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.

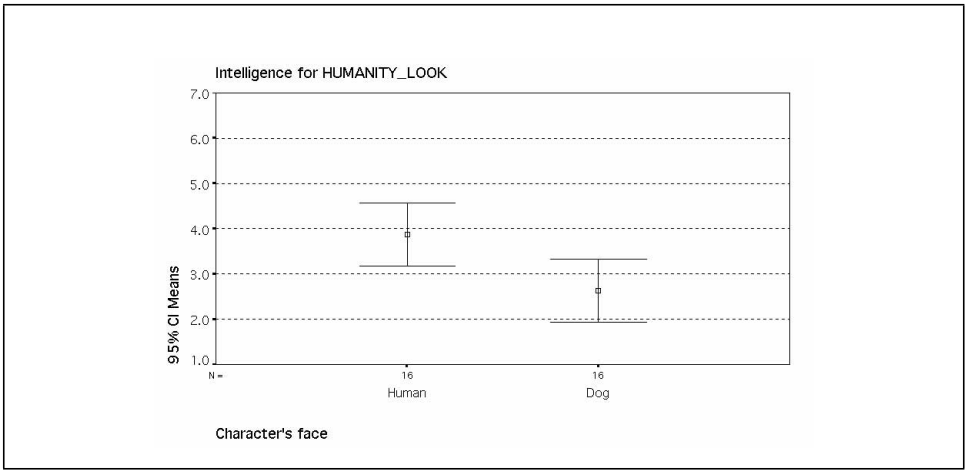

FIGURE 18. The mean value of INT for the Human and Dog's face based on appearance.

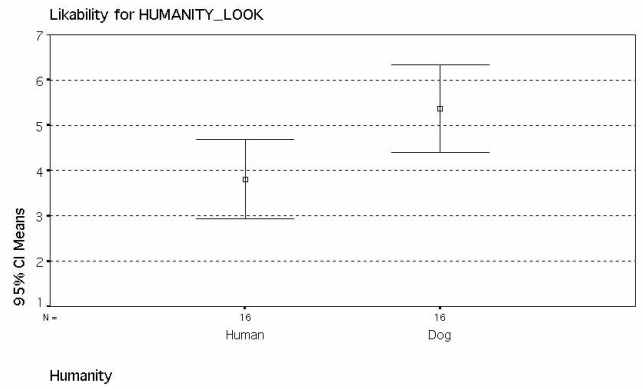

FIGURE 19. The mean value of LIKE for the Human and Dog's face based on appearance.

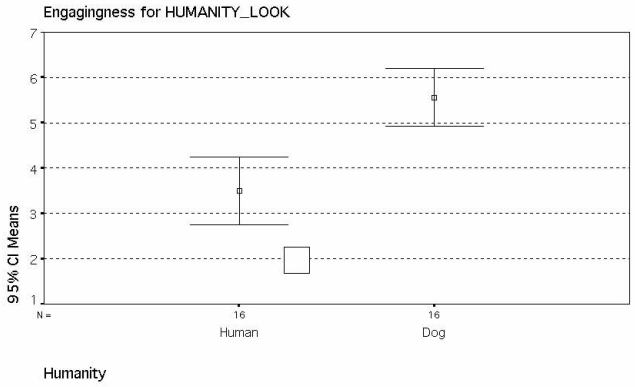

FIGURE 20. The mean value of ENG for the Human and Dog's face based on appearance.

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.

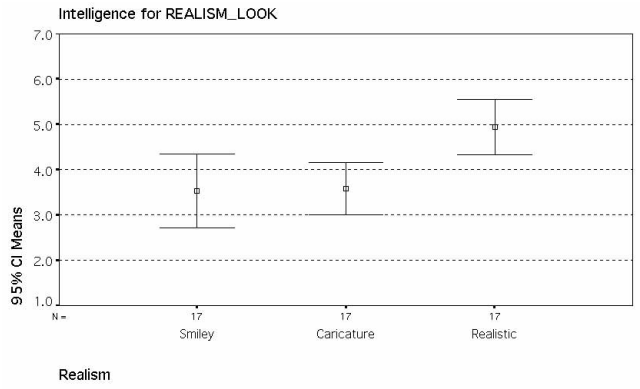

FIGURE 21. The mean value of INT for Smiley, Caricature, and Realistic face based on appearance.

