANITY_LOOK 69
Results of REALISM_LOOK 72
Hypotheses 76
Analysis of Competence 77
Difference between Face and NoFace (Experiment 1) 79
Results 79
Discussion 81
Difference between Characters's Gender (Experiment 2) 82
Results 82

Difference between Humanity (Experiment 3) **85**
Results ***85***
Discussion ***91***
Difference between Realism (Experiment 4) **92**
Results ***92***
Discussion ***95***
Difference between Levels of Expressiveness (Experiment 5) **96**
Results ***96***
Discussion ***99***
Observations and Open Comments **100**
Subjects' Reactions ***100***
Suggestions for Improving the User Interface ***101***

CHAPTER 6 *Conclusions and Future Directions* ***103***

Conclusions **103**
The Effect of Having a Face ***103***
Need for Understanding the Context of an Application ***104***
Need for Understanding the Users ***106***
Toward Making More Efficient Personified Interface **106**
Future Directions **107**

APPENDIX 1 *Facial Expressions of the Poker Playing Agents* ***109***

APPENDIX 2 *Poker Game Instructions* ***113***

Consent Form **113**
Purpose of the Project ***113***
Procedure of the Experiment ***114***
Consent Declaration ***114***
Poker Playing Rules **115**
Procedure ***115***
Poker Hand Ranking (high to low) ***116***
How to Play ***117***
Warnings ***118***

APPENDIX 3 *Questionnaires* ***119***

Common Questionnaire for the Web-based Experiments **120**

Specific Questionnaires for the Web-based Experiments **121**

FACE experiment ***121***

GENDER experiment ***122***

HUMANITY experiment ***123***

REALISM experiment ***124***

EXPRESSIVENESS experiment ***125***

Questionnaire for the Pilot Experiments **126**

Common Introduction ***126***

GENDER_LOOK experiment ***127***

HUMANITY_LOOK experiment ***127***

REALISM_LOOK experiment ***128***

References ***129***

List of Tables

TABLE 1. Facial features of the poker playing agents. 37
TABLE 2. The facial representations used in each experiment 57
TABLE 3. The questions asked in each experiment. 64
TABLE 4. The mean value of variables for GENDER condition (Male vs. Female) based on appearance. 69
TABLE 5. The mean value of variables for HUMANITY condition (Human vs. Dog) based on appearance. 70
TABLE 6. The mean value of variables for REALISM conditions based on appearance. 73
TABLE 7. The mean ending money of subjects and the poker playing agents 77
TABLE 8. The mean value if variables for FACE condition (Face vs. NoFace). . . . 80
TABLE 9. The mean value of variables for GENDER conditions (Male vs. Female). 83
TABLE 10. The mean values of variables for HUMANITY conditions (Human vs. Dog). 87
TABLE 11. The mean value of variables for REALISM conditions (Smiley vs. Caricature vs. Realistic). 93
TABLE 12. The mean value of variables for EXPRESSIVENESS conditions (Honest vs. Deceiving vs. Stoic). 97

List of Figures

FIGURE 1. Two kinds of representation of "Phil" in the Apple Knowledge Navigator. . . . 24
FIGURE 2. The Guides characters. . . . 25
FIGURE 3. The representation used in Newt. . . . 29
FIGURE 4. The facial representation used in Maxims and the calendar agent 30
FIGURE 5. Examples of Chernoff faces 33
FIGURE 6. The graphical representations of the poker playing agents. . . . 38
FIGURE 7. Ten facial expressions of the female Caricature face 39
FIGURE 8. Global structure of emotion types in the OCC model adopted from [Ortony 88]. . . . 41
FIGURE 9. The emotion types for the poker game 42
FIGURE 10. Web interface of the simulated poker table 46
FIGURE 11. Poker game system structure 48
FIGURE 12. Betting strategy used in the poker game. . . . 50
FIGURE 13. The poker game environment for the FACE experiment.. . . . 58
FIGURE 14. The poker game environment for the GENDER experiment 59
FIGURE 15. The poker game environment for the HUMANITY experiment 60
FIGURE 16. The poker game environment for the REALISM experiment 61
FIGURE 17. The poker game environment for the EXPRESSIVENESS experiment.. 62
FIGURE 18. The mean value of INT for the Human and Dog's face based on appearance. . . . 70
FIGURE 19. The mean value of LIKE for the Human and Dog's face based on appearance. . . . 71
FIGURE 20. The mean value of ENG for the Human and Dog's face based on appearance. . . . 71
FIGURE 21. The mean value of INT for Smiley, Caricature, and Realistic face based on appearance. . . . 74
FIGURE 22. The mean value of LIKE for Smiley, Caricature, and Realistic face based on appearance. . . . 74
FIGURE 23. The mean value of ENG for Smiley, Caricature, and Realistic face based on appearance. . . . 75

FIGURE 24. The mean value of INT for Smiley, Caricature, and Realistic face based on appearance, for the AGREE and DISAGREE group.. 75
FIGURE 25. The mean value of LIKE for Face and NoFace. 80
FIGURE 26. The mean value of ENG for Face and NoFace. 81
FIGURE 27. The mean value of COM for Face and NoFace. 81
FIGURE 28. The mean value of INT for the Male and Female face, for both the AGREE and DISAGREE groups. 84
FIGURE 29. The mean value of LIKE for the Male and Female face, for both the AGREE and DISAGREE groups. 84
FIGURE 30. The mean value of LIKE for the Human and Dog's face, for both the AGREE and DISAGREE groups. 88
FIGURE 31. The mean value of COM for the Human and Dog's face, for both the AGREE and DISAGREE groups. 88
FIGURE 32. The mean value of COR for the Human and Dog's face, for both the AGREE and DISAGREE groups. 89
FIGURE 33. The mean value of LIKE for the Human and Dog's face, for both male and female groups. 89
FIGURE 34. The mean value of ENG for the Human and Dog's face, for both male and female groups. 90
FIGURE 35. The mean value of COM for the Human and Dog's face, for both male and female groups. 90
FIGURE 36. The mean value of COM for Smiley, the Caricature, and the Realistic face. 94
FIGURE 37. The mean value of COR for Smiley, the Caricature, and the Realistic face. 94
FIGURE 38. The mean value of INT for the Honest, Deceiving, and Stoic face, for both the AGREE and DISAGREE groups. 98
FIGURE 39. The mean value of COM for the Honest, Deceiving, and Stoic face, for both the AGREE and DISAGREE group. 98
FIGURE 40. The Realistic Male's facial expressions. 110
FIGURE 41. The Realistic Female's facial expressions. 110
FIGURE 42. The Caricature female's facial expressions. 111
FIGURE 43. The Caricature male's facial expressions. 111
FIGURE 44. The Dog's facial expressions. 112
FIGURE 45. Smiley's facial expressions. 112

CHAPTER 1

Introduction

Motivation for This Research -A Need for Agent-based Interfaces

Software agents

The Internet provides us with enormous amounts of information, and it is involving more people in the world of bits. Computers are becoming essential tools for entertainment, socializing, as well as getting information and news. Improvements in computer technology allow us to have access to information from anywhere at anytime. We need help in dealing with this information overload to separate information from noise. Software agents are computer programs that provide assistance to a user dealing with computer based applications, such as information filtering, meeting scheduling, selecting music, etc. Software agents are different from current programs in three ways: 1) they are **proactive**--they initiate actions, 2) they are **adaptive**--they learn from the user's preference and habits, 3) they are **personalized**--they change their way of helping the user according to what they have learned about

the user [Maes 95a]. Software agents work autonomously without waiting for a user's command. The user and the agent both communicate to each other to complete a task. As we depend more on computers to complete our daily tasks, we will have to rely for some part of our productivity on software agents, or "digital assistants." As Maes [Maes 94] says, we evaluate such digital assistants using two criteria: their ability to perform a task on behalf of the user, and their trustworthiness, or the assurance that allows the user to delegate tasks to an agent without worrying. We will only adopt agents if we can trust them and control their actions.

Agents assist users in various ways and domains. The tasks suited for computer agents are the same as those to which we employ human agents: tasks that require expertise, skill, resources, or labor to accomplish some goal, and tasks that we don't want to do or cannot do ourselves. Laurel and Maes categorize those applications as follows [Laurel 90, Maes 94]:

- **information**: navigation and browsing, information retrieval, sorting and organizing, filtering
- **work**: reminding, programming, scheduling, advising, managing mail
- **learning**: coaching, tutoring, providing help
- **entertainment**: playing against, playing with, performing.

Agent-based Interface

Despite the increasing interest in software agents, very little research on the representation of agents has been done. The dominant user interface is still GUI with direct manipulation, which was introduced in the '70s [Goldberg 84]. GUIs have been refined over the past 20

years to make them more usable [Norman, D. 88], but the original idea of using a direct manipulation metaphor has not been changed. As Lieberman says [Lieberman 96b], a direct manipulation interface provides a user high degree of control and the user can initiate tasks when the interface is simple enough to understand intuitively. As software becomes more powerful, the supporting interface becomes more complex. Consequently, the direct manipulation interface no longer provides high degree of control.

Since software agents change the way we interact with computers, there is a need for building a new interface to support software agent applications. Negroponte suggests a new paradigm of the "social" interface. He writes, "What we today call 'agent-based interface' will emerge as the dominant means by which computers and people talk with one another.... the interface will need size, shape, color, tone of voice, and all the other sensory paraphernalia" [Negroponte 95]. This "agent-based" interface should support completing cooperative tasks based on the user's delegation of tasks to agents. Thus, the representation of software agents will be an important issue for agents' implementation. Maes addresses the important issues that need to be resolved for building a successful software agent [Maes 94]:

- How the agent "does the right thing" to assist users with their tasks.
- How we guarantee that users feel comfortable delegating tasks to the agent.

Most previous research in agents has focused on the former. The latter issue has not been studied much in agent research. Maes [Maes 94] lists several remaining issues in software agents research among which are:

- What is the best metaphor for software agents?

- What characteristics make an agent trustworthy and intelligent?
- Should agents use facial expressions and other means of personification?

With few exceptions (see Chapter 2 for discussion), the bulk of HCI research has not been on personified interfaces, but on direct manipulation interfaces. The most commonly used interface for personification is a human face. However, it is still an open question whether software agents should be anthropomorphized in the interface. Several arguments against the use of agents have been offered. For example, Lanier argues that agents make users lazy and narrow-minded [Lanier 95]. Shneiderman observes anthropomorphic representations and concepts are rejected by users since they want to feel in control and do their own tasks [Shneiderman 95]. One study is not the final word, and in addition, we still don't know what features of the personified interfaces contribute to this rejection.

I believe there is a strong need for a theoretical study of visual representations for agents, which would result in guidelines for designing agent-based interfaces. In particular, this research is necessary before agents soon flood the world. I explore in this thesis the pros and cons of software agents with faces. Then I will discuss the effects of having faces and facial expressions in the interface on the basis of the results obtained by a series of experiments to compare subjects' responses to and evaluation of different faces and facial expressions of entertaining agents.

Contributions of this Research

The important findings of this research are as follows:

- Personified interfaces help users engage in a task, and are well suited for an entertainment domain.
- People's impressions of a face in a task are different from ones of the face in isolation. Perceived intelligence of a face is determined not by the agent's appearance but by its competence.
- The experimental results obtained demonstrate that there is a dichotomy between user groups which have opposite opinions about personification. Thus, agent-based interfaces should be flexible to support the diversity of users' preferences and nature of tasks.

Overview of This Document

Chapter 2 highlights related research on personified interfaces over the last decade, including Apple Knowledge Navigator [Laurel 90], Apple Guides [Oren 90], Newt [Sheth 94], and Maxims [Lashkari 94].

Chapter 3 describes the system design, user interface design, and character design of the poker game used for the experiments. Character design is an important step for creating appropriate representation for the experiments.

Chapter 4 describes the design of five experiments conducted on the World Wide Web. This chapter includes lessons learned from the test experiment, description about the procedure, the conditions, and the questionnaire for the web-based experiments.

Chapter 5 is the heart of this thesis. It describes the hypotheses of each experiment and discusses the results of the pilot experiment and the five web-based experiments--Face vs. NoFace, Male face vs. Female face, Human face vs. Animal face, Simple face vs. Caricature face vs. Realistic face, and three modes of Expressiveness.

Chapter 6 summarizes the findings of this research and discusses the future direction of studies on personified interfaces.

CHAPTER 2

Related Work in Personified Interfaces

This chapter describes related research in personified interfaces and agent-based interfaces. The related research stems from four areas: interface metaphors, social interface, believable agents, and software agents. The second half of this chapter describes psychological research on personification.

Metaphors with Character

Metaphors allow us to take advantage of our knowledge about objects and events in the real world to understand more abstract concepts. When used in an interface, they help us to understand the capability of the system, and provide us with expectations about what will happen when we take certain actions.

Apple Knowledge Navigator

"Phil", shown in Figure 1, a semi-intelligent agent who appeared in the Apple Knowledge Navigator video, could talk to the user. Phil came in two versions: a human facial image and a cartoon face, which may set up different user expectations. The common feature of Phil is the bow tie which is a salient and iconic costume element, making the character recognizable in both representations [Laurel 90].

FIGURE 1. Two kinds of representation of "Phil" in the Apple Knowledge Navigator [Laurel 90]

Apple Guides

Apple Guides is a multi-media educational database that contains American historical information. A primary goal of Guides is to "reduce the cognitive load on users that is created by 'navigating' while trying to learn" [Oren 90]. Different characters with different facial features and costumes, called "Guides", are used to distinguish points of view that can be used to navigate through a historical database (Figure 2). Users assumed or wanted to attribute personality to each iconic Guides figure. This means that human figures induce characterization. Users rated the Guides characters as engaging and responded emotionally not to the computer but to the Guides characters. Oren describes the value of the Guides as follows:

personification helped users reduce the cognitive load to navigate themselves, made the interface more engaging, and suspended disbelief. [Oren 90]

FIGURE 2. The Guides characters [Oren 90]

The above research is based on the hypothesis that anthropomorphism in the interface is effective because we can easily predict how a character is likely to think and behave by its external trait [Don 92]. We know how to communicate with other people, and utilize that knowledge to anthropomorphize animals, objects, and computer software [Nass 94]. Laurel also states in defense of anthropomorphism that the key for applying character metaphors or anthropomorphism to an agent is to provide the appropriate set of traits for the given agent [Laurel 90]. The danger is that people may expect human-like intelligence, understanding, and actions from anthropomorphized characters [Laurel 90].

Research Related to "Social Interface"

The work of Nass *et al.*'s [Nass 94] has proven that people apply social rules to human-computer interaction, thus "computers are social actors". Their laboratory studies show that peo-

ple treat computers as human, even when the interface is not anthropomorphized. People prefer interacting with computers that demonstrate the same personality as they have, be it dominant or submissive [Nass 95]. Their work has been used as a basis for several other studies on personified interfaces. If the interaction with computers is social, using a character in the interface is perhaps a natural way to support the social aspect of the interaction.

Microsoft Persona Project

The Persona project [Bell 96] is one of the efforts by Microsoft to use lifelike characters in an interface. The Persona project is based on Nass *et al.*'s social interface paradigm. Persona uses natural spoken dialogue and expressive anthropomorphic figures (a parrot called "Peedy") for a computer assistant which interacts with the user. The goal is "to achieve a level of conversational competence and visual reactivity that allows a user to suspend disbelief and interact with the assistant in a natural fashion" [Bell 96]. Bell lists several requirements for an assistive interface: 1) it should support interactive give and take, 2) it should have some degree of decision making ability without bothering the user every time it needs confirmation, 3) it should know when to interrupt, 4) it should learn appropriate behavior depending on the user's mood, its task, etc. just as a human assistant does [Bell 96].

Microsoft "Bob"

Another product based on Nass *et al.*'s "social interface" is Microsoft "Bob". Bob is an animated personal guide used as a primary interface to the computer. Bob communicates with the user through speech balloons and provides help and possible options for the user to

choose from. The social metaphor helps the user to concentrate on a single source of information without being overwhelmed by too many options [Bell 96].

Research on Believable and Lifelike Agents

This section discusses pioneering research on believable agents, which focuses on building animated characters with agent technology. Believability is another important issue that makes software agents more usable and trustworthy.

MIT Media Lab's "ALIVE" System

The Media Lab's ALIVE system [Maes 95b] is a pioneering system that allows wireless, full-body interaction between a human and a graphical world inhabited by lifelike autonomous agents. One of the goals of the ALIVE system is to provide an "emotional" experience in a virtual world, by allowing users to interact with lifelike characters using s natural interface. Blumberg's autonomous creatures [Blumberg 94] in the ALIVE system are based on concepts from Ethology. This Ethological approach has proved to be effective in making believable and robust agents.

Bates' "Oz project"

Bates' Oz project [Bates 94] aims to build a believable character by conveying the emotions of the character. Bates believes that believability does not come from "honest or reliable character", but from providing the "illusion of life", and "permitting the audience's suspension of disbelief." In doing this, "emotion is one of the primary means to achieve believabil-

ity." Their work is based on the OCC's emotion model (discussed in Chapter3) and studies of traditional animation techniques [Bates 94].

Elliott's "The Affective Reasoning Project"

Clark Elliott's Affective Reasoner [Elliot 94a] is a real-time simulation platform for emotional agents. These agents have the ability to react emotionally to each other and to the user. The platform has an "emotional engine", speech recognition package, music selection and playback mechanism, and facial representation generator. Elliott states that in some cases cartoon-like faces are more effective to express emotional responses than sophisticated 3D faces [Elliot 95]. To test the believability of emotional characters, three groups of subjects were shown a set of videos of the Affective Reasoner, in which the agents interpret an ambiguous sentence with a line-drawn face with facial expressions, synthesized speech, and music. The results show that 1) subjects are able to detect the intended emotion with over 70% accuracy, and 2) music is better than facial expressions in helping subjects identify the right emotion [Elliott 94a]. As the film industry has known for decades, music can engage users, because it "gives additional cues to the user about the emotional states of the characters," and it is "a powerful arousal, and mood-induction tool" [Elliott 95]. Elliott [Elliott 94b] has suggested that a poker game is a good testbed for enhancing believability of agents through emotional expressions. Some model of emotion or personality is essential to make a realistic computer model of a game such as poker.

Personifying Software Agents

MIT Media Lab's "Newt"

Newt [Sheth 94] is a personalized news filtering system developed at the MIT Media Laboratory. A user can create a set of agents, represented by cartoon-like characters, that assist the user with the filtering of an on-line news source. These characters have different costumes depending on the topic they are interested in, and they change their postures when the user is looking at their specific news topics (Figure 3). Although the main focus of this work was on refining the prediction algorithm, what is interesting to the field of personified interface are findings that users who had knowledge about the system's internal mechanism had no difficulty in correlating the agent's behavior to its internal state and could trust their agent. While those who had no knowledge did not understand the agent's behavior, nor did they trust the agent. Another issue that could be improved was to give immediate feedback to users after they responded to the agents. This result shows that visualizing the agent's internal state and explaining its behavior is essential for making a trustworthy and acceptable agent [Sheth 94].

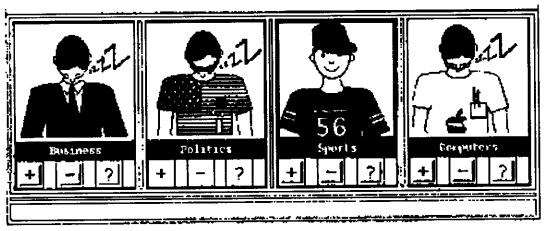

FIGURE 3. The representation used in Newt [Sheth 94]

MIT Media Lab's "Maxims" and the Calendar Agent

Maxims [Lashkari 94] is an agent that assists the user with handling e-mail, and the calendar agent [Kozierok 93] is an agent that schedules meetings on behalf of the user, developed by the MIT Media Laboratory. In both studies, a caricature face with facial expressions is used to convey to the user the degree of certainty when the agent makes a prediction (Figure 4). The agent gives immediate feedback by its caricature face when the user takes an action. This visualization contributes to the user's trust of the agent [Kozierok 93, Maes 94].

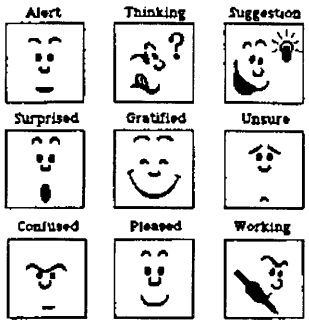

FIGURE 4. The facial representation used in Maxims and the calendar agent [Kozierok 93]

In summary, the above two studies hypothesize that users interact with agents more easily when they are able to predict the agents' characteristics and behaviors from their external traits. A face with facial expressions helps the user understand the agent's behavior and internal states, when used properly. The results are obtained by observation or interviews. None of the studies discussed so far are based on a systematic experiment and quantitative analysis. If the user does not understand the meaning of the visual cues, then personification would confuse the user more. Thus, there is a need to study the following issues:

- Whether faces are the best way to represent the agent's state.
- If yes, what types of faces are appropriate to represent the agent's internal state.
- Whether facial expressions are useful for explaining the agent's behavior.

Research on Users' Impression of Personified Interfaces

The research described so far reports positive effects of personified interfaces. However delightful a personified interface seems, it is important to verify these intuitions. Recently, researchers have begun quantitative analyses of users' experiences with personified interfaces. Interestingly, some of this research indicates that adding faces to an interface does not necessarily provide better human-computer interactions.

Walker *et al.*'s Study of Using a Human Face in a Questionnaire Survey

At Dec, Walker *et al.* [Walker 94] studied how having a face and facial expressions for a computer application affects users' performance and productivity. They compared subjects' responses to an interview survey under three conditions: questions spoken by a synthesized face with a neutral expression, spoken by a face with stern expression, or text-only. Subjects who responded to the spoken face made more effort in answering the question by spending more time, writing more comments, and making fewer mistakes. Those who responded to the stern face spent more time, wrote more comments, and made fewer mistakes than those who dealt with the neutral face. Walker *et al.* reports that having a face is engaging and takes more effort and attention from the user. A face with more expression in an interface leads to greater engagement, but does not always makes the experience for users better if it is added

incautiously. She raises questions about personifying an interface: 1) how does the face in a computer affect the behavior and attitudes of the user interacting with it? 2) how should the faces delivering information be different from those used for soliciting information, and what kind of face should appear? She also states that more work should be done to determine when a personified interface is appropriate [Walker 94].

Takeuchi *et al.*'s "Situated Agent"

At Sony, Takeuchi *et al.* [Takeuchi 95] compared users' impressions of an agent which helped them to win a card game. The agent was represented either as an arrow or a synthesized face. The result shows that users respond differently to systems having a face than to those without. The arrow was recognized as useful and reliable, while the face was rated as fun and entertaining. A face in an interface takes more effort from the user and leads to a lack of concentration on the task because of these reasons: a face in an interface catches more attention, and people try to interpret the meaning of the face. However, Takeuchi argues that it is not necessarily a negative effect in terms of involving the user [Takeuchi 95].

Brennan *et al.*'s Study on Effects of Message Style

Brennan *et al.* [Brennan 94] studied effects of message style with an airline reservation system. There are three message styles: Telegraphic (i.e., "Enter first request."), Fluent (i.e., "Please enter your first request."), and Anthropomorphic (i.e., "How may I help you?"). Users are equally successful in all three groups. Those who interacted with the Anthropomorphic agent treated the computer more like a human than did the others, such as using

indirect phrasing and conversational politeness. There was no difference in attributing intelligence to each agent [Brennan 94].

King *et al.*'s Study on Appraisal of Agents

King *et al.* [King 96] conducted an experiment on people's reactions to different objects (geometric shapes, etc.) and faces, such as Chernoff faces (Figure 5 [Chernoff 73]), caricatures, and realistic 3D human faces. Faces were rated to have more intelligence and agency than any objects. Within faces, subjects attributed minimal agency and intelligence to Chernoff faces, and maximum intelligence to a 3D human face with blinking eyes. King's results show that there are differences in the initial appraisal of visual representations based on whether they are objects or human forms, symbolic or realistic, or static or lively. King's research also points out that users' initial impression of an agent affects how they interact with the agent, and affects the user's model of the agent's ability. Thus a face of an agent has to be carefully chosen to represent the agent's ability properly [King 96].

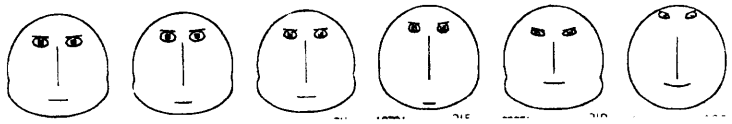

FIGURE 5. Examples of Chernoff faces [Chernoff 73]

Isbister's Study on Perceived Intelligence

Isbister's literature study [Isbister 95] shows that people's judgement of other people's and of TV characters' intelligence is not based on their true intelligence but rather their per-

ceived intelligence. She proposes to apply this tendency to computer interfaces to make agents look intelligent. There are several dimensions which affect people's judgement of another's intelligence: "appearance--judging a book by its cover; language cues; actions/ behavior; and group membership" [Isbister 95]. These dimensions should be taken into careful consideration when designing software agents by using the right visual clues (attractiveness, dress, and other physical features), adding verbal cues, and applying social competence.

Need for User Interface Studies in Actual Applications

The studies mentioned so far contrast with the study in this thesis in several important ways.

The research in the first category (Guides, Persona, Bob, Newt, Maxims) is "implementation-oriented". The focus is on implementation but not on interface and user reactions. In the latter category of research--"psychological/literature studies"--little in terms of implementations was made (except Takeuchi), and the implementations are far from realistic. There is a need to study the effect of personified interfaces in an actual application, not in a controlled laboratory or unrealistic environment.

This thesis is unique in two ways: an application to test agent-based interfaces, namely a poker game, was actually built, and experiments were performed on the World Wide Web so as to have as many and versatile subjects as possible, and the results were analyzed quantitatively and qualitatively. The design of the application is described in the next chapter, and the design of experiments and analysis are described in the subsequent chapters.

CHAPTER 3

Poker Game Design

This chapter describes the character design, interface design, and system design of the poker game, which is an experimental environment used to study the effect of personification of agents. Designing the facial representation of the poker playing agents was one of the most important parts of this study. The design of the experiments are described in the next chapter.

Why a Poker Game?

The way people play poker has two aspects: 1) mathematical probability and 2) psychology. In the first category, players concentrate mainly on their hands and how many cards are dealt to others. In the second category, players play against people they know and have some idea of their opponents' playing styles and use that information to help formulate betting strategies.

In this research, the poker game was chosen to observe to what extent subjects concentrate just on their hands, or pay attention to the faces of opponents and try to interpret their facial expressions. Since subjects do not know anything about the agent's playing style--risk-taker, aggressive, timid, calculating, etc.--at first, they cannot infer anything from looking at their faces or mannerisms which they can do in a real poker game. Do they get into a "psychological" mode after playing poker for a while against the same agents? Do subjects notice the meaning of facial expressions in a game where players should have "poker faces"?[1] The poker game interface was designed to support both modes. If users do not want to see the opponents' faces, they can just concentrate on the math. If they want to observe the faces, they can keep watching them while playing.

The poker game is written in Java to utilize the World Wide Web so that users participate in the experiment exactly the same was as they work in normal conditions. There is a need to conduct experiments on the web because of two reasons. One is that the nature of the experiment needs participants from various backgrounds and age ranges. The other is that current software agents are mostly used on the web.

Designing Characters

The purpose of the experiment is to understand 1) the effect of having a face in an interface, 2) what kind of facial features (gender, humanity, realism) make the agent look intelligent,

1. To examine the appropriateness of using a poker game for an experiment, a test experiment was conducted prior to the main experiment. See Chapter 4 for a detailed description.

likable, and comfortable to work with, and 3) the effect of using facial expressions. Thus, the representations for the agents are the important part of this study.

The poker playing agents used in the experiments have seven different graphical representations: Realistic Human faces, Caricature Human faces, a Caricature Dog's face, a line-drawn Smiley face, and the Invisible Man, which has no face (see Figure 6). Realistic images are taken from photographs of a male and a female face with ten facial expressions. Both faces have dark hair. Caricature Human faces were drawn based on the Realistic faces. The purpose for using the same faces for the Realistic images and the Caricatures is to keep the similar expression of the faces. Table 1 shows the categorization of the faces. Figure 6 shows the graphical representations of the poker playing agents.

TABLE 1. Facial features of the poker playing agents

HUMANITY	REALISM	GENDER
Human	Realistic	Male / Female
	Caricature	Male / Female
	Line-drawn Smiley	-
Non-Human (Dog)	Caricature	-

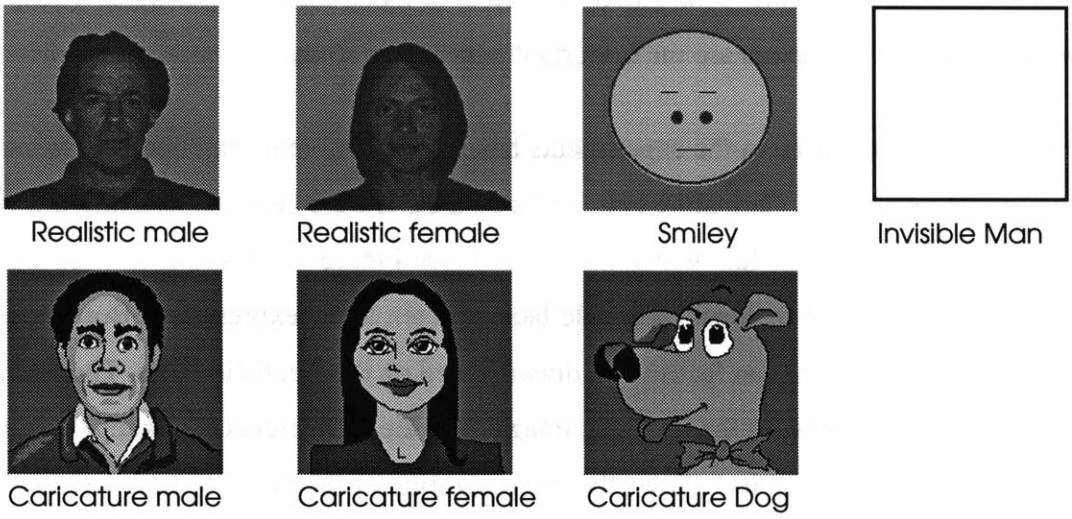

FIGURE 6. The graphical representations of the poker playing agents

Designing Facial Expressions

Each poker playing agent has ten facial expressions. They are neutral, pleased, displeased, satisfied, disappointed, surprised, relieved, excited, anxious, and very excited expressions. Figure 7 shows the ten facial expressions of the female Caricature face. The order of the expressions are the same as stated above. The facial expressions for each character were pre-tested to validate that they express the right emotions. Ten subjects were asked to make pairs of emotion types with the ten facial expressions. Ninety five percent of the answers were correct. Appendix 1 shows all character's facial expressions.

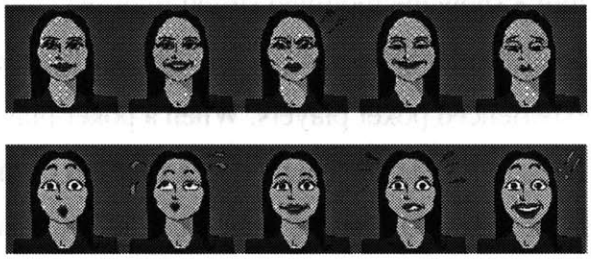

FIGURE 7. Ten facial expressions of the female Caricature face

Selecting Emotion Types -the OCC model-

The ten expressions are derived from the Ortony, Clore, and Collins (OCC) model of emotions [Ortony 88]. There are many approaches in defining and modeling emotions in psychology. Among them, the OCC model is designed as the basis for a computational system reasoning about emotions. The OCC model assumes that emotions are the result of valanced--positive or negative--reactions to situations which a person experiences. For example, when we compare the winning lottery number and the lottery we have, this event leads to a pleased emotion when we win, or a displeased emotion when we lose. In this model, the world consists of events, agents, and objects. Valanced reactions to combinations of these result in emotions. The model leads to different emotional states depending or whether the reaction is positive or negative (see Figure 8). A more complicated emotion structure exists for reaction to events and agents, and depending on whether the situation affects the self or someone else.

In the poker game situation, I assume that the world consists of events. To keep the model simple, reactions to other agents are not considered in the poker game. Thick arrows and gothic characters in Figure 8 show the modified structure of the emotion types used in designing the possible emotions for the poker playing agents. The modifications were made after interviewing two experienced poker players. When a poker playing agent sees its hand, that is an *event*, which leads to a valanced reaction of a **pleased** or **displeased** emotional state (see A in Figure 8 and Figure 9). Betting is an *event*, which focuses on the consequences for *self-prospects*. The betting event derives a valanced reaction of **hope** or **fear**, which is represented by an **excited** or **anxious** expression (see B in Figure 8 and Figure 9). The emotion intensifies as another betting round goes on, especially when the agent is bluffing. At the final stage, the agent shows a **satisfied** expression when it wins (*hope confirmed*), **disappointed** expression when it loses (*hope disconfirmed*), **relieved** expression when it folds (*fear disconfirmed*), or a **surprised** expression when the result is not what was expected (*unexpected*) (see C in Figure 8 and Figure 9). The OCC model does not include "surprise" in their emotion types. However, "surprise" is a common emotion in a real poker game environment.

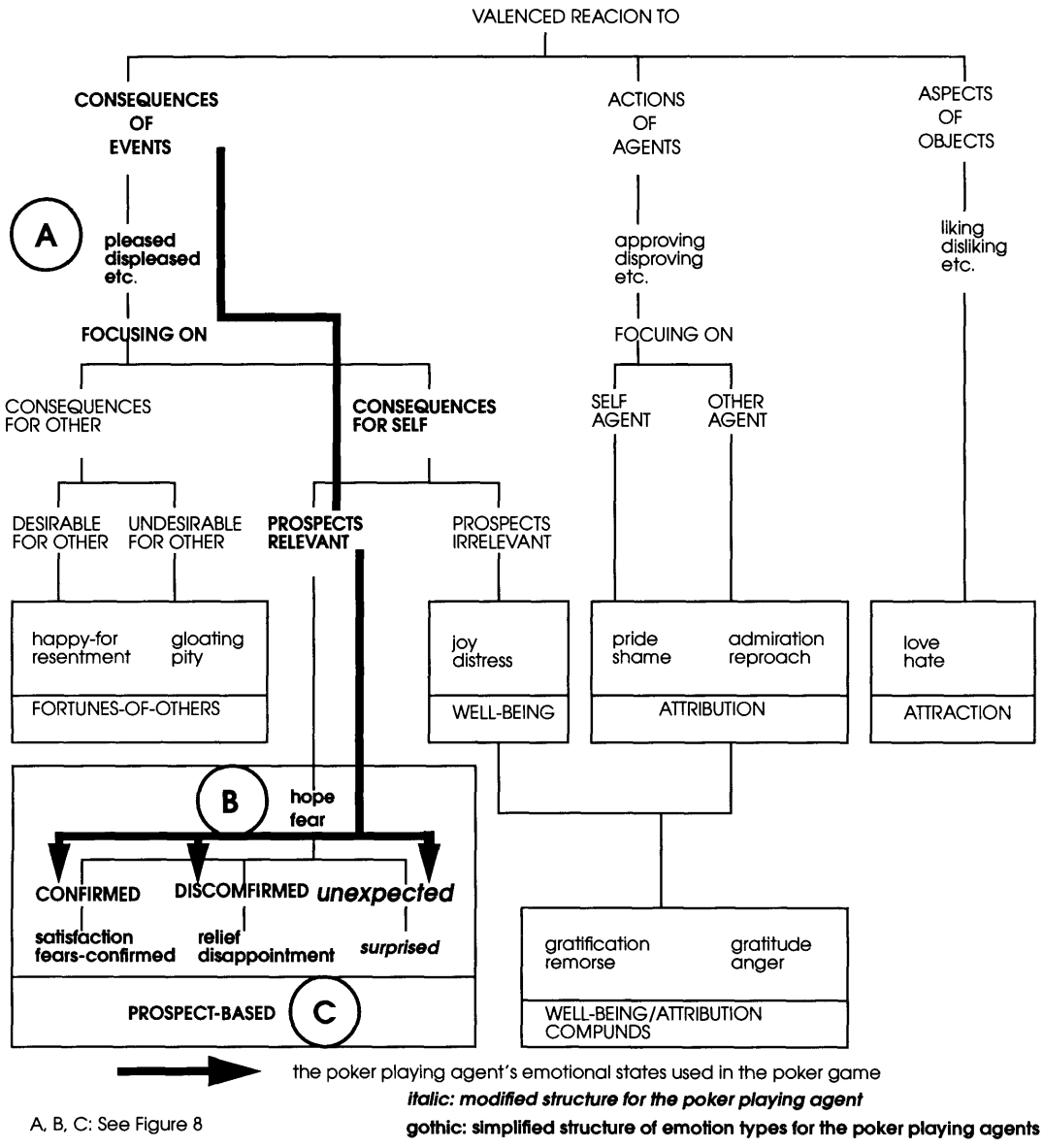

FIGURE 8. Global structure of emotion types in the OCC model adopted from [Ortony 88]

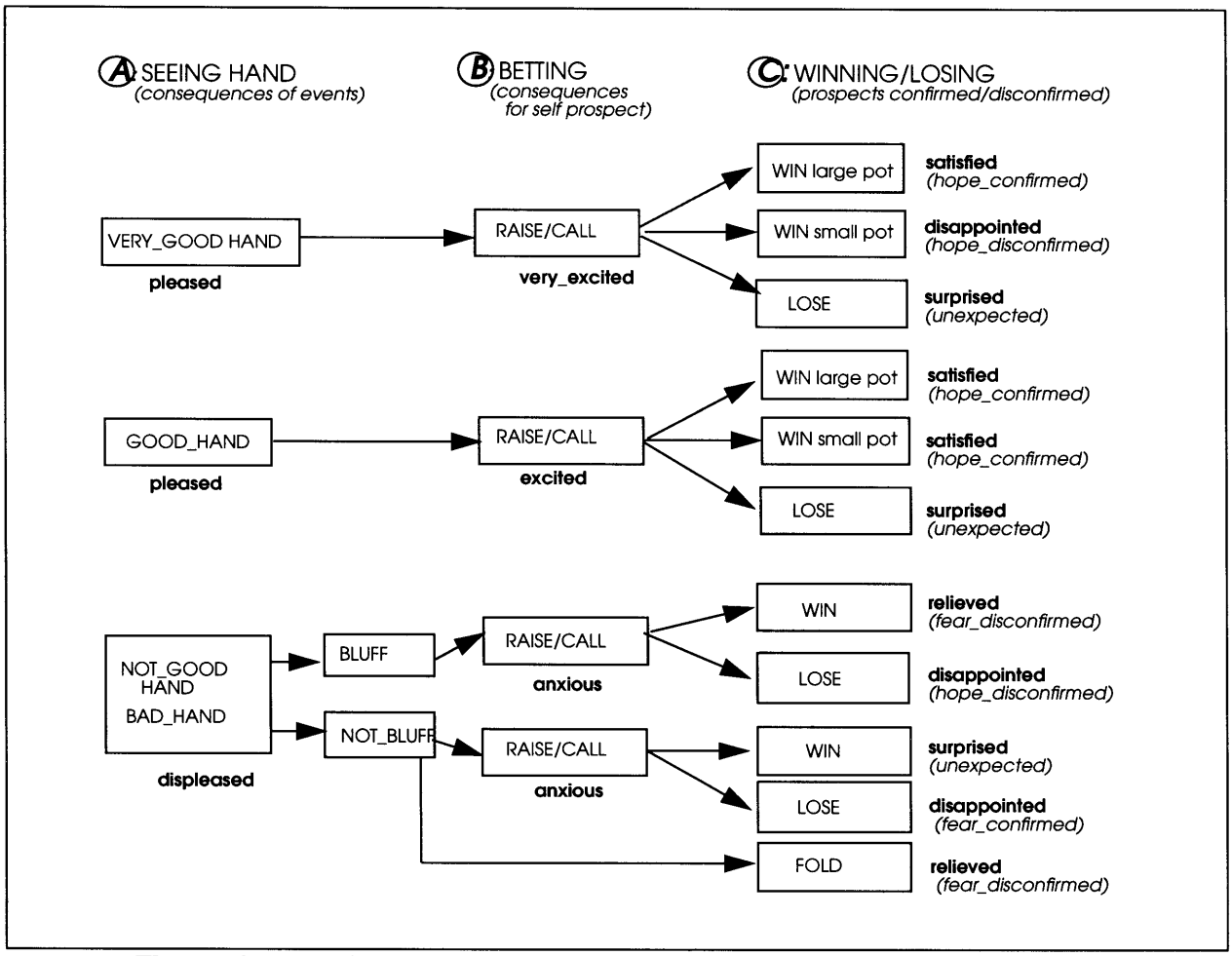

FIGURE 9. The emotion types for the poker game

Applying the Emotion Model to the Poker Game Environment

Emotional states of a poker playing agent are determined based on the situation it is in, as illustrated in Figure 9. After dealing or discarding and when the agent sees its hand (see the next section for the categorization of a hand), it has one of the following facial expressions:

- **Pleased** when its hand is "VERY_GOOD," or "GOOD".
- **Displeased** when its hand is "NOT_GOOD" or "BAD".

While betting, the agent has one of the following expressions and intensifies the emotional level as another betting round goes on:

- **Very_Excited** when its hand is "VERY_GOOD" and it raises or calls.
- **Excited** when its hand is "GOOD" and it raises or calls.
- **Anxious** when its hand is "NOT_GOOD" or "BAD" and it raises or calls.

During the winning/losing stage, the player has one of the following expressions:

- **Satisfied** when its hand is "VERY_GOOD" or "GOOD" and wins large pot (*hope_confirmed*).
- **Satisfied** when its hand is "GOOD" and wins small pot (*hope_confirmed*).
- **Disappointed** when its hand is "VERY_GOOD" but nobody makes a big bet or folds (*hope_disconfirmed*).
- **Disappointed** when it is bluffing and it loses (*hope_disconfirmed*).
- **Disappointed** when it is not bluffing and it loses (*hope_disconfirmed*).

- **Surprised** when its hand is "VERY_GOOD" or "GOOD" but it loses (unexpected).
- **Surprised** when its hand is "NOT_GOOD" or "BAD" but it wins when it is not bluffing (unexpected).
- **Relieved** when its hand is "NOT_GOOD" or "BAD", is bluffing, and wins (*fear_disconfirmed*).
- **Relieved** when its hand is "BAD" and it folds (*fear_disconfirmed*).

The later section (System Design, Poker Playing Strategy) describes the strategy in more detail.

Different Modes of Expressiveness

The poker playing agents have three different modes of expressiveness.

- The first one is the honest mode, in which a player is honest about its hand in its facial expressions.
- The second is the deceiving mode, in which a player is expressive, but its facial expressions do not correspond to its actual hand. A poker playing agent in deceiving mode shows one of the set of facial expressions in each stage described above. For example, the agent shows either pleased or displeased expression when dealing, regardless of its hand.
- The last mode is the stoic mode, in which a player has stoic expressions, a poker face, so to speak. The poker playing agents show only **neutral** and **pleased** expressions when they see their hands and bet in stoic mode.

Designing the Interface

The web interface of the poker game has a poker table surrounded by up to five poker players including the user (see Figure 10). The user is sitting in front of the table, and his/her hand and the amount of money is shown under the table. There is a message area in the center of the table. This area is used to instruct the user (i.e., "Please discard."), to tell other player's actions (i.e., "Bruce raises $10."), and to tell the result (i.e., "Pattie won!"). The current pot is shown under the message area.

The poker game situation changes every second. In other words, users see their opponents make an action every second[1]. A new status message appears every second, and the agent who took action at that time changes its facial expression. The game stops when it waits for the user's action, such as call, raise, or discard. There are interaction buttons at the bottom of the screen to let users play the poker game. They are "open", "fold", "discard", "pass", "raise", "call", continue", and "questionnaire" (see APPENDIX 2 for instructions for playing the poker game and Chapter 4 for the questionnaires).

There were several user interface modifications after hearing test users' opinions. One user worried about his hand "being peeked on" by other poker playing agents. The poker table was redesigned to make it look like 3D by tilting it and giving it some depth to solve this problem. Several users reported that it was easy to get lost in the game because they did not know when to play and where to start. To solve this problem, two modifications were made.

1. During the test experiment, different playing speeds were tested. Some subjects preferred slower speed to observe what other players were doing, while other subjects preferred playing faster and did not want to wait for their turn.One second was the appropriate speed in general. See Chapter 4 for the description of the test experiment.

The first modification was to change the color of messages to a user to blue, while status messages are shown in black. The second modification was to disable/enable buttons in each stage. For example, the "discard" button is clickable only when it is the user's turn to discard. The third suggestion was to move status messages near each players, so that users do not have to move their eyes from the message area to each of the faces. However, the message position should be kept in the center, since the purpose of the experiment is to let the user choose which information they try to pay attention to, the messages and/or the faces.