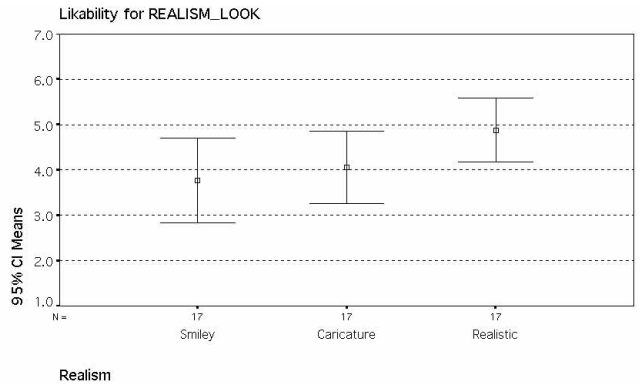

FIGURE 22. The mean value of LIKE for Smiley, Caricature, and Realistic face based on appearance.

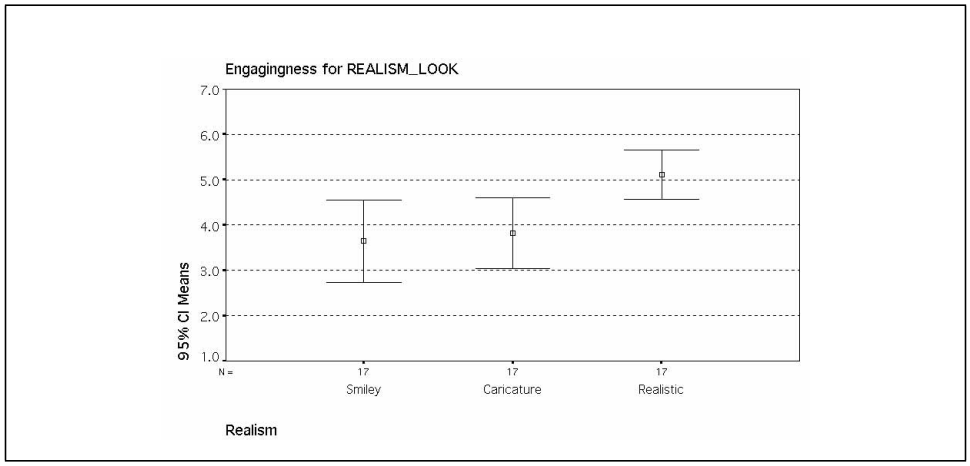

FIGURE 23. The mean value of ENG for Smiley, Caricature, and Realistic face based on appearance.

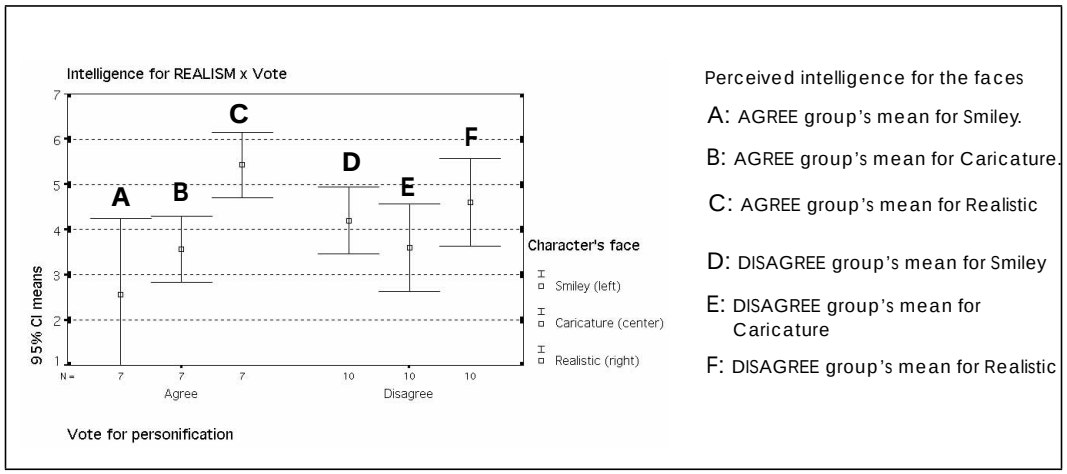

FIGURE 24. The mean value of INT for Smiley, Caricature, and Realistic face based on appearance, for the AGREE and DISAGREE group.

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	-	-	-

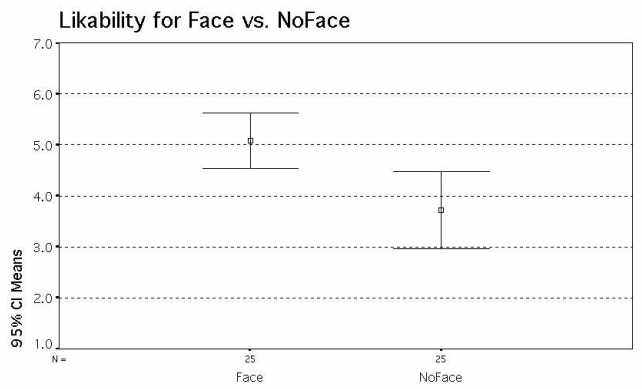

FIGURE 25. The mean value of LIKE for Face and NoFace.

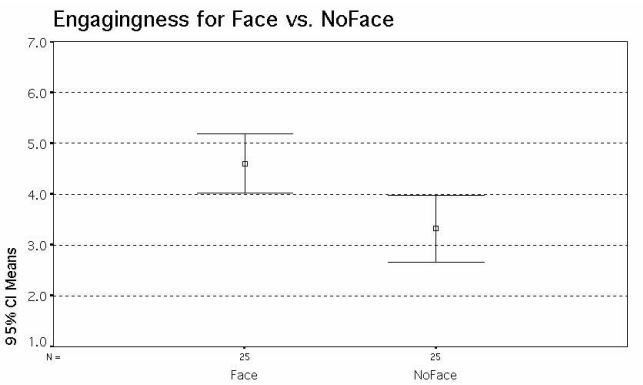

FIGURE 26. The mean value of ENG for Face and NoFace.

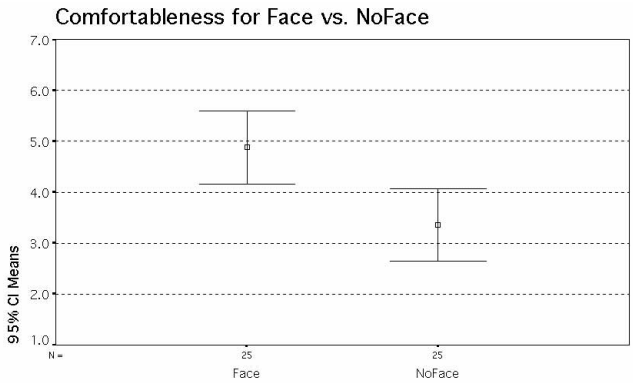

FIGURE 27. The mean value of COM for Face and NoFace.

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.