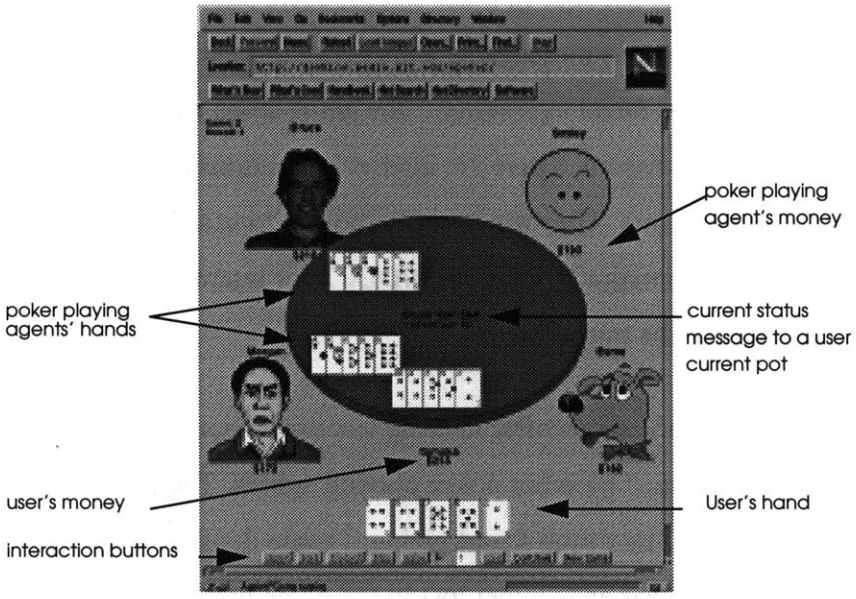

FIGURE 10. Web interface of the simulated poker table

System Design

Overall Architecture

Each user downloads the same poker java applet, which runs in the client's site independently from the poker server. The java applet contains the poker game and the facial images. The applet randomly assigns one of five experiments to users in an randomized order. A java application in the server continuously logs the players' hands, actions, and the amount of money that the player has in a disk while users are playing the poker game. After finishing one condition, the applet sends the experiment name and the user's name to the server, and then calls a CGI script which dynamically generates a corresponding questionnaire in HTML format. Another CGI script collects users' answers to a questionnaire in a server disk for future analysis. Another CGI script checks log files and generates a high-score list. Figure 11 shows the poker game system structure.

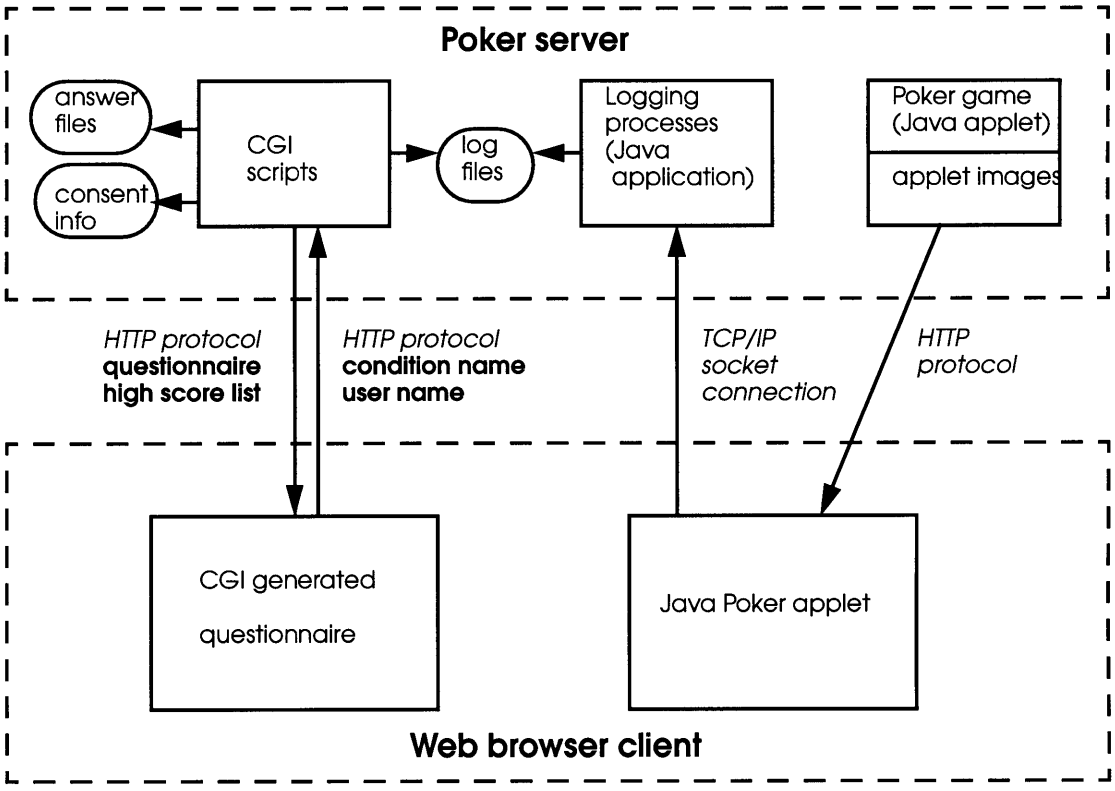

FIGURE 11. Poker game system structure

Poker Playing Strategy

The purpose of building the poker game was not to make the best or most realistic game available, but more importantly to make one with expressive poker playing agents. Thus, the implementation of the poker playing strategy is minimal. The player's hand is categorized as follows:

- "VERY_GOOD": Straight flush, 4 of a kind, Full house, Flush, Straight, 3 of a kind
- "GOOD": Two-pair, High-pair
- "NOT_GOOD": Low-pair
- "BAD": Nothing

The betting strategy is as follows:

There is a total raising limit (the pot should be under $150) in the poker game. When a poker playing agent has a "VERY_GOOD" hand, it raises up to a maximum raising limit ($30) per betting round when its hand is not new, or raises up to $20 when its hand is new. "Up to" means that the actual bet is determined by a random number within the limit, in order not to let users notice the playing style too easily. When it has a "GOOD" hand, the poker playing agent raises up to $20 when its hand is not new, calls when the money it has to pay to call ("dif") is more than $20, or raises when "dif" is less than $20. When the agent has a "NOT_GOOD" hand or a "BAD" hand, it "decides" whether to bluff or not by a randomized procedure. It raises up to the maximum limit when it is bluffing, calls when "dif" is less than $5 or $10 and when not bluffing, or folds when "dif" is bigger than $5 or $10. Figure 12 shows the poker playing strategy described above. All the poker playing agents have the same strategy.

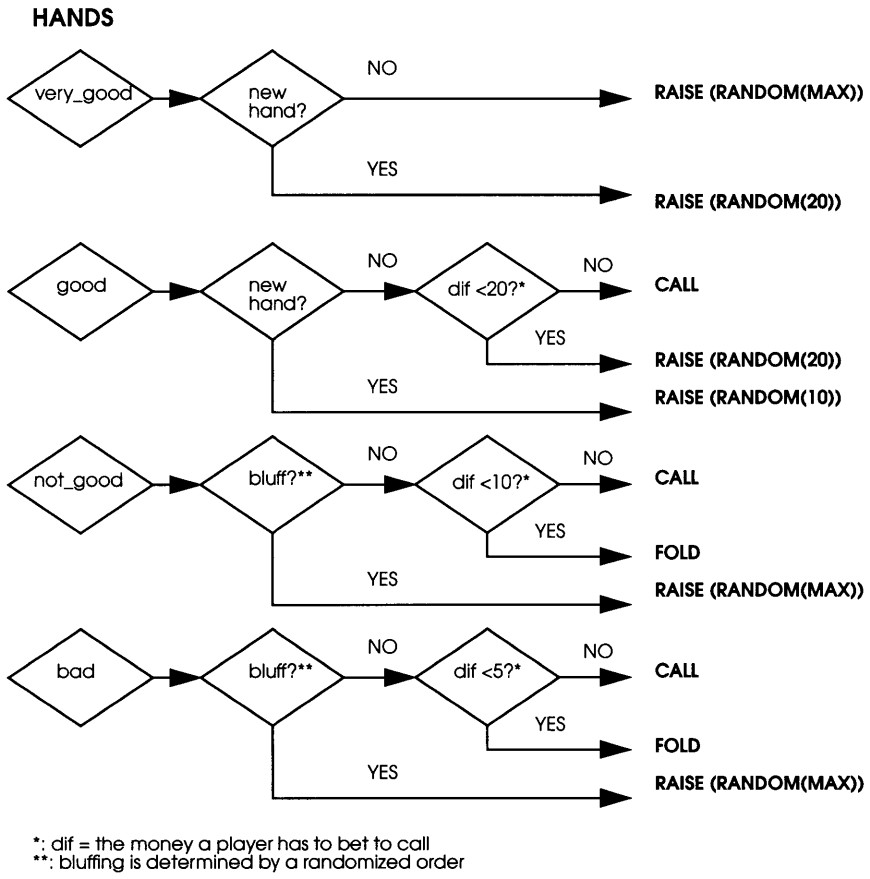

FIGURE 12. Betting strategy used in the poker game

CHAPTER 4

Design of Experiments

This chapter describes trial and error for designing experiments, then describes the purpose, procedure, and conditions of the five web-based experiments performed. Hypotheses are described in Chapter 5.

Lessons Learned from the Test Experiment

Before conducting the web-based experiments, a test experiment was conducted within the Media Lab to refine the design of the experiment. There were several important lessons about how to design an experiment to get meaningful and significant results.

The first lesson concerned controlling the conditions. The test experiment was designed to test all conditions (the effect of having a face, the difference between gender, realism, humanity, and expressiveness) with only one experiment. The poker game had four poker

playing agents which were randomly chosen out of 18 combinations.[1] The results were hard to analyze, since there were too many factors which might affect the users' impressions of the faces. Thus, the web-based experiments were designed to have only one factor per experiment. For example, in order to study the difference between gender, two faces which differ in terms of gender only (Caricature Male and Caricature Female, or Realistic Male and Realistic Female) were used.

Secondly, I found that there was a need for some initial training to let subjects familiarize themselves to the facial expressions of the poker playing agents. During the test experiment, subjects reported that they could not understand the meaning of each facial expression. Thus, the web-based experiments started with familiarization.

The third lesson concerned designing questionnaires. Yes/no questions should not be used in a questionnaire survey in order to have quantitative measurements. Instead of asking yes/no questions, fixed scale questions should be used to maximize the quantitative data. The scale should have odd numbers to permit the neutral opinion. As for the manner of asking questions, questions that are too direct should be avoided. For example, instead of asking, "How intelligent do you think player A is?", "Do you think the player is good at playing poker?" should be used.

The test experiment was useful to make sure whether the poker game was appropriate for the experiment. Since the main experiment was conducted on the World Wide Web, there was a

1. There were six facial representations (Realistic Male/ Realistic Female/Caricature male/Caricature Female/Smiley/ Dog) and three modes of expressiveness (Honest/Deceiving/Stoic). This means that there are 18 possible combinations of representation for the poker playing agents.

need to evaluate whether people consider the poker game as a serious experiment and to make sure subjects respond to the questionnaire. Some subjects were video taped and observed during the test experiment. They took the game as a serious experiment though they enjoyed playing the game. The response rate for the questionnaire was more than 30%. The test subjects made many comments and suggestions on the interface design, which was used to modify the interface before starting the web-based experiments (see Chapter 3 Designing the Interface). The test experiment was also useful to determine the appropriate number of rounds for playing poker and speed of playing in turn.

Design of Web-based Experiments

Purpose

The first purpose of the experiments is to understand the effect of using faces and facial expressions in the interface of the poker game environment. The second purpose is to analyze whether people's impression of an agent is determined by its representation or by its performance or both. All agents use exactly the same algorithm for playing poker, which the subjects were not aware of.

The questions tested in the experiments are as follows:

- Do people pay attention to a face or facial expressions of an agent?
- Are people distracted by the face or facial expressions?
- Does having a face increase people's engagement to a task?

- Do people use information from the face to interpret the agent's behavior?
- What kind of facial features make the agent look intelligent, likable, and comfortable to interact with?
- Does people's impression of an agent is determined based on its representation or on its performance or both?
- Does people's opinion about personification affect their impression of the faces?
- Is there any difference between the subject's gender in their impression of faces?

The purpose of conducting experiments on the World Wide Web is to have as many subjects with various backgrounds as possible in a condition where software agents are mostly used.

Subjects

Intended subjects are people who accessed the poker game site on the World Wide Web. Anyone who knows the URL (http://poker.www.media.mit.edu:81/epoker) can play the poker game. The poker game site was advertised to two poker related news groups, added to the Gamelan Java directory, Yahoo, the Media Lab web page, and the MIT fun stuff web page.

Procedure

A typical scenario of an experiment is as follows:

- A subject accesses the poker web site.

- The subject reads the consent form and instructions. The instructions contain the poker playing rules, how to play, and the procedure (see Appendix 2 for more details).
- The subject agrees to participate in the experiment and submits an electronic agreement form.
- The subject sees the poker playing agents faces to familiarize himself/herself with the agents' faces and facial expressions.
- The subject's name, email address, and login time are sent to the server and stored in a disk.
- The poker java applet is downloaded to the subject's computer.
- The subject plays a poker game with one out of five conditions for 15 rounds, which takes from 10 to 15 minutes to complete.
- The subject's and agents' hands, actions, and the ending moneys are logged into the server.
- A questionnaire appears after playing 15 rounds[1].
- The subject answers and submits the questionnaire.
- The subject's answers are stored in the server.
- The subject can continue another poker game with one out of five conditions or quit the game.

1. Prior to the web-based experiment, a test experiment was conducted to observe how long it takes for subjects to master how to play the poker game. According to my observation and subjects' comments, it takes about 5 rounds to become familiar with the poker game. I assume playing for 15 rounds is long enough to observe the opponents' strategies, but not long enough to get bored of playing the game.

The poker game was designed so that subjects can continue playing the poker game with another condition only after submitting their answers. This is to encourage subjects to answer the questionnaire[1]. Conditions are randomly assigned to each subject. Appendix 2 shows the instructions presented to the subjects.

Conditions for the Web-based Experiments

The poker game contains five experiments. Table 2 shows the facial representations used in each experiment. There are two or three poker playing agents in each experiment. All facial images used in the experiments have the same size, same resolution, and the same background color. Except for experiment 5, all poker playing agents are in Honest mode.

Experiment 1 (FACE experiment) is designed to study the effect of having a face or not. There are two poker playing agents in this experiment. Their representations are the Caricature Female face which has a white square background, and a white square with the same size which is called "Invisible Man". Figure 13 shows the poker game environment of experiment 1. Experiment 2 (GENDER experiment) studies the difference in subjects' impressions of characters' gender. Once again there are two poker playing agents. The representations used in this experiment are the Realistic Male and the Realistic Female face. Figure 14 shows the poker game environment for experiment 2. Experiment 3 (HUMANITY experiment) is designed to compare the difference in subjects' impressions between a human face and a non-human face. The representations used are the Caricature Male face

1. Some subjects experienced the same condition and answered the same questionnaire more than once. Only the first answer of those was analyzed in order to design within-subjects (or repeated-measures) experiments. In a within-subject design, a given subject participates in all experimental conditions only once.

and the Dog's face. Figure 15 shows the poker game environment for experiment 3. Experiment 4 (REALISM experiment) uses three faces to compare the effects of three different levels of realism. The representations used are the Realistic Male face, the Caricature Male face, and the Smiley face. Figure 16 shows the poker game environment for experiment 4. Experiment 5 (EXPRESSIVENESS experiment) has three players with the same facial representation. This experiment was designed to evaluate the difference in subjects' impression between three modes of expressiveness. The Caricature Female face with different color of clothes is used to distinguish each face. One agent has Honest expressions, the other has Deceiving expressions, and the last one has Stoic expressions. Figure 17 shows the poker game environment for experiment 5.

TABLE 2. The facial representations used in each experiment

The expressive mode is Honest unless noted.

	Player 1	Player 2	Player 3
FACE	Caricature Female	White square (Invisible man)	-
GENDER	Realistic Male	Realistic Female	-
HUMANITY	Caricature Male	Caricature Dog	-
REALISM	Realistic Male	Caricature Male	Smiley
EXPRESSIVENESS	Caricature Female with Honest expressions	Caricature Female with Deceiving expressions	Caricature Female with Stoic expressions

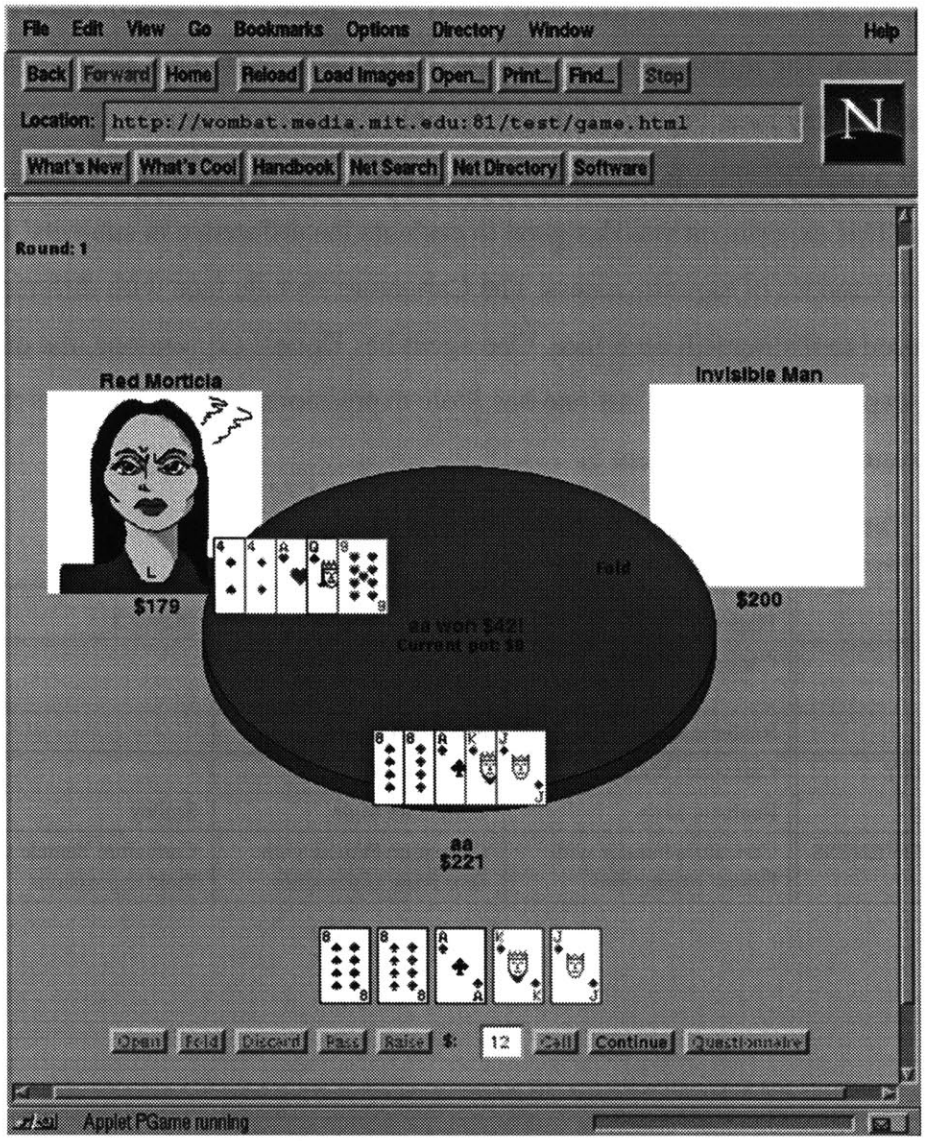

FIGURE 13. The poker game environment for the FACE experiment.