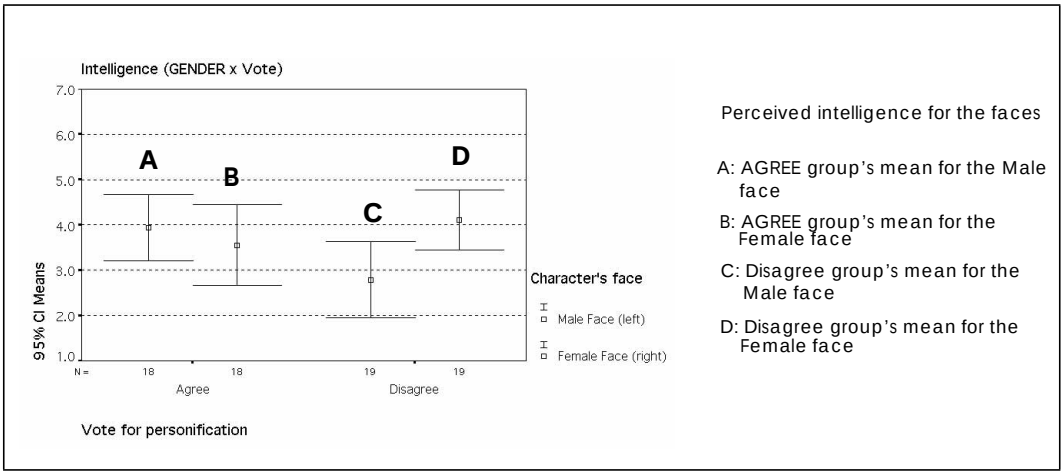

FIGURE 28. The mean value of INT for the Male and Female face, for both the AGREE and DISAGREE groups.

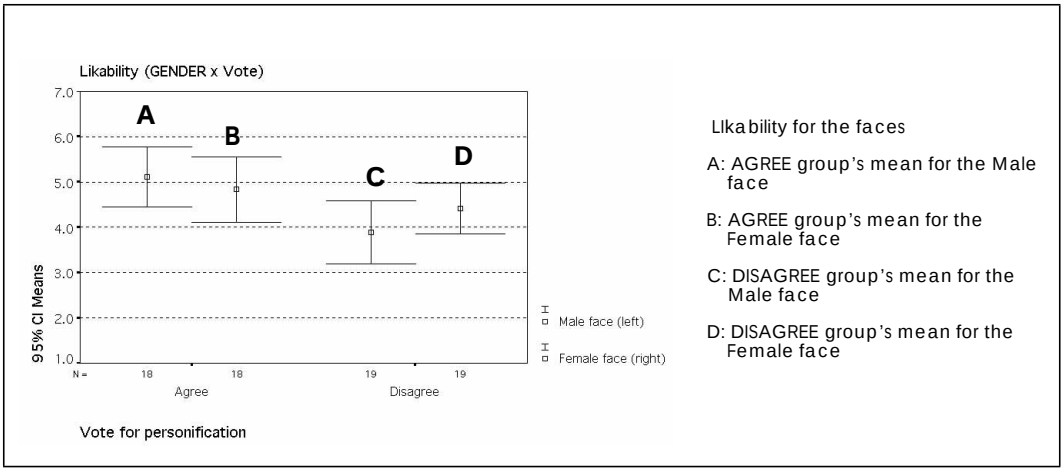

FIGURE 29. The mean value of LIKE for the Male and Female face, for both the AGREE and DISAGREE groups.

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		

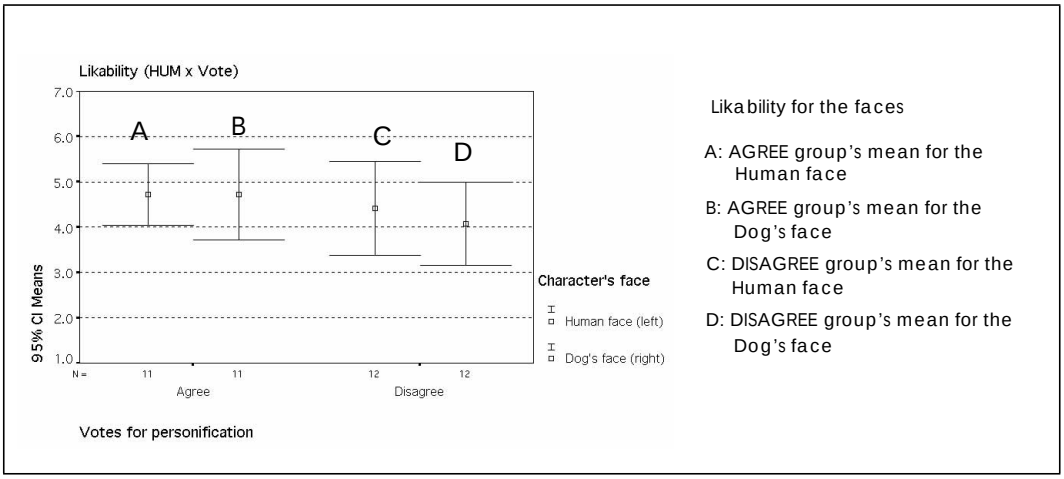

FIGURE 30. The mean value of LIKE for the Human and Dog's face, for both the AGREE and DISAGREE groups.