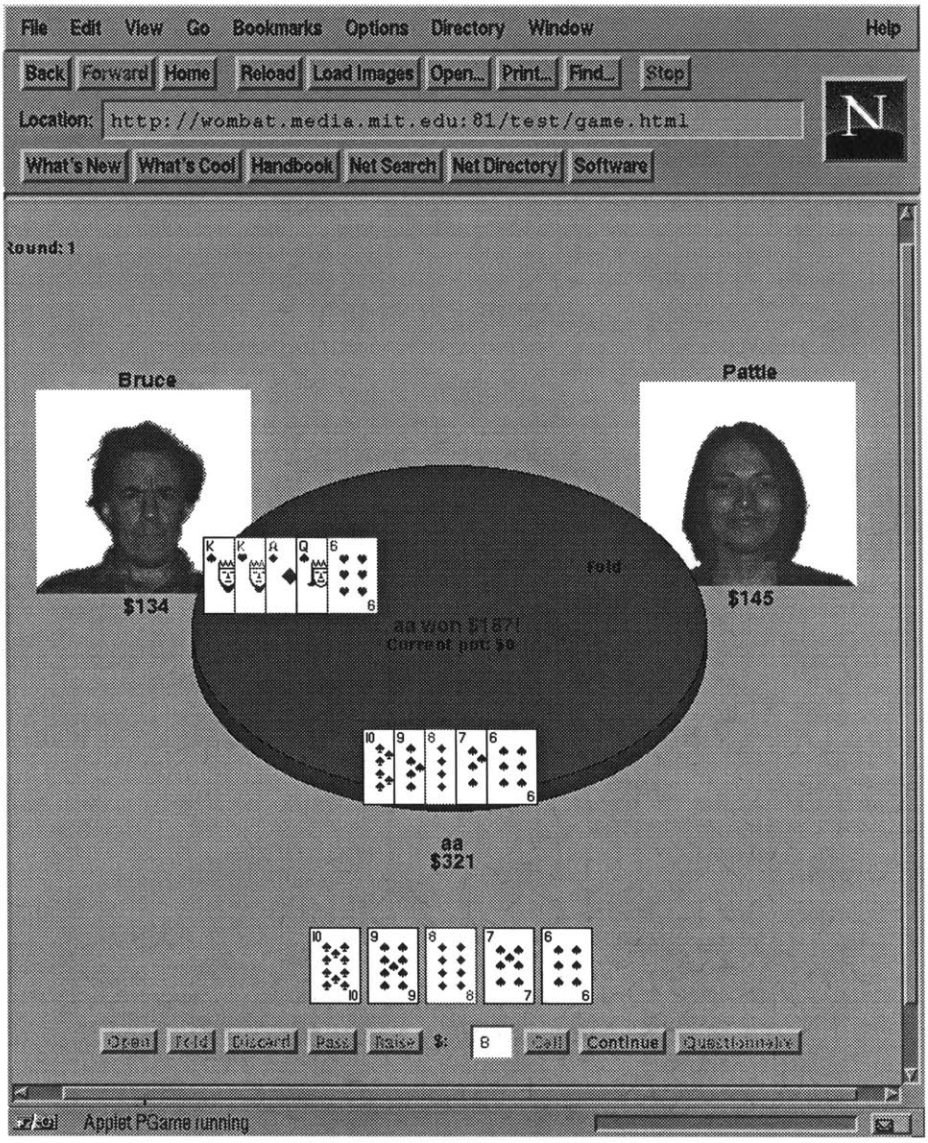

FIGURE 14. The poker game environment for the GENDER experiment

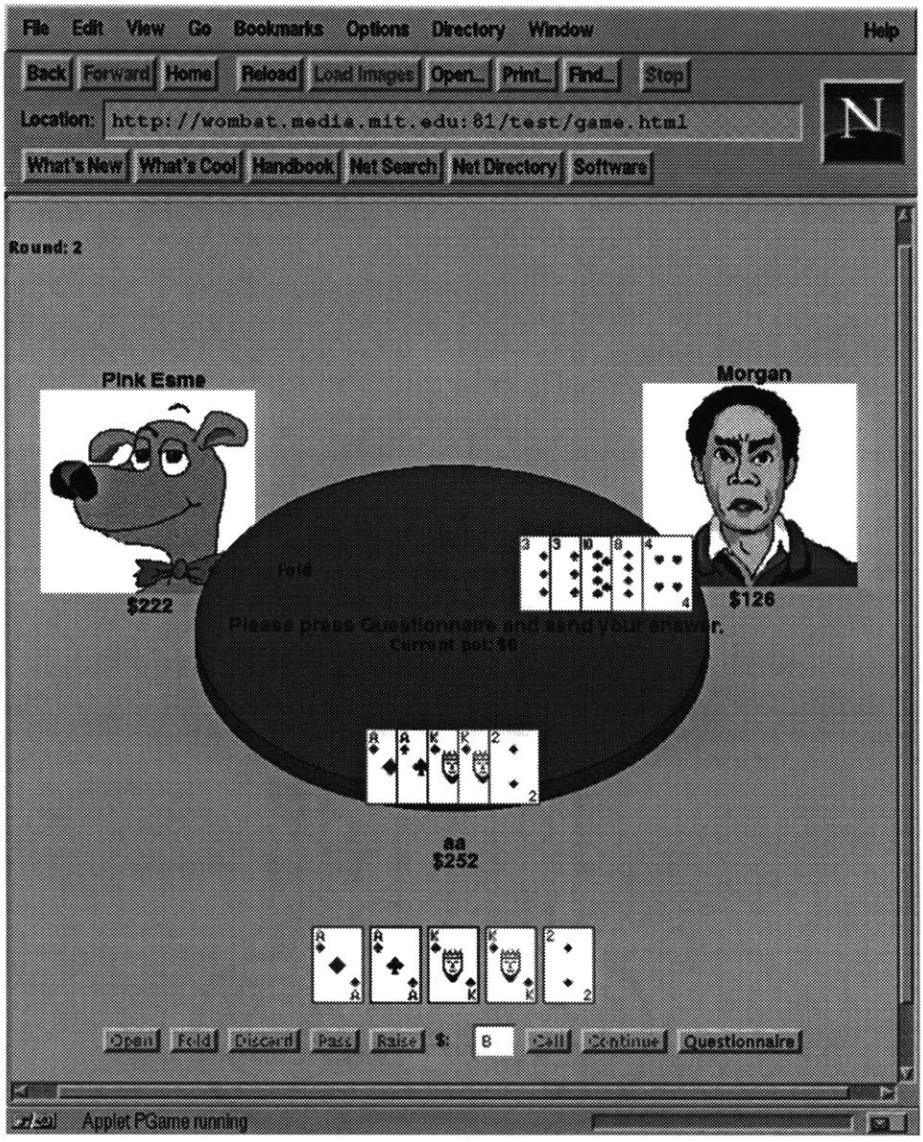

FIGURE 15. The poker game environment for the HUMANITY experiment

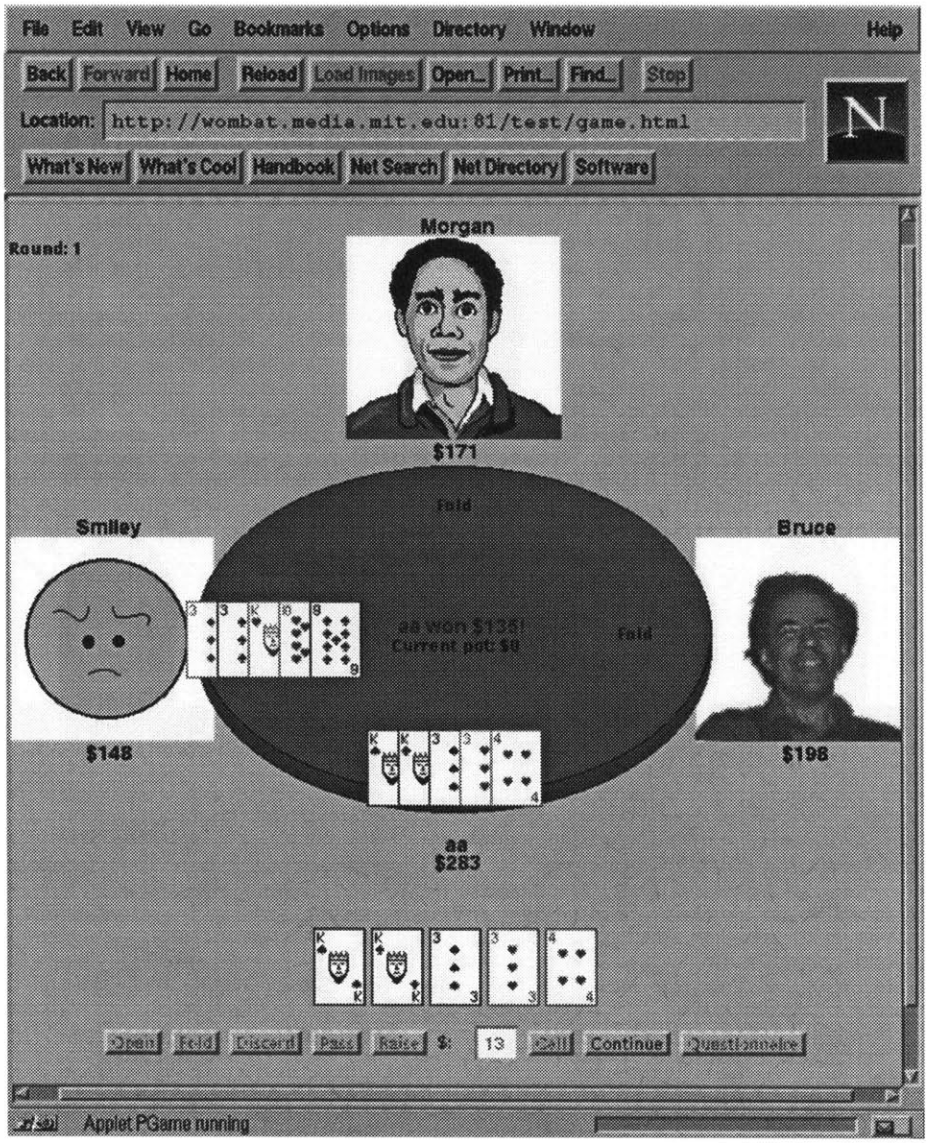

FIGURE 16. The poker game environment for the REALISM experiment

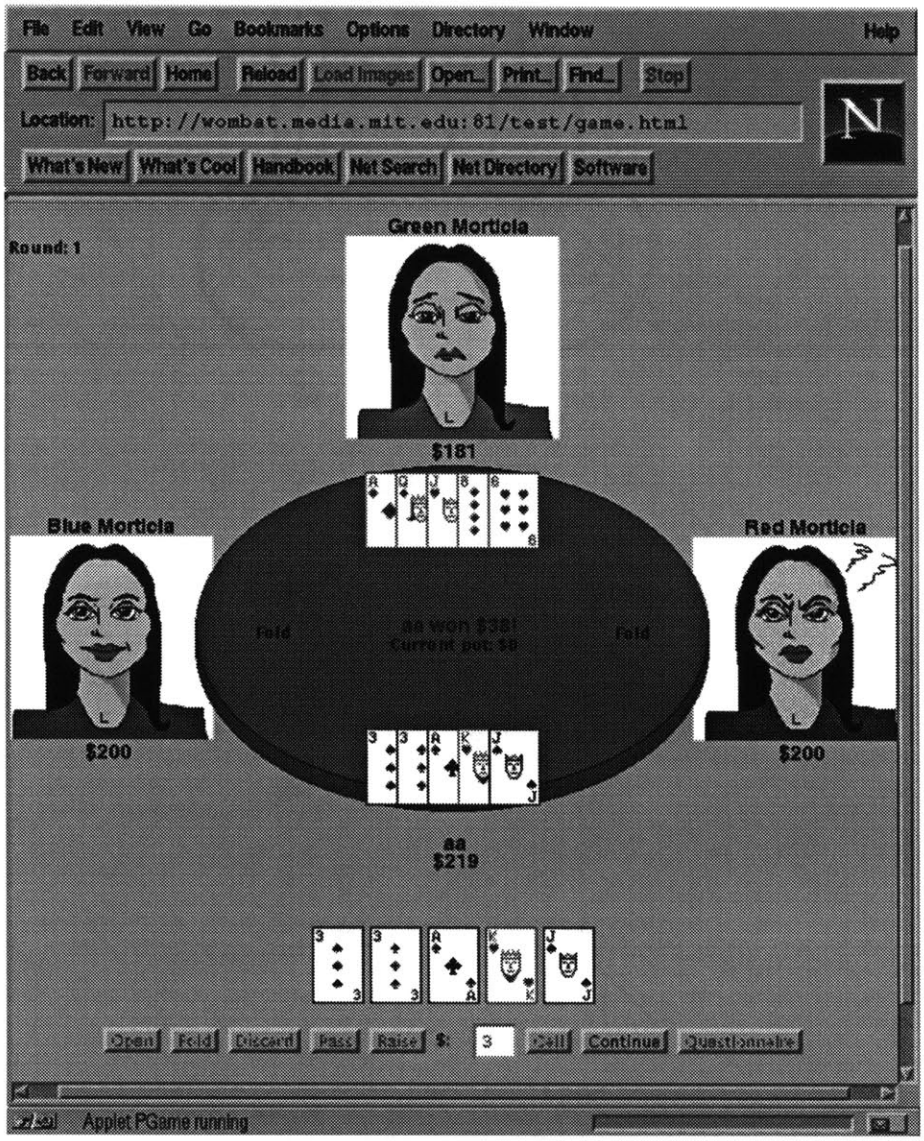

FIGURE 17. The poker game environment for the EXPRESSIVENESS experiment.

Questionnaire

A questionnaire contains 8 to 12 questions. Questions vary according to the experiment. There are 4 background questions (subjects's gender, age, computer expertise, and their opinion about personification), and 4 to 8 questions about the subjects' impressions of the faces against which they played the poker game. The questions about their impressions are seven point scale (1: strongly disagree to 7: strongly agree). Appendix 3 shows all questions presented to subjects.

The common questions across the experiments are perceived intelligence, likability, engagingness, and level of comfort for each face. Intelligence is measured by how good at playing poker each player is perceived to be. Likability is measured by how much the user enjoyed player poker against each player. Engagingness is measured by how engaging playing poker against each player was. Comfortableness is measured by how much the user wants to play poker with the same player.

Specific questions are attention required and level of distraction (for FACE and EXPRESSIVENESS experiments), level of correspondence to the actual intelligence of the player (for HUMANITY and REALISM experiments), believability of the facial expressions, and the usefulness of the facial expressions (for EXPRESSIVENSS experiment). To measure level of distraction, subjects are asked to rate how much they were distracted by the existence of the face or facial expressions with 7 point scale. Level of correspondence is measured by how suitable each face was in representing the actual poker playing skills of each agent. Believability is measured by how much the user believed each player was honest in its facial expressions about its poker hands. Usefulness is measured by how helpful the

facial expressions of each player were to understand the agent's strategy. Table 3 shows the experiment name and the questions asked in each experiment.

TABLE 3. The questions asked in each experiment

experiment name	FACE	GENDER	HUMANITY	REALISM	EXPRESSIVENESS
conditions dependent measurements	Face/ NoFace	Male/ Female	Human/Dog	Smiley/ Caricature/ Realistic	Honest/Deceiving/ Stoic
Intelligence	Y	Y	Y	Y	Y
Likability	Y	Y	Y	Y	Y
Engagingness	Y	Y	Y	Y	Y
Comfortableness	Y	Y	Y	Y	Y
Required attention	Y				Y
Distraction	Y				Y
Correspondence		Y	Y	Y	
Believability					Y
Usefulness					Y

Pilot Experiment

A pilot experiment was conducted before starting a web-based experiment. The purpose of the pilot experiment was to collect subjects' impressions of each face based solely upon its visual appearance. Seventeen MIT undergraduate and graduate students participated in the pilot experiment. Subjects were shown one of the sets of faces used in the GENDER, HUMANITY, and REALISM experiments with a neutral expression. The name of the experiments are GENDER_LOOK, HUMANITY_LOOK, and REALISM_LOOK, respec-

tively. The subjects answered questions about each face's intelligence, likability, and engagingness as a opponent poker player, just by looking at the facial images. The questions asked in the pilot experiments are shown in Appendix 3.

CHAPTER 5

Results and Discussion

This chapter discusses the results of the pilot experiment and the five web-based experiments. All statistical analyses were done with SPSS[1]. Analyses consist of examining the main effect of having different faces, interactions between facial conditions (FACE, GENDER, HUMANITY, REALISM, EXPRESSIVENESS) and subjects' gender (SS_GEN) and subjects' opinion about personification (VOTE). To verify that all poker playing agents had the same level of poker playing skill, the amount of money that each player had at the end of each experiment was also analyzed.

Experiments continued for one and a half months, from early June to mid July, 1996. Subjects participated in the experiments from all over the world using the World Wide Web. More than 1,000 people accessed the poker game site and 157 of them answered the questionnaire, for a response rate of 15%. Of these subjects, 78% were male. The age range of

1.SPSS is a software package for microcomputer data management and analysis developed by SPSS inc.

the subjects was from 10 to 50 years old. Fifty-seven percent of them were in their 20's, 26% were teens, and 14% were in their 30's. Fifty-two percent of them were advanced computer users, 40% were intermediate users. When asked about personifying an interface, 51% of them supported having a face on the screen ("AGREE" group), the rest were against having a face ("DISAGREE" group).

Impressions Based on Visual Appearance (Pilot Experiment)

This section describes the results of the pilot experiments. There are three experiments: GENDER_LOOK, HUMANITY_LOOK, and REALISM_LOOK. The results show that peoples' impressions of faces differ based on appearance.

Results of GENDER_LOOK

Table 4 shows the results of the GENDER_LOOK pilot experiment. According to the result of paired-t test[1], there is no main effect of GENDER in the measurements. There are no 2-way interactions between GENDER (the agent's gender), SS_GEN (the subject's gender), and VOTE (the user's opinion about personification). This result indicates that there is no difference between people's impression of the Male face and Female face in terms of its intelligence (INT), likability (LIKE), and engagingness (ENG).

1.The t-test is used for measured variables, in comparing two means. The paired t-test compares two paired observations on the same individual or on matched individuals [Norman, S. 86]. In this experiment, the two paired observations are the ratings of the Male and Female face for INT, LIKE, and ENG. The two paired observations are (INT1. INT2), (LIKE1, LIKE2), (ENG1, ENG2), where 1 means the rating for the Male face, and 2 means the rating for the Female face.

TABLE 4. The mean value of variables for GENDER condition (Male vs. Female) based on appearance.

Variables	Mean for Male (n=15)	Mean for Female (n=15)	t (paired)[a]	significance
INT (Intelligence)	4.73	5.13	t(14) = -.88	n.s.
LIKE (Likability)	4.80	4.73	t(14) =.19	n.s.
ENG (Engagingness)	5.13	4.80	t(14) =.77	n.s.

a. T (degree of freedom) = mean difference / standard error of differences.

Cells contain the mean rating on a 1-7 scale where 1 was the negative extreme and 7 was the positive extreme.

Results of HUMANITY_LOOK

The result of the HUMANITY_LOOK pilot experiment shows a significant main effect of HUMANITY, as shown in Table 5. The Human face is rated to be more intelligent than the Dog's face ($t(15)=3.10$, $p<.01$[1]), but less likable ($t(15)=-3.13$, $p<.01$) and less engaging ($t=-5.57$, $p<.01$) as an opponent for a poker game (see Figure 18, Figure 26, and Figure 20, respectively). There is no 2 way or 3 way interaction between HUMANITY, SS_GEN, and VOTE. This result shows that the Human face is perceived as more intelligent than the Dog's face based on their visual appearances, but less likable and engaging as a representation for a poker player.

1.P is the probability that the difference we do observe could be due to chance variation alone. When $p<.05$, we reject the null hypothesis and conclude that there is some difference between the groups [Norman, S. 86].

TABLE 5. The mean value of variables for HUMANITY condition (Human vs. Dog) based on appearance.

Variables	Mean for Human (n=16)	Mean for Dog (n=16)	t (paired)	significance
INT (Intelligence)	3.88	2.63	t(15) = 3.10	p <.01
LIKE (Likability)	3.81	5.37	t(15) = -3.13	p <.01
ENG (Engagingness)	3.50	5.56	t(15) = -5.57	p <.01

Cells contain the mean rating on a 1-7 scale where 1 was the negative extreme and 7 was the positive extreme.

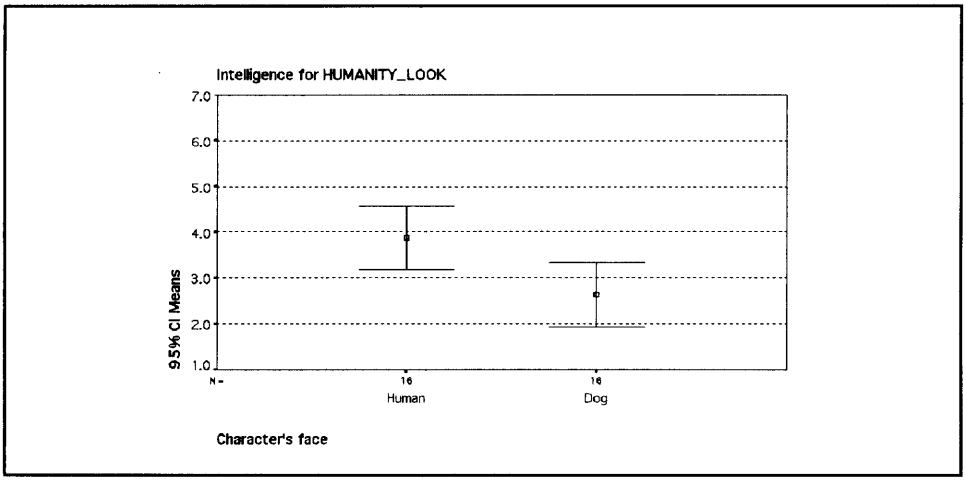

FIGURE 18. The mean value of INT for the Human and Dog's face based on appearance.

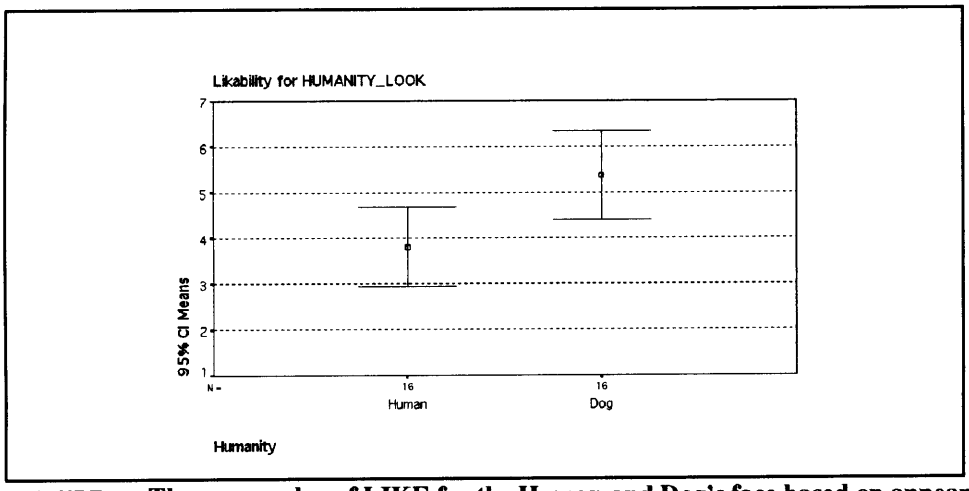

FIGURE 19. The mean value of LIKE for the Human and Dog's face based on appearance.

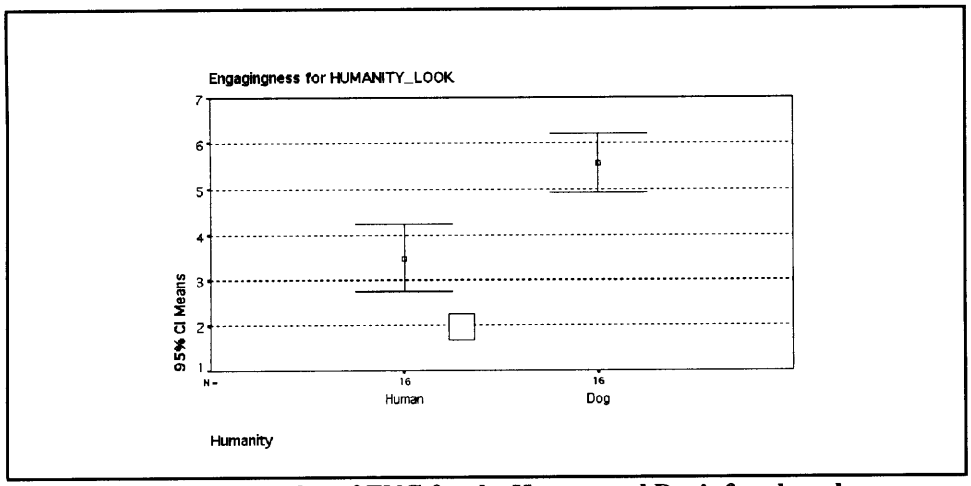

FIGURE 20. The mean value of ENG for the Human and Dog's face based on appearance.

Results of REALISM_LOOK

Comparisons between individual means were done with repeated measures analysis of variance (ANOVA) tests[1], and are summarized in Table 6. The result of the REALISM_LOOK experiment shows that there are main effects of REALISM in INT (perceived intelligence) ($F(6.12, 2)$[2], $p<.01$), LIKE (likability) ($F(5.06,2)$, $p<.05$), and ENG (engagingness) ($F(6.25, 2)$, $p<.01$). When tested by paired t-test[3], significant differences lie between Caricature and Realistic, as well as between Smiley and Realistic. There is no significant difference between Smiley and Caricature. Subjects rated the Realistic face to be more intelligent, likable, and engaging than the Caricature face and Smiley (see Figure 21, Figure 22, and Figure 23, respectively). There is also an interaction between VOTE (subjects' opinion about personification) and INT (perceived intelligence) ($F(3.57, 2)$, $p<.05$). The AGREE group attributes higher intelligence to a more realistic face, while the DISAGREE group rated the Caricature face to be least intelligent. Figure 24 illustrates this interaction.

1.Analysis of variance (ANOVA) is used to compare among more than two sample means [Norman, S. 86]. Since each subject rated three faces in this experiment, repeated measures ANOVA (equivalent with paired t-test when there are two measurements) was used to compare the means.

2.F (F ratio, degree of freedom). F ratio is a measure of the relative variation between groups to variation within groups. F=Mean Square (between) / Mean Square (within).

3.Since there are differences in the means of the three faces, the next step is to analyze where the differences are--between face 1 and 2, face 2 and 3, and/or face 1 and 3--by paired t-test.

TABLE 6. The mean value of variables for REALISM conditions based on appearance.

Variables	Mean for Smiley (n=16)	Mean for Caricature (n=16)	Mean for Realistic (n=16)	F(value, df) t (paired)	significance
INT (Intelligence)	3.53	3.59	4.94	F (6.12, 2)	p <.01
Vote = Agree (n = 7)	2.83	3.50	5.33	F (3.57, 2)	p <.05
Vote = Disagree (n = 10)	4.20	3.60	4.60		
Smiley vs. Caricature				t(15) = -.13	n.s.
Caricature vs. Realistic				t(15) = -3.63	p <.01
Smiley vs. Realistic				t(15) = -2.57	p <.05
LIKE (Likability)	3.76	4.06	4.88	F (5.06. 2)	p <.05
Smiley vs. Caricature				t(15) = -.57	n.s.
Caricature vs. Realistic				t(15) = -3.63	p <.01
Smiley vs. Realistic				t(15) = -2.51	p <.05
ENG (Engagingness)	3.65	3.82	5.12	F (6.25, 2)	p <.01
Smiley vs. Caricature				t(15) = -.30	n.s.
Caricature vs. Realistic				t(15) = -2.86	p <.01
Smiley vs. Realistic				t(15) = -3.03	p <.01

Cells contain the mean rating on a 1-7 scale where 1 was the negative extreme and 7 was the positive extreme.

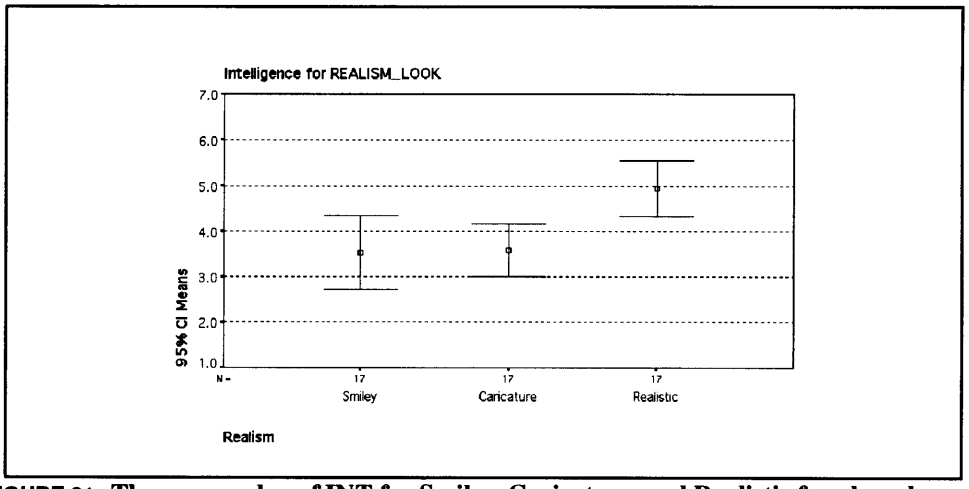

FIGURE 21. The mean value of INT for Smiley, Caricature, and Realistic face based on appearance.

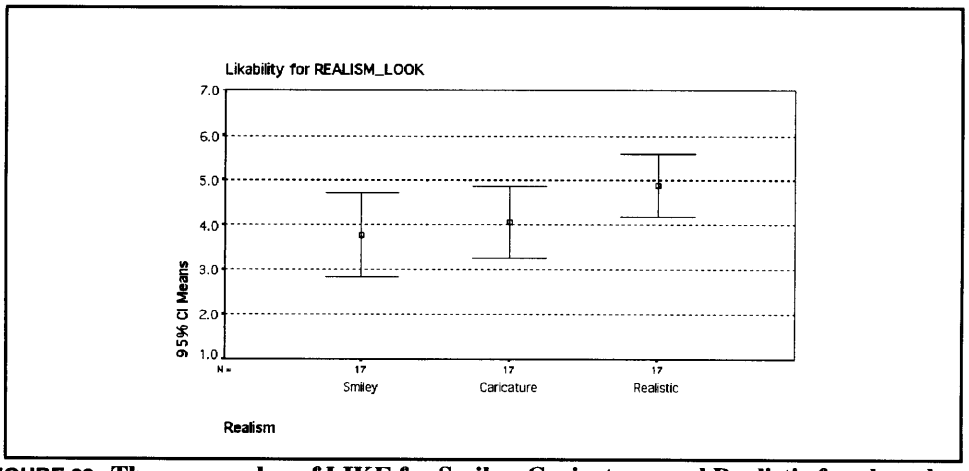

FIGURE 22. The mean value of LIKE for Smiley, Caricature, and Realistic face based on appearance.

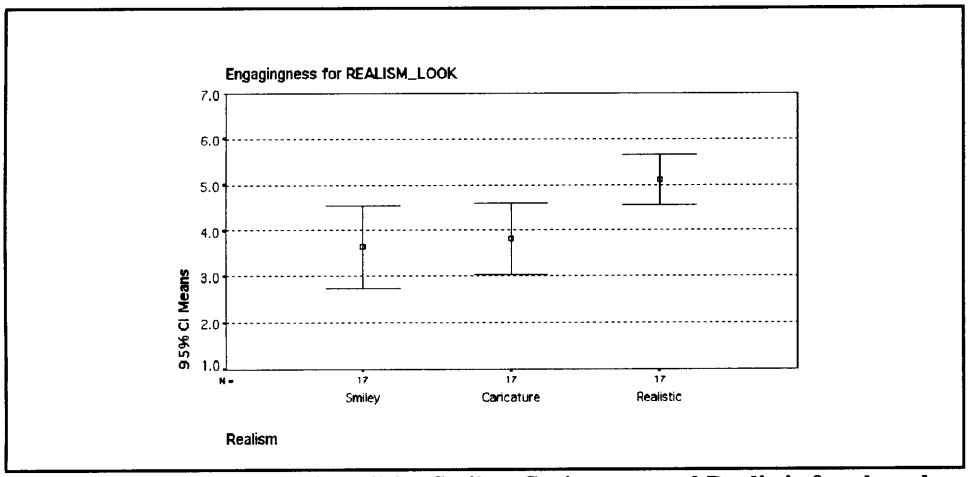

FIGURE 23. The mean value of ENG for Smiley, Caricature, and Realistic face based on appearance.

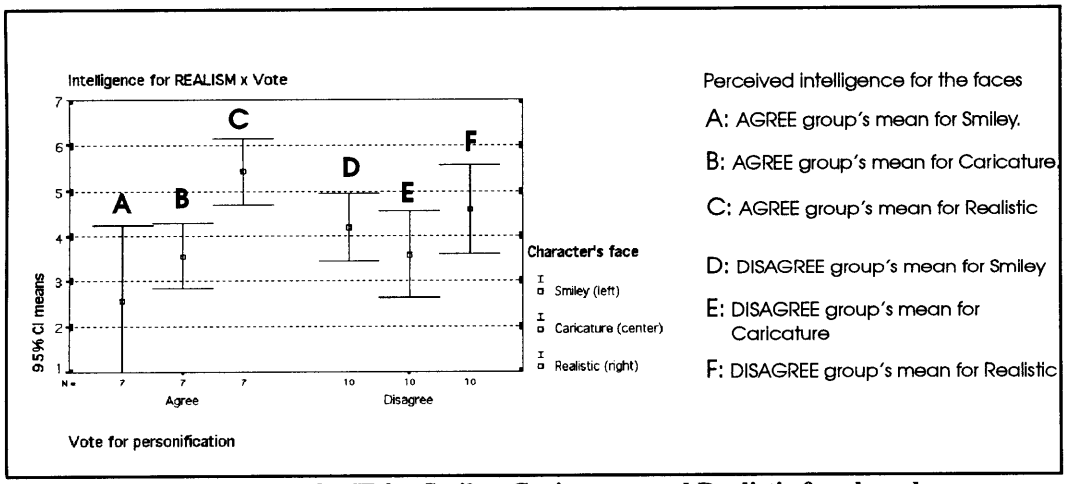

FIGURE 24. The mean value of INT for Smiley, Caricature, and Realistic face based on appearance, for the AGREE and DISAGREE group.

Hypotheses

Six hypotheses were set up based on the results from the pilot experiment.

- *Hypothesis 0*

  There is NO difference in intelligence, likability, and engagingness between the impressions based on appearance (results of the LOOK pilot experiment) and those based on performance (results of the web-based experiments).

- *Hypothesis 1 (FACE)*

  There ARE differences in intelligence, likability, engagingness, and comfortableness between Face and NoFace condition.

- *Hypothesis 2 (GENDER)*

  There are NO differences in intelligence, likability, engagingness, and comfortableness between different genders (Male and Female).

- *Hypothesis 3 (HUMANITY)*

  There ARE differences in intelligence, likability, engagingness, comfortableness, correspondence between the Human and Dog's faces.

- *Hypothesis 4 (REALISM)*

  There ARE differences in intelligence, likability, engagingness, comfortableness, and correspondence between three levels of realism (Smiley, Caricature, and Realistic).

- *Hypothesis 5 (EXPRESSIVENESS)*

  There ARE differences in intelligence, likability, engagingness, comfortableness, required attention, level of distraction, believability, and usefulness between three different modes of expressiveness (Honest, Deceiving, and Stoic). This hypothesis was set up according to the Walker *et al.*'s study [Walker 94], which indicates that a face with more expression leads to greater engagement.

Analysis of Competence

As described in Chapter 4, all poker playing agents have the same strategy. Furthermore, there was no significant variation among the mean winnings of differentiated groups of agents. In other words, all poker playing agents played at the same level of ability. Table 7 shows the mean ending money of subjects and the poker playing agents.

TABLE 7. The mean ending money of subjects and the poker playing agents

Experiments		Agents	Agents			
FACE	Subjects	Face	NoFace	t (paired)	Sig.	-
The means of money	297	140	170	-	-	-
Subjects vs. Face	-	-	-	t(29, 3.92)	p=0.0005	-
Subjects vs. NoFace	-	-	-	t(29, 2.85)	p=0.008	-
Face vs. NoFace	-	-	-	t(29, 2.85)	p=0.23	-
GENDER	Subjects	Male	Female	t (paired)	Sig.	-
The means of money	285	154	179	-	-	-
Subjects vs. Male	-	-	-	t(41, 4.02)	p = 0.0002	-
Subjects vs. Female	-	-	-	t(41, 3.43)	p = 0.001	-
Male vs. Female	-	-	-	t(41, 0.86)	p = 0.39	-

TABLE 7. The mean ending money of subjects and the poker playing agents

Experiments		Agents	Agents			
HUMANITY	Subjects	Human	Dog	t (paired)	Signifi-cance	-
The means of money	260	160	184	-	-	-
Subjects vs. Human	-	-	-	t(28, 2.99)	p = 0.006	-
Subjects vs. Dog	-	-	-	t(28,2.24)	p = 0.03	-
Human vs. Dog	-	-	-	t(28, 0.81)	p = 0.43	-
REALISM	Subjects	Smiley	Caricature	Realistic	t (paired)	Sig.
The means of money	287	159	167	187	-	-
Subjects vs. Smiley	-	-	-	-	t(32, 3.66)	p=0.0009
Subjects vs. Caricature	-	-	-	-	t(32, 3.55)	p=0.001
Subjects vs. Realistic	-	-	-	-	t(32, 2.57)	p=0.01
Smiley vs. Caricature	-	-	-	-	t(32, -0.29)	p=0.77
Smiley vs. Realistic	-	-	-	-	t(32, -0.76)	p=0.45
Caricature vs. Realistic	-	-	-	-	t(32, -0.54)	p=0.59
EXPRESSIVENESS	Subjects	Honest	Deceiving	Expressive	t (paired)	Sig.
The means of money	285	156	191	155	-	-
Subjects vs. Honest	-	-	-	-	t(22, 3.17)	p=0.004
Subjects vs. Deceiving	-	-	-	-	t(23, 1.92)	p=0.06
Subjects vs. Stoic	-	-	-	-	t(23, 3.19)	p=0.004
Honest vs. Deceiving	-	-	-	-	t(23, -1.06)	p=0.30
Honest vs. Stoic	-	-	-	-	t(23, 0.04)	p=0.97
Deceiving vs. Stoic	-	-	-	-	t(23, 1.10)	p=0.28

Difference between Face and NoFace (Experiment 1)

Results

Twenty-five subjects returned answers for the FACE experiment. Of these subjects, 21 of them were male, and 10 of them were in favor of personification. The subjects' ages ranged from 11-49 years old. Ten of them were intermediate computer users and the rests rated themselves as experts. Dependent variables for this experiment are intelligence (INT), likability (LIKE), engagingness (ENG), comfortableness (COM) for each condition (FACE: Face vs. NoFace), and required attention to the face (ATT_FACE), level of distraction by the face (DIS_FACE).

Overall, the results **supported** *hypothesis 1*--there ARE differences in intelligence, likability, engagingness, and comfortableness between Face and NoFace condition. Subjects paid attention to the face (ATT = 5.28, where 4.0 is neutral), but were not distracted by the existence of the face (DIS = 2.84, where 4.0 is neutral). Comparisons between individual means were done by paired t-tests, and are summarized in Table 8. There is a main effect of FACE (whether there is a face or without) for LIKE, ENG, and COM. Subjects rated Face condition (the Caricature Female face) to be significantly more likable ($t(24)=2.76$, $p<.05$), more engaging ($t(24)=2.85$, $p<.01$), and more comfortable to play with ($t(24)=3.36$, $p<.01$), than NoFace condition (the Invisible Man). Figure 25, Figure 26, and Figure 27 illustrate the effects of having a face in terms of likability, engagingness, and comfortableness, respectively. However, there is no difference in intelligence between Face and NoFace. There are no interactions between FACE and subjects' gender (SS_GEN) or subjects' opinion about personification (VOTE).

TABLE 8. The mean value if variables for FACE condition (Face vs. NoFace).

Cells contain the mean rating on a 1-7 scale where 1 was the negative extreme and 7 was the positive extreme.

Variables	Mean for Face (n=25)	Mean for NoFace(n=25)	t (paired)	significance
INT (Intelligence)	3.64	3.80	t(24) = -.27	n.s.
LIKE (Likability)	5.08	3.72	t(24) = 2.76	p <.05
ENG (Engagingness)	4.60	3.32	t(24) = 2.85	p <.01
COM (Comfortableness)	4.88	3.36	t(24) = 3.06	p <.01
ATT (Attention to the face)	5.28	-	-	-
DIS (Distracted by the face)	2.84	-	-	-

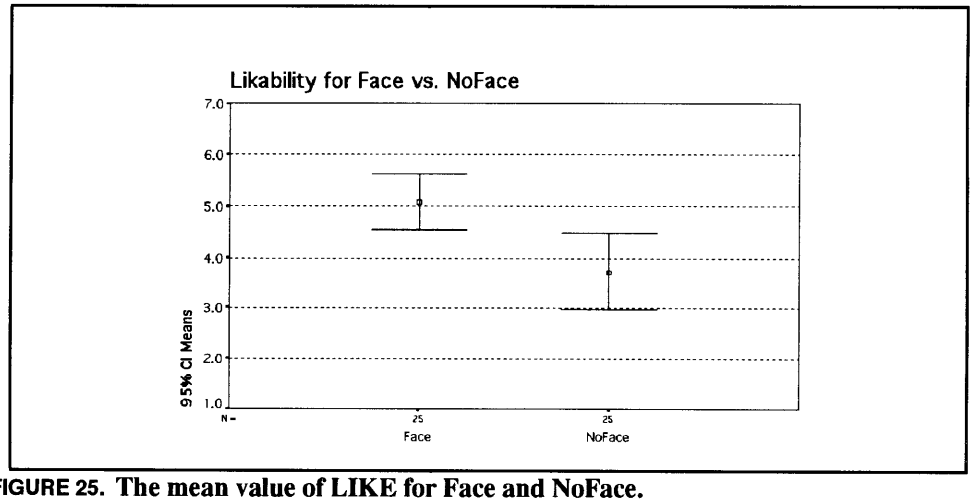

FIGURE 25. The mean value of LIKE for Face and NoFace.

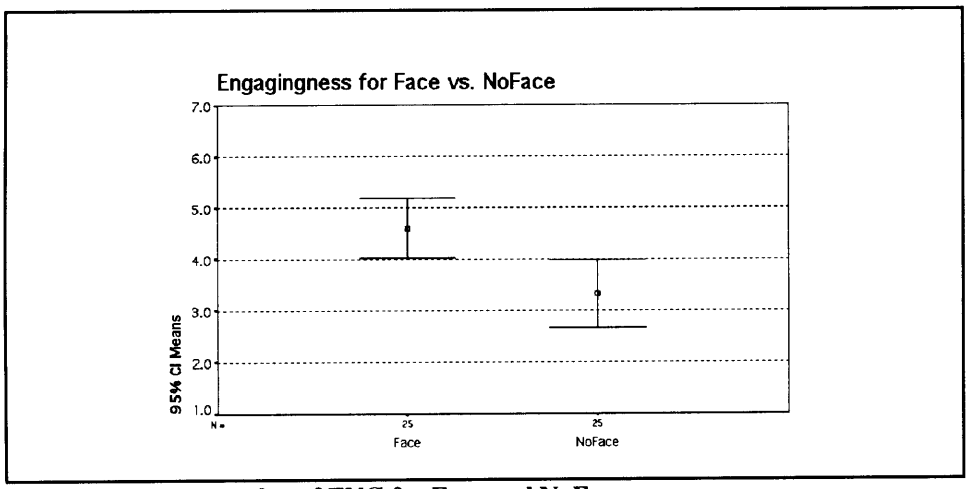

FIGURE 26. The mean value of ENG for Face and NoFace.

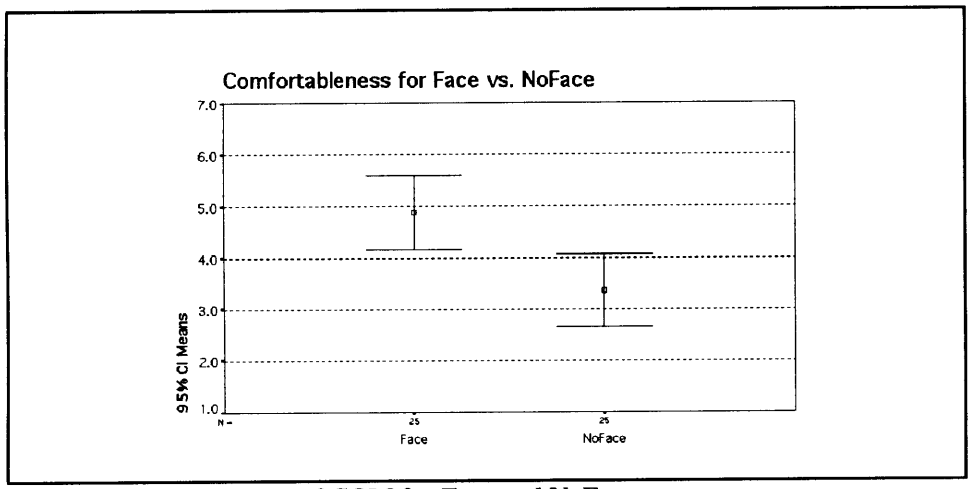

FIGURE 27. The mean value of COM for Face and NoFace.

Discussion

Subjects rated a poker playing agent with a face and without as equally intelligent. Attaching a face did not add any intelligence to the poker player. However, having a face is more

likable, engaging, and comfortable to play against regardless of their opinion about personification. As described in Chapter 2, Takeuchi's [Takeuchi 94] and Walker's [Walker 94] studies show a similar result, that having a face is engaging. It is encouraging that the face did not create a negative effect on the subjects' likability, even to those who have a negative opinion about personification, contrary to Walker's study [Walker 94] (see Chapter 2). This difference may be caused by the nature of the tasks and faces used in these studies. Walker used a mean-looking old woman's synthetic talking face in a questionnaire survey, while this study used a female caricature face[1] in a poker game environment.

Difference between Characters's Gender (Experiment 2)

Results

Thirty-seven subjects returned answers for the GENDER experiment. Of these subjects, 27 of them were male, and 18 of them were in favor of personification. The subjects' ages ranged from 11-39 years old. Thirteen of them are intermediate computer users and the rests rated themselves as experts. Dependent variables for this experiment are intelligence (INT), likability (LIKE), engagingness (ENG), comfortableness (COM) for each condition (GENDER: Male vs. Female).

Hypothesis 2--There is NO difference in perceived intelligence, likability, engagingness, and comfortableness between gender--is **accepted**. *Hypothesis 6*--there is NO difference in perceived intelligence, likability, and engagingness between the impressions based on appear-

1.One subject "fell in love" (from his comments) with the Female Caricature agent. He played the poker game for more than two hours until he could see the Female Caricature Face again and again.

ance and those based on performance--is also **accepted**. Comparisons between individual means were done with paired t-tests, and are summarized in Table 9. There is no main effect of agent's gender (GENDER) in any variables. However, there are significant interactions between VOTE and INT ($F(4.73, 1)$, $p < .05$), and slight interactions between VOTE and LIKE ($F(3.62, 1)$, $p < .10$). As shown in Figure 28 and Figure 29, the AGREE group (i.e., those who are for personification) rated intelligence and likability for the Male face higher and the Female face lower. While the DISAGREE group (i.e., those who are against personification) rated the Male and the Female face in the opposite way. No other 2-way or 3-way interactions were found.

TABLE 9. The mean value of variables for GENDER conditions (Male vs. Female).
Cells contain the mean rating on a 1-7 scale where 1 was the negative extreme and 7 was the positive extreme.

Variables	Mean for Male (n=37)	Mean for Female (n=37)	t (paired) F(value, df)	significance
INT (Intelligence)	3.35	3.83	t(36) = -1.33	n.s.
Vote = Agree (n=18)	3.94	3.55	F(4.73, 1)	p <.05
Vote = Disagree (n=19)	2.78	4.10		
LIKE (Likability)	4.49	4.62	t(36) = -.55	n.s.
Vote = Agree (n = 18)	5.11	4.83	F(3.62,1)	0 <.10
Vote = Disagree (n = 19)	3.89	4.42		
ENG (Engagingness)	4.32	4.43	t(36) = -.55	n.s.
COM (Comfortableness)	4.00	3.92	t(36) =.30	n.s.

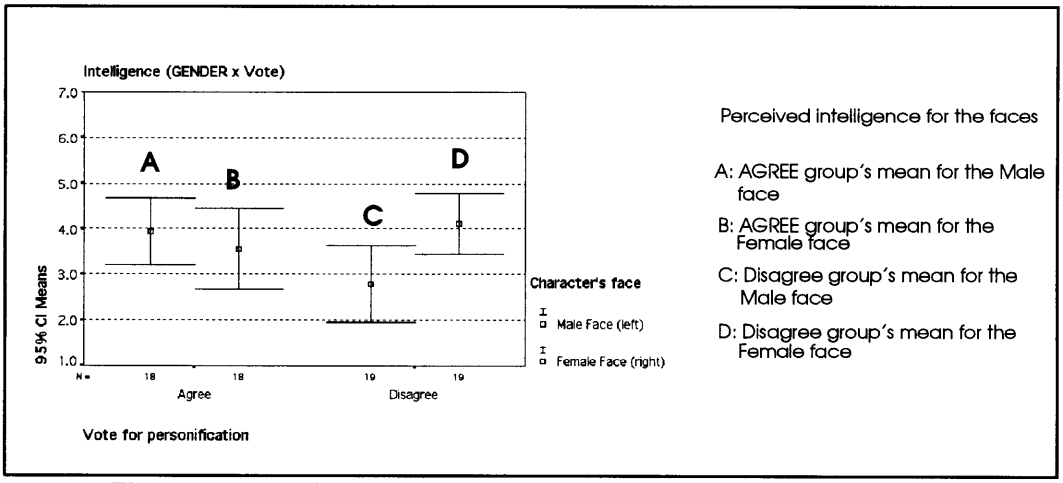

FIGURE 28. The mean value of INT for the Male and Female face, for both the AGREE and DISAGREE groups.

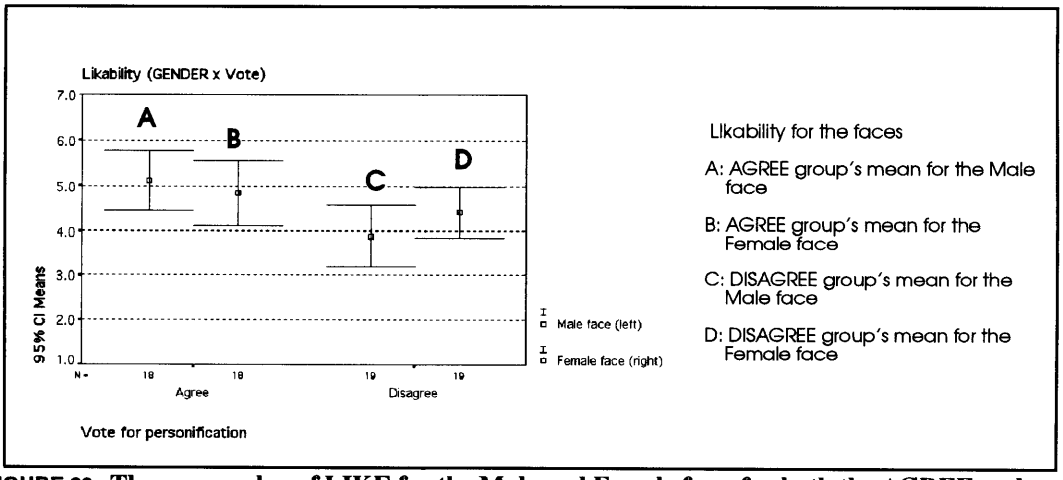

FIGURE 29. The mean value of LIKE for the Male and Female face, for both the AGREE and DISAGREE groups.

Difference between Humanity (Experiment 3)

Results

Twenty-three subjects returned answers for the HUMANITY experiment. Of these subjects, 16 of them were male, and 50% of them were in favor of personification. The subjects' ages ranged from 11-39 years old. Fifteen of them were intermediate computer users and the rests rated themselves as experts. Dependent variables for this experiment are intelligence (INT), likability (LIKE), engagingness (ENG), comfortableness (COM), and correspondence (COR) for each condition (HUMANITY: Human vs. Dog).

Hypothesis 0--There is NO difference in perceived intelligence, likability, and engagingness between the impressions based on appearance and those based on performance--is **rejected**. *Hypothesis 3*--There ARE differences in perceived intelligence, likability, engagingness, comfortableness, correspondence between the "Human" and "Dog" condition--is **accepted**. Comparisons between individual means were done with paired t-tests, and are summarized in Table 10. There is no main effect of HUMANITY (Human or Dog's face) in any variables. However, there are significant interactions between VOTE and LIKE ($F(5.40, 1)$, $p < .05$), COM ($F(11.42, 1)$, $p < .01$), and COR ($F(6.97, 1)$, $p < .05$). As shown in Figure 30, Figure 31, and Figure 32, the AGREE group rated likability, comfortableness, and correspondence of the Human face significantly lower while they rated the Dog's face significantly higher. The DISAGREE group rated the Human and the Dog's face in the opposite way. There are also interactions between SS_GEN and LIKE ($F(9.24, 1)$, $p < .01$), ENG ($F(5.94, 1)$, $p < .05$), and COM ($F(7.26, 1)$, $p < .05$). Due to the small number of female subjects (6 out of 23 subjects) and wide variances, further study is needed to conclude that the

differences are truly significant, even though the analysis shows significance. The result suggests male subjects rated the Dog's face as slightly more likable, engaging, and comfortable than the Human face, while female subjects rated in the opposite way (see Figure 33, Figure 34, and Figure 35, respectively).

TABLE 10. The mean values of variables for HUMANITY conditions (Human vs. Dog).
Cells contain the mean rating on a 1-7 scale where 1 was the negative extreme and 7 was the positive extreme.

Variables	Mean for Human (n=23)	Mean for Dog (n=23)	t (paired) F(value, df)	significance
INT (Intelligence)	3.73	3.82	t(22) = -.18	n.s.
LIKE (Likability)	4.56	4.39	t(22) =.53	n.s.
Vote = Agree (n = 11)	4.72	4.73	F(5.40, 1)	p <.05
Vote = Disagree (n = 12)	4.45	4.09		
SS_Gen = Male (n = 17)	4.47	4.65	F(9.24, 1)	p <.01
SS_Gen = Female (n = 6)	4.83	3.67		
ENG (Engagingness)	3.95	4.13	t(22) = -1.00	n.s
SS_Gen = Male (n = 17)	3.88	4.29	F(5.94, 1)	p <.05
SS_Gen = Female (n = 6)	4.17	3.67		
COM (Comfortableness)	4.08	4.04	t(22) =.13	n.s.
Vote = Agree (n = 11)	3.91	4.55	F(11.42, 1)	p <.01
Vote = Disagree (n = 12)	4.27	3.55		
SS_Gen = Male (n = 17)	4.29	4.41	F(7.26, 1)	p <.05
SS_Gen = Female (n = 6)	3.50	3.00		
COR (Correspondence of the face to actual intelligence)	3.91	3.50	t(22) = 1.16	n.s.
Vote = Agree (n = 11)	3.82	4.18	F(6.97, 1)	p <.05
Vote = Disagree (n = 12)	4.00	2.82		

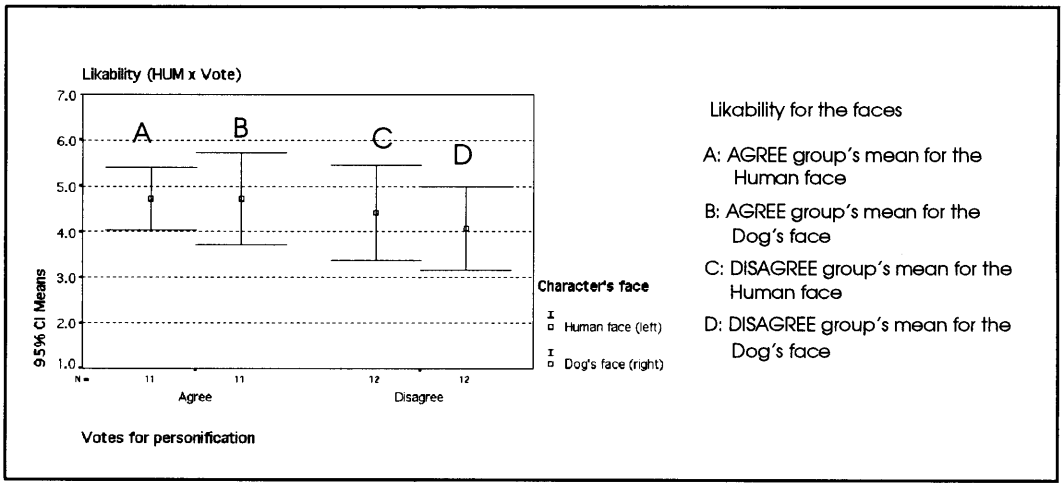

FIGURE 30. The mean value of LIKE for the Human and Dog's face, for both the AGREE and DISAGREE groups.

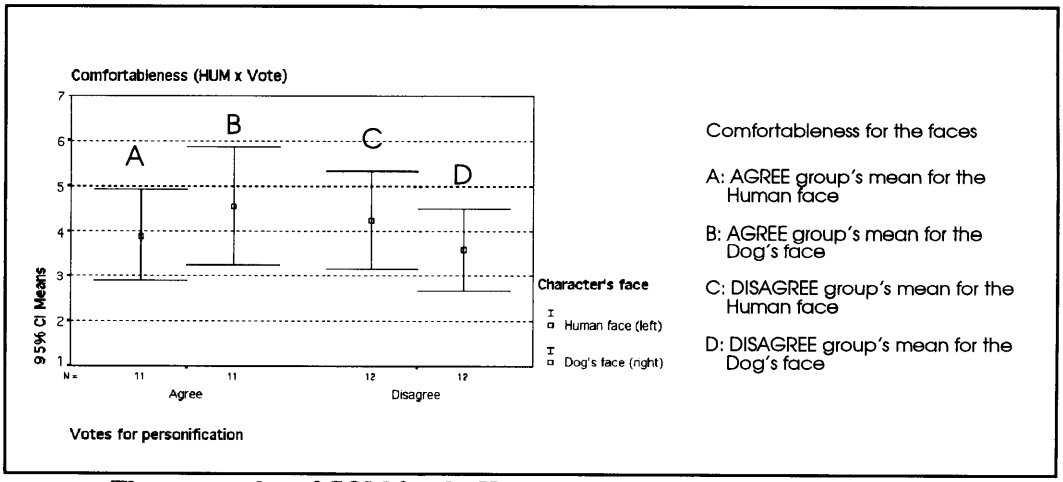

FIGURE 31. The mean value of COM for the Human and Dog's face, for both the AGREE and DISAGREE groups.

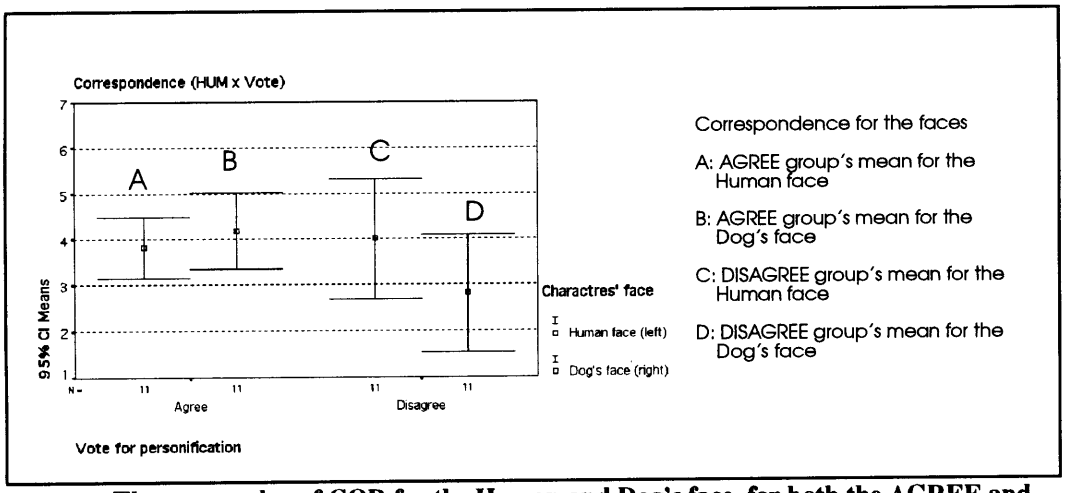

FIGURE 32. The mean value of COR for the Human and Dog's face, for both the AGREE and DISAGREE groups.

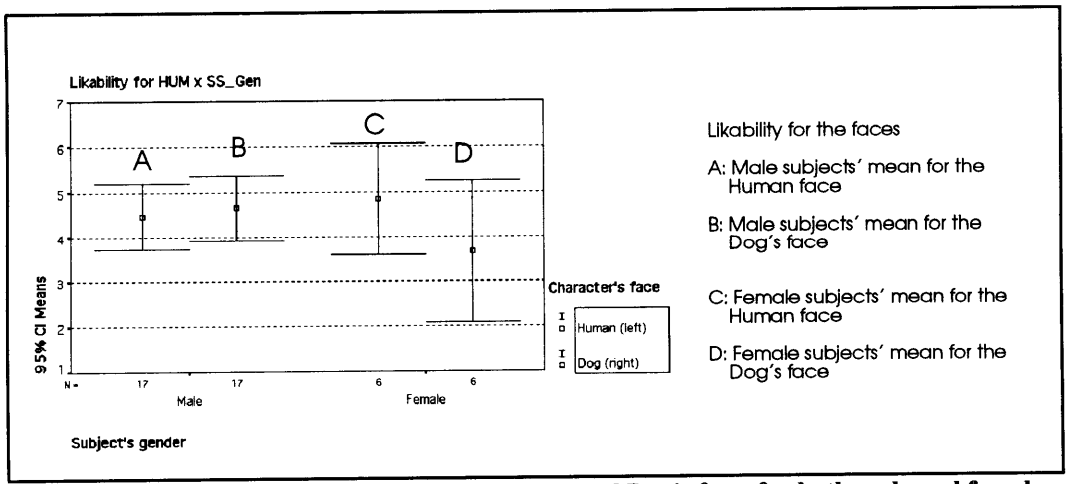

FIGURE 33. The mean value of LIKE for the Human and Dog's face, for both male and female groups.

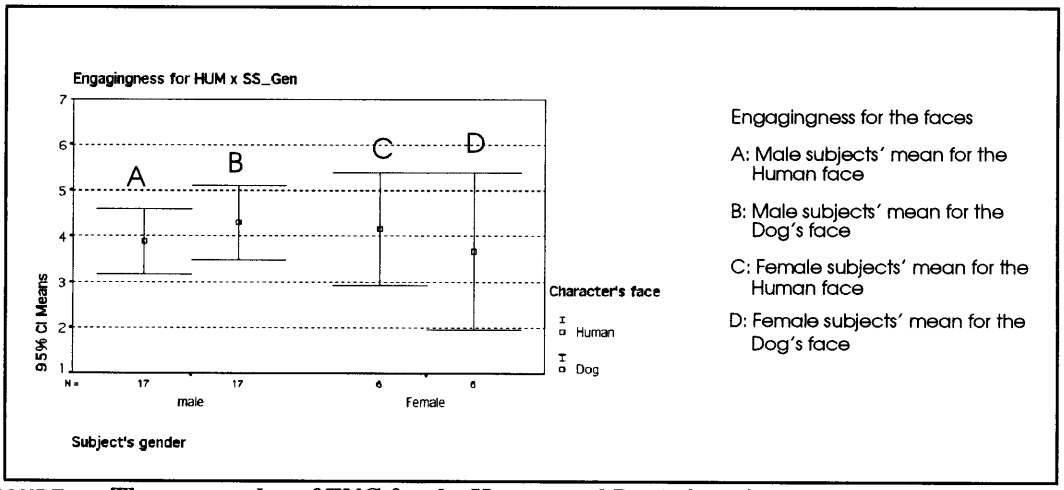

FIGURE 34. The mean value of ENG for the Human and Dog's face, for both male and female groups.

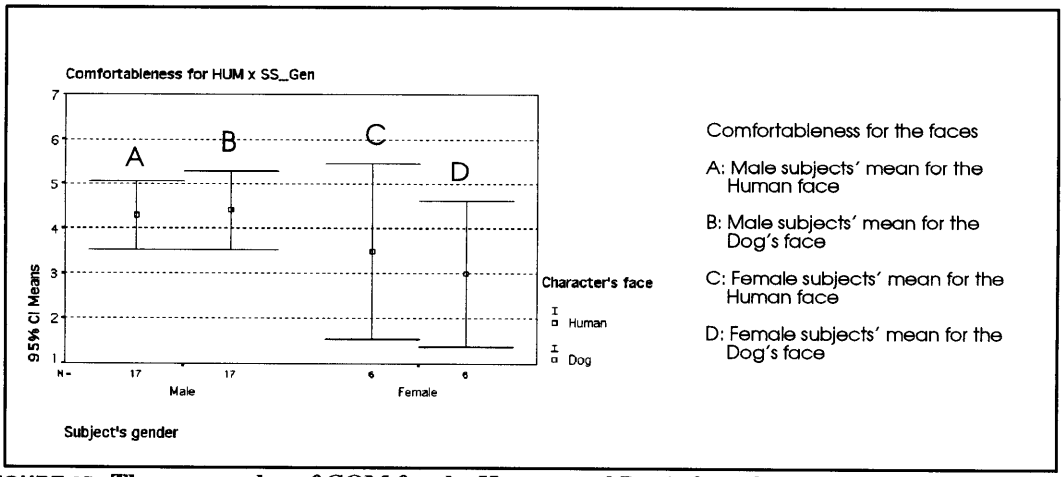

FIGURE 35. The mean value of COM for the Human and Dog's face, for both male and female groups.

Discussion

Subjects' responses to the LOOK_HUMANITY experiment (see Table 5 on page 70) show reasonable expectations from the appearance of each face--humans are regarded as more intelligent than dogs, but playing a poker game against a dog in a computer game might be likable and engaging--, regardless of their opinion about personification. King's study [King 96] shows a similar result in perceived intelligence, that human forms are perceived to be more intelligent than other forms. However, subjects rated the Human and Dog's faces to be equally intelligent (or unintelligent) in the poker game. This means that they did not rate each player's intelligence based on its appearance. Their impressions are based on the player's actual competence in playing poker. However, their opinions about personification affects their attitude toward each face. The AGREE group are more "strict" about personification, since they prefer the Dog's face to the Human face, felt more comfortable with the Dog's face, and considered the Dog's face more appropriate for representing the actual competence of the player in playing poker. On the other hand, the DISAGREE group are more "generous" about personification. They tend to insist on using the Human face for representing a poker player. The subject's gender works in the same way as the subjects' opinion about personification. Males are more "strict", while female are more "generous" about personification. However, due to the small number of female subjects, further study is needed to conclude that users' gender effects their impressions of agent's humanity.

Difference between Realism (Experiment 4)

Results

Thirty subjects returned answers to the REALISM experiment. Of these subjects, 22 of them were male, and 16 of them were in favor of personification. The subjects' ages ranged from 11-39 years old. Nine of them were intermediate computer users and the rests rated themselves as experts. Dependent variables for this experiment are intelligence (INT), likability (LIKE), engagingness (ENG), comfortableness (COM), and correspondence (COR) for each condition (REALISM: Smiley vs. Caricature vs. Realistic).

Hypothesis 0--There is NO difference in perceived intelligence, likability, and engagingness between the impressions based on appearance and those based on performance-- is **rejected** again. *Hypothesis 4*--there ARE differences in perceived intelligence, likability, engagingness, comfortableness, correspondence between the three levels of realism--is **rejected.** Comparisons between individual means were done with repeated measures analysis of variance (ANOVA) tests, and are summarized in Table 11. There is no main effect of REALISM in INT, LIKE, and ENG. The result shows that the Realistic face is slightly more intelligent, likable, engaging than other faces, though not significant. There is a main effect of REALISM in COM ($F(.05, 2)$, $p = .067$) and a significant main effect in COR ($F(.18, 2)$, $p < .05$). Comparing two of the three faces by paired t-test, there is a significant difference in comfortableness between Smiley and Realistic face. The Realistic face is rated as more comfortable than the Smiley face. In terms of correspondence to the actual poker playing skill, subjects rated the Smiley face as the most appropriate representation, and the caricature and

Realistic faces are less appropriate. These main effects are illustrated in Figure 36 and Figure 37. No 2-way or 3-way interactions are found.

TABLE 11. The mean value of variables for REALISM conditions (Smiley vs. Caricature vs. Realistic).
Cells contain the mean rating on a 1-7 scale where 1 was the negative extreme and 7 was the positive extreme.

Variables	Mean for Smiley (n=30)	Mean for Caricature (n=30)	Mean for Realistic (n=30)	F(value, df) t (paired)	significance
INT (Intelligence)	3.37	3.53	3.70	F (.30, 2)	n.s.
LIKE (Likability)	4.20	4.10	4.70	F (.59, 2)	n.s.
ENG (Engagingness)	4.17	4.13	4.76	F (.58, 2)	n.s
COM (Comfortableness)	4.07	4.00	4.63	F (.05, 2)	p =.067
Smiley vs. Caricature				t(29) =.14	n.s.
Caricature vs. Realistic				t(29) = -1.60	n.s.
Smiley vs. Realistic				t(29) = -1.71	p <.1
COR (Correspondence of the face to actual intelligence)	4.67	3.90	4.03	F (.18, 2)	p =.074
Smiley vs. Caricature				t(29) = 2.25	p <.05
Caricature vs. Realistic				t(29)= -.36	n.s.
Smiley vs. Realistic				t(29) = 2.00	p =.055

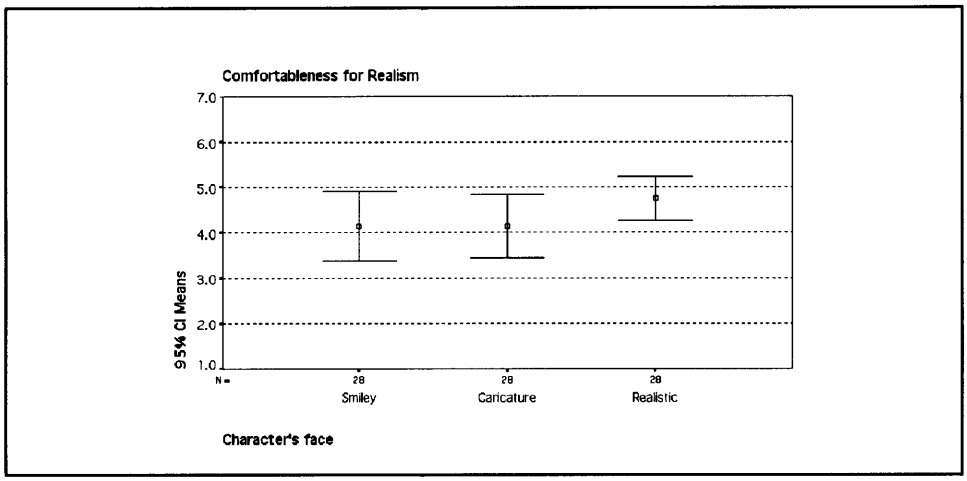

FIGURE 36. The mean value of COM for Smiley, the Caricature, and the Realistic face.

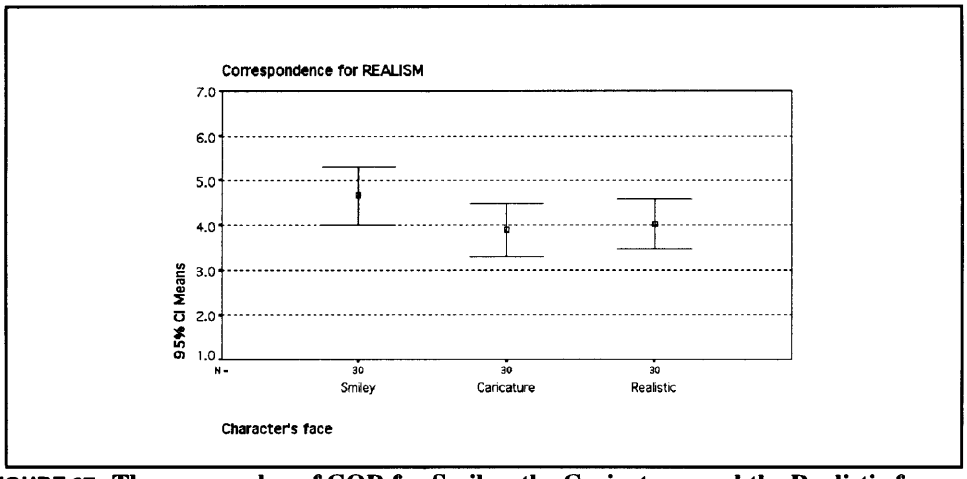

FIGURE 37. The mean value of COR for Smiley, the Caricature, and the Realistic face.

Discussion

The REALISM_LOOK experiment shows that the Realistic face looks more intelligent, engaging, and likable than the less realistic faces (see Table 6). Subjects categorized the faces in two ways--realistic face and not-realistic (Caricature and Smiley) when they evaluate the faces based on appearance. The level of realism does not affect the overall impression of faces, unless they are not realistic. This result is the same as King's study [King 96] of subjects' appraisal of intelligence and agency of various images. King's study indicates fully articulated human forms are rated to have significantly higher intelligence and agency than either the caricatures or the Chernoff faces (corresponds to Smiley face in this experiment), and caricatures and Chernoff faces are rated to have similar intelligence and agency. There is an evidence that people's opinion about personification affects the impression of faces again in this REALISM_LOOK experiment. The AGREE group tends to attribute more intelligence to the Realistic face than the DISAGREE group does.

However, subjects rated the three faces differently after playing poker. They judged the level of perceived intelligence of a face based not on its appearance but instead on its competence, as seen in experiment 3 (HUMANITY experiment). Subjects rated the three faces to be equally intelligent, and they thought that the Smiley face (less realistic) face represents the true level of competence more appropriately. Though the effort for personification focuses on synthesizing a human face as realistic and as possible [Takeuchi 93, Lee 95], realism may not be the main issue to make a character look more intelligence or likable.

Difference between Levels of Expressiveness (Experiment 5)

Results

Twenty subjects returned answers to the EXPRESSIVENESS experiment. Of these subjects, 16 of them were male, and 50% of them were in favor of personification. The subjects' ages ranged from 11-39 years old. Ten of them were intermediate computer users and the rests rated themselves as experts. Dependent variables for this experiment are intelligence (INT), likability (LIKE), engagingness (ENG), comfortableness (COM), believability of the player's facial expressions (BEL), required attention for the facial expressions (ATT), distraction by the facial expressions (DIS), and usefulness of the facial expressions (USE) for each condition (EXPRESSIVENESS: Honest vs. Deceiving vs. Stoic).

Hypothesis 5--there ARE differences in perceived intelligence, likability, engagingness, comfortableness, required attention, distraction, believability, and usefulness between three different expressiveness (Honest, Deceiving, and Stoic)--is **partly accepted**. Comparisons between individual means was done with repeated measures ANOVA, and are summarized in Table 12. There is no main effect of EXPRESSIVENESS in any variables. However, there are significant interactions between VOTE and INT ($F(8.91, 2)$, $p < .01$) and COM ($F(4.47, 2)$, $p < .05$). As shown in Figure 38, the AGREE group rated Honest face as most intelligent, while the DISAGREE group rated the Stoic face as the most intelligent and the Honest face as the least. The AGREE group rated the Stoic face as least comfortable, while the DISAGREE rated the same face as most comfortable, as shown in Figure 39.

TABLE 12. The mean value of variables for EXPRESSIVENESS conditions (Honest vs. Deceiving vs. Stoic).

Cells contain the mean rating on a 1-7 scale where 1 was the negative extreme and 7 was the positive extreme.

Variables	Mean for Honest (n=20)	Mean for Deceiving (n=20)	Mean for Stoic (n=20)	F(value, df)	significance
INT (Intelligence)	3.84	3.37	3.89	F(1.98, 2)	n.s.
Vote = Agree (n = 10)	4.20	3.10	3.30	F(8.91, 2)	p <.01
Vote = Disagree (n = 10)	3.44	3.66	4.55		
LIKE (Likability)	4.37	4.37	4.68	F(1.28, 2)	n.s.
ENG (Engagingness)	3.95	4.21	4.21	F(.56, 2)	n.s.
COM (Comfortableness)	3.95	3.74	3.95	F(.14, 2)	n.s.
Vote = Agree (n = 10)	3.90	4.10	3.60	F(4.47, 2)	p <.05
Vote = Disagree (n = 10)	4.00	3.33	4.33		
BEL (Correspondence of the face to actual intelligence)	4.79	4.79	4.63	F(6.44, 2)	n.s.
ATT (Required attention to the facial expressions)	4.32	4.47	4.37	F(1.46, 2)	n.s.
DIS (distraction by the facial expressions)	3.26	3.21	3.37	F(5.26, 2)	n.s.
USE (usefulness of the facial expressions)	4.21	4.42	4.11	F(1.81, 2)	n.s.

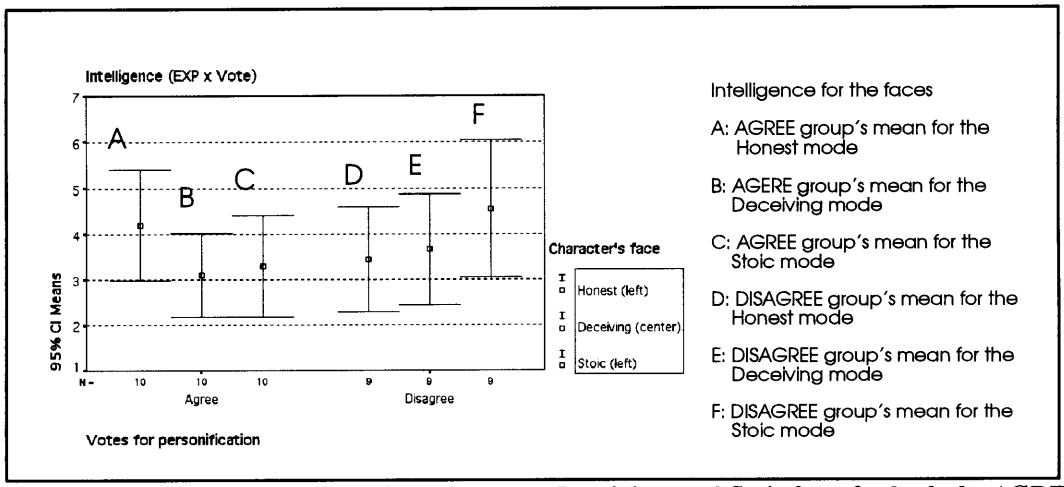

FIGURE 38. The mean value of INT for the Honest, Deceiving, and Stoic face, for both the AGREE and DISAGREE groups.

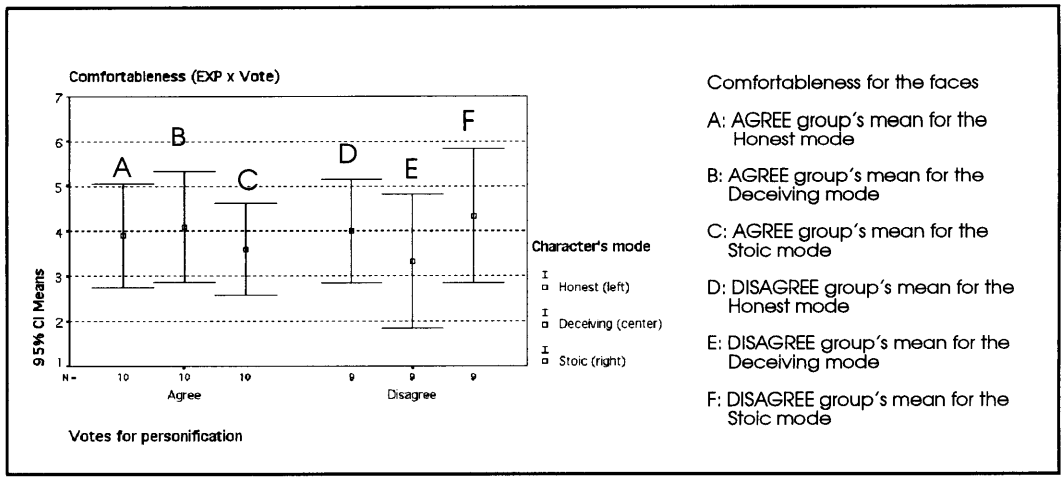

FIGURE 39. The mean value of COM for the Honest, Deceiving, and Stoic face, for both the AGREE and DISAGREE group.

Discussion

It is interesting that people's opinion about personification changes the way they feel about expressiveness. The AGREE group rated expressive faces (Honest and Deceiving) to be more comfortable, and the Honest face to be most intelligent. The DISAGREE group rated the stoic face as most intelligent and comfortable. This means the DISAGREE group is more strict about simulating a real poker situation, where players are not expected to show honest expressions. Subjects who wrote comments about using a poker face to make the game more realistic were all in the DISAGREE group. One DISAGREE subject called the Stoic player "sophisticated." Though there are no significant differences between likability, engagingness, believability, required attention, distraction, and usefulness, the result shows that subjects paid attention to facial expressions (ATT > 4 where 4 is neutral) but were not distracted by them (DIS < 4 where 4 is neutral). They believed facial expressions (BEL > 4 where 4 is neutral), and rated the Stoic face to be less believable about its honesty. Subject's comments show that facial expressions changed so quickly that they sometimes could not catch up with the changes, though the appropriate turn-taking speed was tested beforehand (see Chapter 4). This might be one reason that there was no difference in the above measurements.

Observations and Open Comments

Subjects' Reactions

Interviews and observations suggest that subjects tried to read the agent's poker playing strategy from its face. Then they attributed different personalities and characteristics to each face and thought each agent had a different playing strategy.

Seven subjects were called in to the Media Lab, participated in all experiments and had a interview. All subjects answered that the poker players' strategies are too predictable from their facial expressions. They all thought that the players have different strategies and personalities. The following comments support this belief: "Morgan (the Caricature Male) bluffs a lot, and he is too extrovert. I paid attention to his face particularly." (in HUMANITY experiment); "Morticia (the Caricature Female) is hard to play against, because she is very stoic and her playing style is sophisticated. She reminds me of my girlfriend." (in GENDER experiment); "Bruce (the Realistic Male) changed his strategy during the game to beat me. I want to continue this game until I beat him!" (in REALISM experiment) When they were asked about their preferable faces, they all had their own preferable character. Subjects responded emotionally to the agents. They laughed at the player's facial expressions and got upset when they lost a game.

Subjects wrote many comments at the end of the questionnaire. Their comments can be divided into two categories: their experience with the players and how to make a more sophisticated and realistic poker game.

In the first category, some of the subjects used "attitude" or "personality" to describe the characteristics of the players in their comments. Comments show that the subjects first tried

to figure out whether the faces gave them clues on the player's hands and strategies. Then they noticed that the players' facial expressions gave information about their hands. Some subjects associated a person they know to a player, and attributed the person's personality to be that of the player. In the second category, some subjects criticized that the poker players should not be expressive in order to make a successful poker game.

Suggestions for Improving the User Interface

Subjects submitted many suggestions to improve the user interface of the poker game.

• The position of the message

Some of the subjects were confused by moving their eyes between a message in the center of the poker table and the players' faces. Several subjects suggested that the message should appear close to each player. They were forced to select which information they should focus on. Most subjects reported that they paid more attention to the faces than to messages. As described in Chapter 3, my intention was to make users decide which information they should focus on, faces or messages, worked well; they chose to watch the players' faces.

• Animation

Some subjects suggest adding smooth transition from one facial expression to another to make the characters more lively. They also suggested adding eye movements to the player's hand or to other players, and adding lip movement in synchronizing with audio to help making the poker game more engaging and realistic.

• Audio feedback

As mentioned above, adding audio feedback is also suggested by some subjects. If they can hear players sigh or laugh, it would make the game more engaging and realistic. Though making a realistic poker game was not the goal of this research, studying the effects of animation and audio would be interesting topics for research in personified interface.

CHAPTER 6

Conclusions and Future Directions

There are three main findings of this study. In this chapter, I will summarize these findings and address future agent-based applications with personified interfaces.

Conclusions

The Effect of Having a Face

The first finding is that having a face is considered more likable, engaging, and comfortable to see in a poker game environment. It is encouraging that people are favorable to having a face in an interface regardless of their opinion about personification. People said that they were not distracted by the presence of a face or facial expressions. Moreover, people tried to interpret faces and facial expressions, which makes the users pay attention to the face and engage in the task. Walker [Walker 94] describes the advantages and disadvantages of per-

sonification as follows: "This can lead to improved performance if the task is not very complex or to degraded performance if the task is complex." It is clear that faces are useful for entertainment purposes, since engagement is essential for games and people don't care about taking more effort for entertainment. Hence it may also be useful for applications which require an engaged user for success, such as education and training. An example of a training system might be Picard's "the effective piano teacher" [Picard 95]. In her paper "affective computing"[1], she addresses the possibility of making a computer piano teacher which reads users' emotions while practicing the piano through the users' gestural input and facial expressions. The teacher tries to encourage the users when they are frustrated or gives them more challenging exercises when the users are motivated. By attaching a face and facial expressions to the tutor would help the user engage in the training.

Need for Understanding the Context of an Application

The second finding is that people's impressions of a face are different when they see a face in isolation versus when they interact with a face within a task. People evaluate a face not based on appearance but its competence or performance. For example, though people attributed different levels of perceived intelligence, likability, and engagingness to the Human face and the Dog's face, or Smiley, the Caricature, and the Realistic face when they evaluate the faces based on their appearance, there were no differences when they rated the same faces in the poker game. Most psychological HCI studies use static facial images separately from applications. Considering what we understand from this study, we have to evaluate

1. Affective computing is "computing that relates to, arise from, or deliberately influences emotions" [Picard 95] by equipping a computer to detect human emotions and express emotions by itself.

effects of personification within a context--software agent applications. Of course not all software agent applications require personified interfaces. The goal of HCI work should be to understand when a personified interface is appropriate. One example of the nature of the task or context is whether people are used to do a task with other people. In the poker game, the nature of the task is social. People accept seeing their opponents' faces in a computer poker game as they see in a real poker game. On the other hand, people might be annoyed to see a face in an automatic teller machine.[1]

When attaching a face in an interface, the face should not be "the icing on the cake." The face should provide useful information which helps us understand the content better. For example, weather forecast on TV uses "personified weather", such as an image of a person who has an umbrella when it rains or a image of a person with sunglasses when it is sunny. There is a translation process between seeing a image and knowing the content of the image. As Norman says [Norman, D. 86], good interfaces should make this translation process easier and encourage users to develop a mental model of the application that is similar to the model of the task domain. This is also true in agent-based interface. Lieberman [Lieberman 96a] says, an agent-based interface must support forming a user's corresponding mental model of the application to that of the agent.

1. However, the author feels comfortable with seeing a caricature face in an ATM. In Japan, a female caricature face in an ATM vows, smiles, and advertises while users are waiting for a transaction to finish. However, in the states, people who have been using non-personified ATMs for more than 30 years, don't expect to see a face while interacting with an ATM (through discussions and observations). We should also consider cultural differences when applying a personified interface.

Need for Understanding the Users

The third finding is that there is a dichotomy between user groups which have opposite opinions about personification. Differences in facial features--such as character's gender, humanity, expressiveness--cause opposite evaluations to these two subject groups. For example, the Dog's face is preferred by those who are in favor of personification, while the Human face is preferred by those who are against. Another example is that those who are in favor of personification attributed more intelligence to the expressive face, while those who are against thought the stoic face has higher level of intelligence. Some of the experiments performed indicate that there is also a potential difference in evaluation of a human face and a non-human face between the subjects' gender. The answer for the question--"what kind of facial features makes the agent look intelligent, likable, and comfortable to be with?"--is not the only one. We need to consider the target users when designing a personified interface. The future personified interface should be flexible so that it can provide options to choose a preferred face for each user.

Toward Making More Efficient Personified Interface

Though the focus of computer graphics work on faces is on making more realistic faces [Essa 95, Lee 95, Takeuchi 93], realism might not be the main issue to affect people's evaluation about the face in making a personified interface. This study shows that realism does not affect people's evaluation of faces. Thorisson's humanoid agent [Thorisson 96] at MIT Media Lab has a caricature face with smooth animation. Thorisson argues that realistic syn-

thesized faces look abnormal and repulsive, and a caricature face represents an attractive alternative to a synthesized realistic face for animating software agents.

Instead of realism, smooth animation and audio feedback would make a personified interface more natural and engaging. Gonzalez's study [Gonzalez 96] suggests that smooth animation makes people's decision making more accurate than abrupt animation. Elliott [Elliott 94a] suggests music is more effective in conveying the emotions of an agent than facial expressions. Thorisson [Thorisson 96] argues the importance of giving a computer some command of human multimodal facilities--speech, gesture, facial gesture, and gaze--to enable natural and realistic face-to-face communication with a user. He also suggests that such a computer with a multimodal interface should explicitly be modeled as an interactive agent. The poker players would look more natural if they moved their heads, eyes, and make utterances at the right time. There is a need for further study to understand how effective animation and multimodal interaction would function in an personified interface.

Future Directions

This study raises many fundamental questions about the future of agent-based interface. As stated above, there is a need to evaluate the effects of personification in a different application domain, such as information filtering agents or decision support agents. One good test application might be "Kasbah"--selling and buying agents [Chavez 96] developed by the Autonomous Agent Group at the MIT Media Laboratory. This application requires an agent with a more complex decision making process. Currently Kasbah agents do not have a personified interface. A selling agent could report to a user when it succeeds in selling a car at a

desired price with a satisfied expression. While a buying agent could ask a user what to do with a confused expression when it cannot find an affordable car to buy.

Subjects' computer expertise, age range were not taken into consideration when the data was analyzed in this study. To make a flexible interface, we should consider user diversity other than gender and opinion about personification.

APPENDIX 1

Facial Expressions of the Poker Playing Agents

Figures 40 to 45 show 10 facial expressions of the six poker playing agents--the Realistic Male, the Realistic Female, the Caricature Female, the Caricature Male, the Dog, and Smiley. The order of expressions is neutral, pleased, displeased, satisfied, disappointed, surprised, relieved, excited, anxious, very excited.

The order of expressions is neutral, pleased, displeased, satisfied, disappointed, surprised, relieved, excited, anxious, very excited

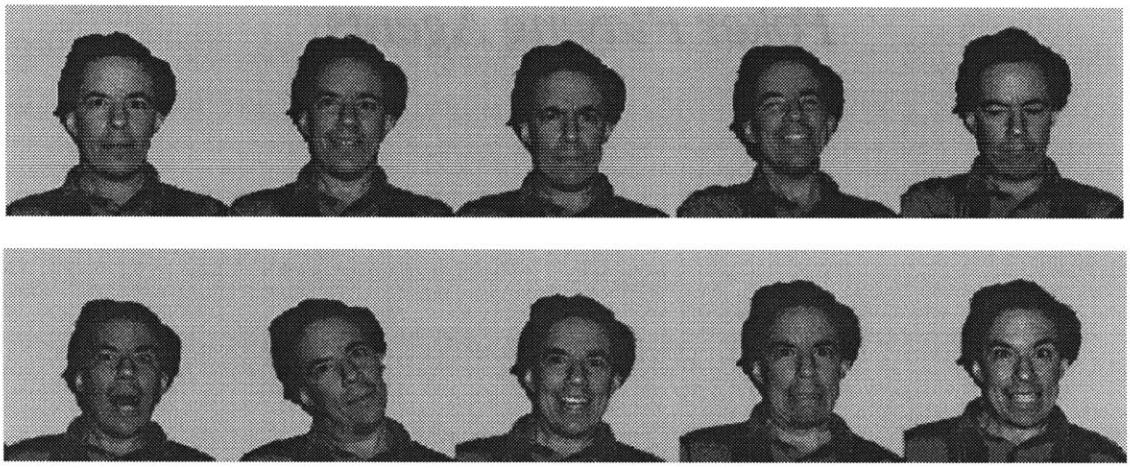

FIGURE 40. The Realistic Male's facial expressions

FIGURE 41. The Realistic Female's facial expressions

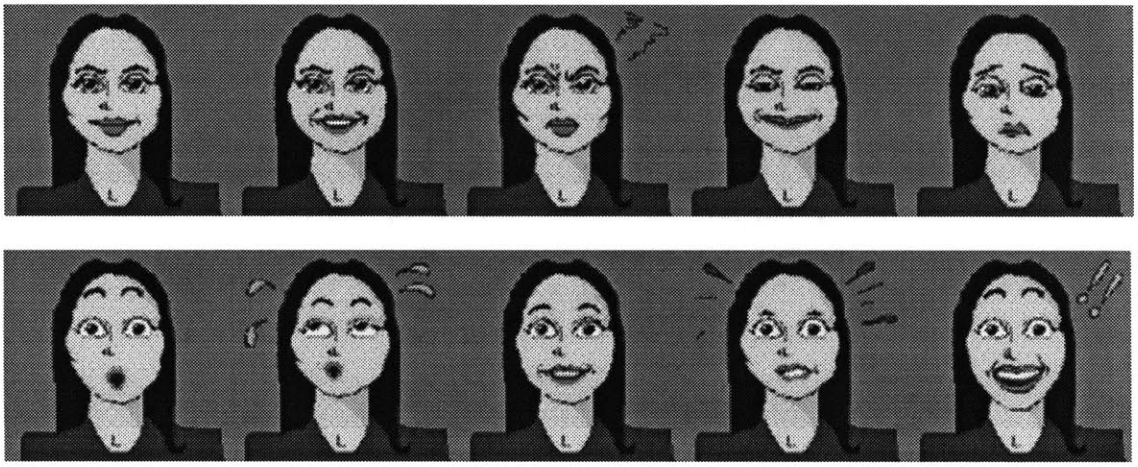

FIGURE 42. The Caricature female's facial expressions

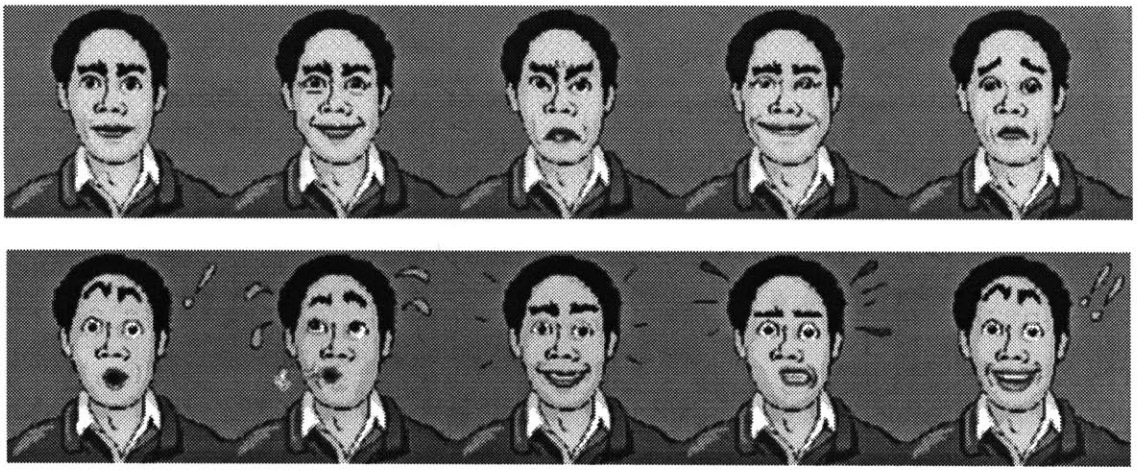

FIGURE 43. The Caricature male's facial expressions

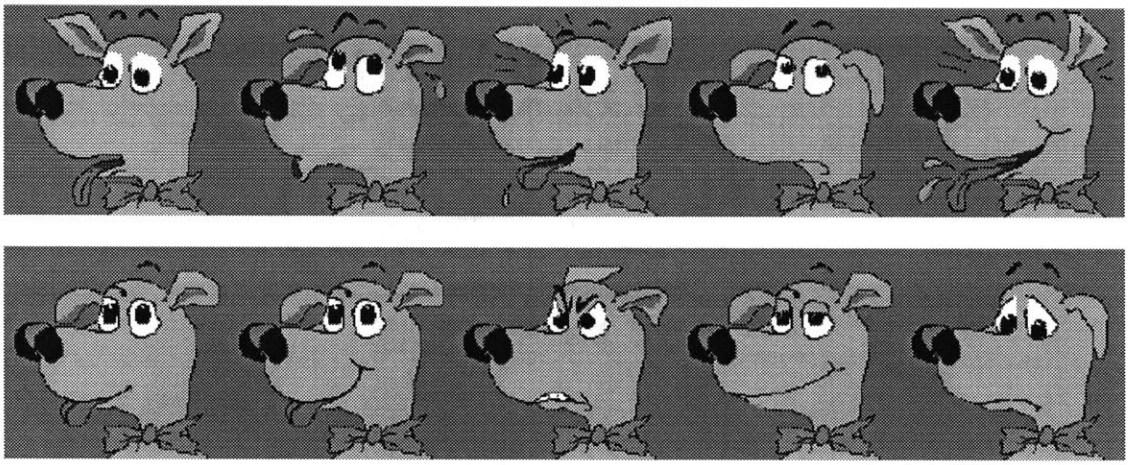

FIGURE 44. **The Dog's facial expressions**

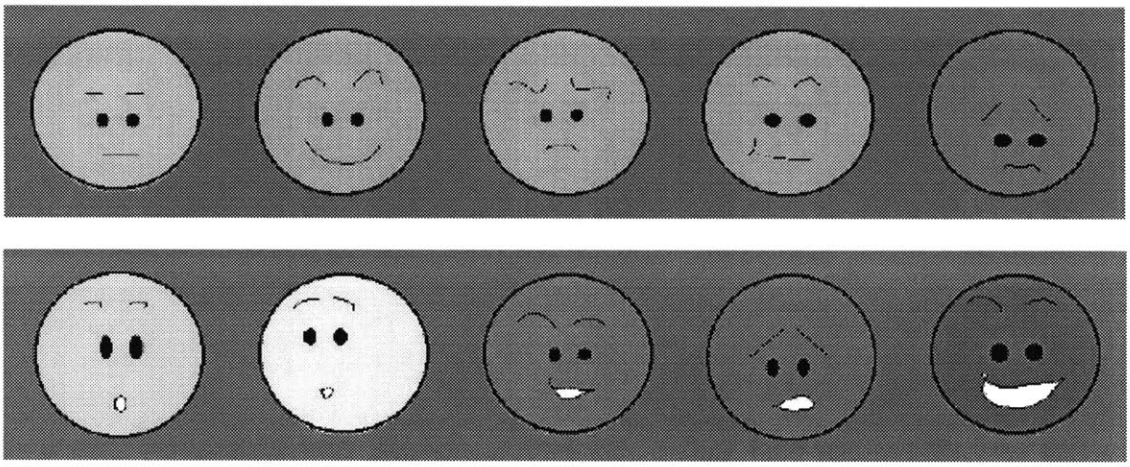

FIGURE 45. **Smiley's facial expressions**

APPENDIX 2

Poker Game Instructions

Below I reproduced the instructions and poker playing rules presented to subjects when they accessed the web site.

Consent Form

Purpose of the Project

The purpose of this project is to study the effect of software agents* with faces-how facial expressions and characteristics can be used in an effective way.

* Software agents: "Software agents" provide active, personalized assistance to a person engaged in the use of a particular computer application.

Procedure of the Experiment

The experiment contains two parts: a poker game and a questionnaire. The game part is done by playing poker against poker playing agents for 15 rounds. It takes approximately 10-15 minutes. After playing poker, you will be asked to complete a questionnaire which takes 1-2 minutes. The questionnaire will ask you to evaluate each poker player's face.

After submitting your answers, you can continue another experiment which also consists of another poker game and questionnaire. Poker playing agents differ in every experiment.

Answering a questionnaire is voluntary, but greatly appreciated. Your activities while playing poker and your answers for a questionnaire are stored in a secured disk, and will be used only in the conduct of this research project or related projects.

Questions can be emailed to the Autonomous Agents Group, the MIT Media Laboratory by clicking HERE.

Consent Declaration

I have read the description of the experiment and I fully understand that my participation is voluntary. I am free to withdraw my consent and discontinue participation at any time without prejudice myself. I understand the procedure and the propose of the experiment. I also understand that I will not be paid for taking part in this experiment.

I give permission to MIT and its research personnel to log my on-line activities while participating in an experiment and to use the logs for the purpose of research. I understand that I

may withdraw this consent at any time, and that by my participation in this study I am not waiving any of my legal rights. I understand that I may also contact the Chairman of the Committee on the Use of Humans as Experimental Subjects, M.I.T. 617-253-6787, if I feel that I have been treated unfairly as a subject.

Poker Playing Rules

Procedure

Media Lab E-Poker is for experimental (and entertainment) purposes only. You can play against different players every time you participate in an experiment. An experiment consists of two parts: a poker game and a questionnaire. Please play poker with the same players for 15 rounds, then answer a questionnaire. After submitting your answers, you can play poker with other players.

The rule is based on five draw poker. Since the purpose is not to make a perfect poker game, the rules of this game might not correspond to the traditional poker rules. At the beginning of the game you will be given 200 dollars in cyber money. There is no real money involved.

A round consists of ante, initial bets, discarding cards, and final bets. The ante cost is 1. You're out if you can't meet this ante.

Initial bets can all be zero. (i.e. no one bets anything). Please type the amount of money to raise in the entry field next to the RAISE button when you raise. There is a betting limit in order not to lose too much money. A message appears when you try to raise over the limit.

Bets and discarding cards are done in order of slot position. A selection can't be undone. In Final bets, if no one makes a bet (i.e. everyone checks), then the round ends. During the game, folding can be done anytime except during discarding.

Poker Hand Ranking (high to low)

- Straight Flush - five cards in one suit in consecutive numerical order.
  Example: 3-D 4-D 5-D 6-D 7-D
- 4-of-a-kind - four cards of the same numerical value.
  Example: A-D A-H A-C A-S 5-S
- Full House - basically a pair and three of a kind in the same hand.
  Example: 9-D 9-S 9-C 3-D 3-S
- Flush - all five cards in the same suit, but not consecutive numerical order.
  Example: 3-D 5-D 7-D 9-D J-D
- Straight - all five cards in consecutive numerical order, not of same suit.
  Example: 4-D 5-S 6-D 7-C 8-D
- 3-of-a-kind - three cards of the same numerical value.
  Example: 5-D 5-S 5-C 3-D 8-H
- 2 pairs - two cards each of two separate numerical values.
  Example: Q-D Q-H 8-S A-C A-D
- 1 pair - two cards of the same numerical value
  Example: 7-D 7-H 2-S 4-H 6-C
- nothing - high card.

How to Play

You are sitting in front of the table. Your hands are shown in front of the table. Opponent player's actions and a message to you appear in the center of the table.

Messages addressed to you are blue. Current pot is shown under the message.

During ante, you decide whether to open or fold by clicking the OPEN or FOLD button.

You can either CALL, RAISE, FOLD, or PASS in each betting round.

When you raise, enter the amount of money you want to raise in the entry field next to the RAISE button, then press RAISE.

To discard, please click the cards you want to discard then press DISCARD.

Press CONTINUE when the "Please press continue" message appears or you want to start another game.

Your current money is shown under your hands. (Don't worry, there is no penalty if you lose all your money.)

After playing 15 rounds, please press QUESTIONNAIRE. A new browser window for a questionnaire appears. After submitting your answer, you can start a new game with different players by pressing CONTINUE button on the poker game window.

If you want to quit, you can do so by moving to another web site.

Warnings

If your browser does not support Java then you will not be able to play. You have to have HotJava browser or Netscape 2.0 or higher, and enable Java support.

Your browser must also allow a socket connection to be opened back to the host from which the applet was loaded, i.e.the server you are now accessing.

Currently, the poker game does not support Macs and PC with Linux.

While you are playing poker (including the login screen), please don't resize the browser window horizontally. This is a limitation for Netscape users.

Dowloading an applet takes up to seven minutes (it depends on where you are). Please be patient during downloading -please don't click the START GAME button again and again even if it looks like it is not responding.

If your display's resolution is not high enough (i.e. 640x480), you may not be able to see the whole poker table. In this case, you don't have to answer a questionnaire. Just enjoy playing poker with not-totally-visible players :)

APPENDIX 3

Questionnaires

This section shows the questionnaires for the pilot experiment and the web-based experiments. The original questionnaires on the web have HTML format. Selection lists are shown as drop down lists. Each player has a nickname in the poker game as follows:

- NoFace: Invisible Man
- Dog: Esme
- Smiley: Smiley
- Caricature Male: Morgan
- Caricature Female: Morticia (Blue: Honest, Red: Deceiving, Green: Stoic)
- Realistic Male: Bruce
- Realistic Female: Pattie

Common Questionnaire for the Web-based Experiments

Thank you very much for visiting Media Lab Poker Room, *user name*. I hope you had fun.

Please take a couple of minutes to complete this questionnaire. This information will not be used for any other than research purpose, and will be kept confidential. After completing the questionnaire, please click "SUBMIT" button. You can continue playing poker after submitting your questionnaire.

First, please tell us about yourself.

Your gender male female

Your age -10 11-20 21-30 31-40 41- 50 51- 60 61-

Your computer experience beginner intermediate advanced

Suppose you have a computer program which works on your behalf. For example, you have an email filtering agent which sorts your incoming messages and tells you which you should read first and which you don't have to read. Do you think it a good idea to show the agent's face on the screen?

Yes, I want to have a face of my agent on the screen.

No, I don't want to have a face of my agent on the screen.

The purpose of the questionnaire is to evaluate your subjective impression of each face shown while playing poker, not to evaluate your performance. Please rate your agreement with the following statements, comparing the players you just played poker against. The scale is 1 to 7, in which 1 is the least and 7 is the best. Please give your instantaneous response. Your evaluations don't have to be consistent.

Please tell us more comments on this experiment. Give us your comments on how to improve the interface on the poker game, what kind of face do you want to see as your opponent player? *(A blank text area for comments follows.)*

Specific Questionnaires for the Web-based Experiments

Each selection list has 7 scales: 1-Strongly disagree; 2-Disagree; 3-Slightly disagree; 4-Neutral; 5-Slightly agree; 6-Agree; 7-Strongly agree.

FACE experiment

1. The player was good at playing poker.

    Morticia
    Invisible Man

2. You liked playing poker with the player.

    Morticia
    Invisible Man

3. Playing poker with the player was engaging.

   Morticia
   Invisible Man

4. You want to continue playing poker with the same player.

   Morticia
   Invisible Man

5. You paid attention to the player's face. *(1-7 scale)*

6. You were distracted by the player's face. *(1-7 scale)*

GENDER experiment

1. The player was good at playing poker.

   Bruce(Male)
   Pattie(Female)

2. You liked playing poker with the player.

   Bruce(Male)
   Pattie(Female)

3. Playing poker with the player was engaging.

   Bruce(Male)
   Pattie(Female)

4. You want to continue playing poker with the same player.

Bruce(Male)
Pattie(Female)

HUMANITY experiment

1. The player was good at playing poker.

Morgan(Male)
Esme(Dog)

2. You liked playing poker with the player.

Morgan(Male)
Esme(Dog)

3. Playing poker with the player was engaging.

Morgan(Male)
Esme(Dog)

4. You want to continue playing poker with the same player.

Morgan(Male)
Esme(Dog)

5. The face corresponded to the player's true level of intelligence.

Morgan(Male)
Esme(Dog)

REALISM experiment

1. The player was good at playing poker.

   Smiley
   Morgan
   Bruce

2. You liked playing poker with the player.

   Smiley
   Morgan
   Bruce

3. Playing poker with the player was engaging.

   Smiley
   Morgan
   Bruce

4. You want to continue playing poker with the same player.

   Smiley
   Morgan
   Bruce

5. The face corresponded to the player's true level of intelligence.

   Smiley
   Morgan
   Bruce

EXPRESSIVENESS experiment

1. The player was good at playing poker.

    Blue Morticia
    Red Morticia
    Green Morticia

2. You liked playing poker with the player.

    Blue Morticia
    Red Morticia
    Green Morticia

3. Playing poker with the player was engaging.

    Blue Morticia
    Red Morticia
    Green Morticia

4. You want to continue playing poker with the same player.

    Blue Morticia
    Red Morticia
    Green Morticia

5. You believed the player was honest in its facial expressions about its poker hands.

    Blue Morticia
    Red Morticia
    Green Morticia

6. You paid attention to the player's facial expressions.

Blue Morticia
Red Morticia
Green Morticia

7. You were distracted by the player's facial expressions.

Blue Morticia
Red Morticia
Green Morticia

8. The facial expressions of the player helped you to determine your strategy.

Blue Morticia
Red Morticia
Green Morticia

Questionnaire for the Pilot Experiments

Common Introduction

The purpose of the questionnaire is to evaluate your impression of each face shown below. Please rate your agreement with the following statements, comparing the two or three faces. The scale is 1 to 7, in which 1 is the least and 7 is the best. Please give your instantaneous response. Your evaluations don't have to be consistent. Imagine that you walk into a poker room in a casino. You see two (three) players sitting in a poker table. If you play poker against those two (three) players shown below, how would you think about each player?

GENDER_LOOK experiment

1. The player looks good at playing poker.

   Male
   Female

2. You want to play poker with the player.

   Male
   Female

3. Playing poker with the player will be engaging.

   Male
   Female

HUMANITY_LOOK experiment

1. The player looks good at playing poker.

   Male
   Dog

2. You want to play poker with the player.

   Male
   Dog

3. Playing poker with the player will be engaging.

   Male
   Dog

REALISM_LOOK experiment

1. The player looks good at playing poker.

    Smiley
    Caricature
    Realistic

2. You want to play poker with the player.

    Smiley
    Caricature
    Realistic

3. Playing poker with the player will be engaging.

    Smiley
    Caricature
    Realistic

References

[Bates 94] Bates, Joseph. "The Role of Emotion in Believable Agents." Communications of the ACM. Vol. 37, No. 7 (July, 1994). 122-125.

[Bell 96] Bell, Gene *et al.* "Lifelike Computer Characters: the Persona project at Microsoft Research." in Software Agents, edited by Jeff Bradshaw. Cambridge, MA: MIT Press, 1996.

[Blumberg 94] Blumberg, Bruce. "Action Selection in Hamsterdam: Lessons from Ethology." in Proceedings of the Third International Conference on the Simulation of Adaptive Behavior, 1994.

[Brennan 94] Brennan, Susan. *et al.* "Effect of Message Style on Users' Attributions toward Agents," in CHI '94 Conference Companion Human factors in Computing Systems, Reading, MA: Addison-Wesley. 281-282.

[Chavez 96] Chavez, Anthony *et al.* "Kasbah: An Agent Marketplace for Buying and Selling Goods." in Proceeding in The Practical Application of Intelligent Agents and Multi-Agent Technology, The Practical Application Company Ltd. 1996. 75-90.

[Chernoff 73] Chernoff, Herman. "The Use of Faces to Represent Points in k-Dimensional Space and Graphically." Journal of the American Statistical Association, Vol. 68, No. 342, (June, 1973). 361-368.

[Don 92] Don, Abbe *et al.* "Anthropomorphism: From Eliza to Terminator 2." in CHI '92 Conference Proceedings of Human Factors in Computing Design. Reading. MA: Addison-Wesley, 1992. 67-69.

[Elliot 94a] Elliot, Clark. "Affective Reasoner Project." available at http://condor.depaul.edu/~elliott/arback.html

[Elliott 94b] Elliott, Clark. "Multi-Media Communication with Emotion-driven 'Believable Agents.'" In Working Notes for the AAAI Spring Symposium on Believable Agents. AAAI, American Association for Artificial Intelligence. 16-20.

[Elliott 95] Elliott, Clark *et al.* "Getting to the Point: Emotion as a Necessary and Sufficient Element of Story Construction." In Working Notes for the AAAI Spring Symposium on Interactive story systems: Plot and character, 1995. AAAI, American Association for Artificial Intelligence. 37-40.

[Essa 95] Essa, Irfan. "Analysis, Interpretation and Synthesis of Facial Expressions." Ph.D. Thesis, Media Arts and Sciences, Massachusetts Institute of Technology. MIT Media Laboratory Perceptual Computing Section Technical Report No.256.

[Goldberg 84] Goldberg, A. "Smalltalk-80: The Interactive Programming Environment." Reading, MA: Addison-Wesley, 1984.

[Gonzalez 96] Gonzalez, Cleotilde. "Does Animation in User Interface Improve Decision making?" in CHI '96 Conference Proceedings in Human Factors in Computing Systems. Reading, MA: Addison-Wesley, 1996. 27-34.

[Isbister 95] Isbister, Katherine. "Perceived Intelligence and the Design on Computer Characters," http://www.cyborganic.com/People/kath/lccpaper.html

[King 96] King, William Joseph *et al.* "The Representation of Agents: Anthropomorphism, Agency, and Intelligence." in CHI '96 Conference Companion of Human Factors in Computing Systems. Reading, MA: Addison-Wesley, 1996. 289-290.

[Kozierok 93] Kozierok, Robyn. "A Learning Approach to Knowledge Acquisition for Intelligent Interface Agents." SM Thesis, Department of Electrical Engineering and Computer Science. MIT. May, 1993.

[Lanier 95] Lanier, Jaron. "Agents of Alienation." Interactions. July, 1995. 66-72.

[Lashkari 94] Lashkari, Y., Metral, M., and Maes, P. "Collaborative Interface Agents." In Proceedings of the National Conference on Artificial Intelligence. Cambridge, MA.: MIT Press, 1994

[Laurel 90] Laurel, Brenda. "Interface Agents: Metaphors with Character." In The Art of Human-computer Interface Design, edited by Brenda Laurel. Reading, MA: Addison-Wesley, 1990. 355-365.

[Lee 95] Lee, Yuencheng *et al.* "Realistic Modeling for Facial Animation." In SIGGRAPH 95 Conference Proceedings on Computer Graphics Annual Conference Series, 1995. 35-62.

[Lieberman 96a] Lieberman, Henry. "Attaching Interface Agent Software to Applications." Available as <http://lieber.www.media.mit.edu/people/lieber/Lieberary/Attaching/Attaching/Attaching.html>.

[Lieberman 96b] Lieberman, Henry *et al.* "Instructible Agents: Software that Just Keeps Getting Better." IBM Journal, Vol. 35, No. 3-4, 1996.

[Maes 94] Maes, Pattie. "Agents That Reduce Work and Information Overload". In Communications of the ACM. Vol. 37, No. 7 (July, 1994). 31-40,146.

[Maes 95a] Maes, Pattie. "Intelligent Software." Scientific American. September, 1995. 84-86.

[Maes 95b] Maes, Pattie. "Artificial Life Meets Entertainment: Lifelike Autonomous Agents." in Communications of the ACM, Vol. 38, No. 11 (Nov, 1995).

[Nass 94] Nass, Clifford, *et al.* "Computers as Social Actors." In CHI '94 Conference Proceedings of Human Factors in Computing Systems. Reading, MA: Addison-Wesley, 1994. 72-77.

[Nass 95] Nass, Clifford, *et al.* "Can Computer Personalities Be Human Personalities?" In CHI '95 Conference Proceedings of Human Factors in Computing Systems. Reading, MA: Addison-Wesley, 1995. 228-229.

[Negroponte 95] Negroponte, Nicholas. Being Digital. 101-102. The Alfred A. Knorpf, Inc. 1995.

[Norman, S. 86] Norman, Streiner. "PDQ Statistics." Philadelphia, PA: B.C. Decker Inc., 1986.

[Norman, D. 86] Norman, Donald. "Cognitive Engineering." in User Centered System Design, edited by Donald Norman *et al.* Hillside, NJ: Lawrence Erlbaum Associates. 1986

[Norman, D. 88] Norman, Donald. "The Design of Everyday Things." New York, NY: Doublesday. 1988.

[Oren 90] Oren, Tim. *et al.* "Guides: Characterizing the Interface." In The Art of Human-computer Interface Design, edited by Brenda Laurel. Reading, MA: Addison-Wesley, 1990. 367-381.

[Ortony 88] Ortony, Andrew *et al.* "The Cognitive Structure of Emotions." Cambridge, MA: Cambridge University Press. 1988

[Picard 95] Picard, Rosalind. "Affective Computing." M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 321. 1995.

[Sheth 94] Sheth, Beerud. "A Learning Approach to Personal Information Filtering" Master's Thesis. MIT Media Lab, 1994.

[Shneiderman 95] Shneiderman, Ben. "Looking for the Bright Side of User Interface Agents." Interactions. January, 1995. 13-15.

[Takeuchi 93] Takeuchi, Akikazu *et al.* "Communicative Facial Displays as a New Conversational Modality." in InterCHI '93 Conference Proceedings in Human Factors in Computing Systems, 1993. Reading, MA: Addison-Wesley. 187-193.

[Takeuchi 95] Takeuchi, Akikazu. *et al.* "Situated Facial Displays: Towards Social Interaction," in CHI '94 Conference Proceedings Human Factors in Computing Systems, Reading, MA: Addison-Wesley. 450-454.

[Thorisson 96] Thorisson, Kris. "Communicative Humanoids: A Computational Model of Psychosocial Dialogue Skills." Ph.D. Thesis, Program in Media Arts and Sciences, MIT. July, 1996.

[Walker 94] Walker, Janet. *et al.* "Using a Human Face in an Interface," in CHI '94 Conference Proceedings Human Factors in Computing Systems, Reading, MA: Addison-Wesley. 85-91.