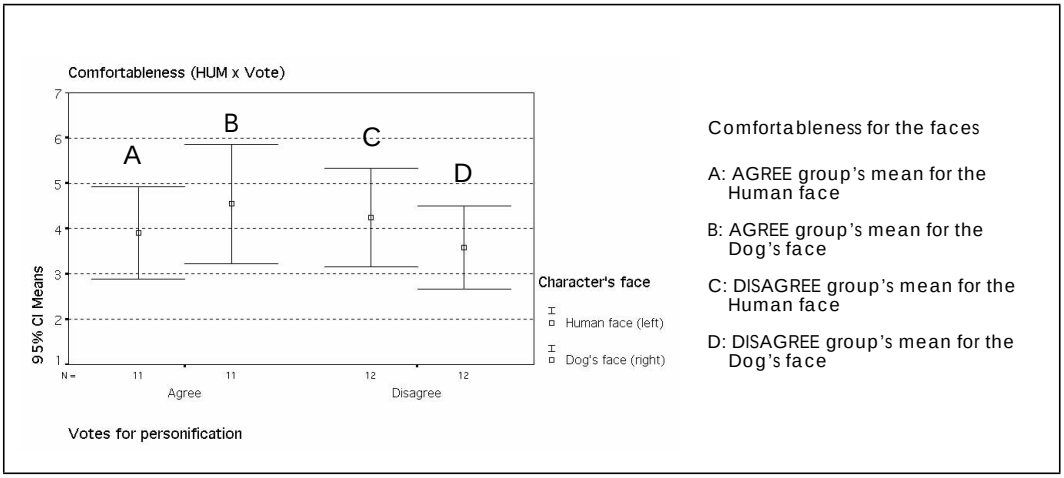

FIGURE 31. The mean value of COM for the Human and Dog's face, for both the AGREE and DISAGREE groups.

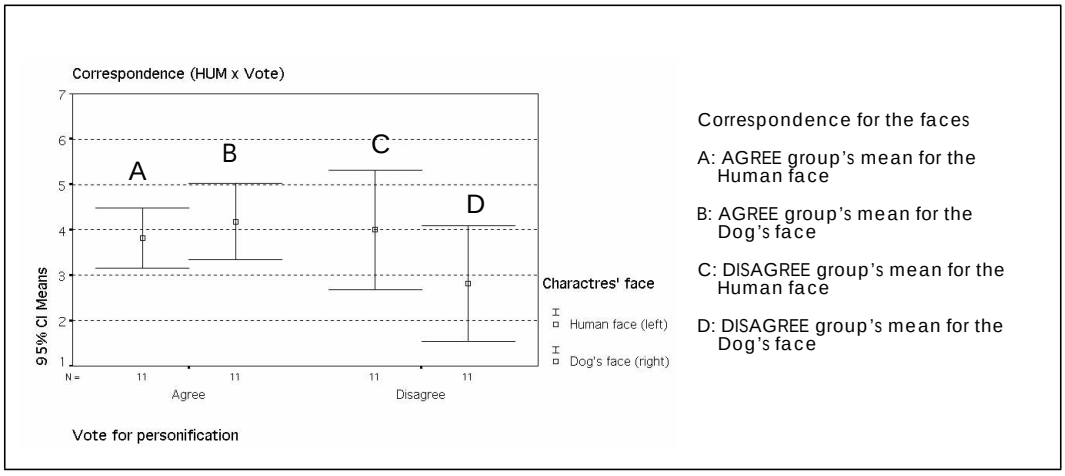

FIGURE 32. The mean value of COR for the Human and Dog's face, for both the AGREE and DISAGREE groups.

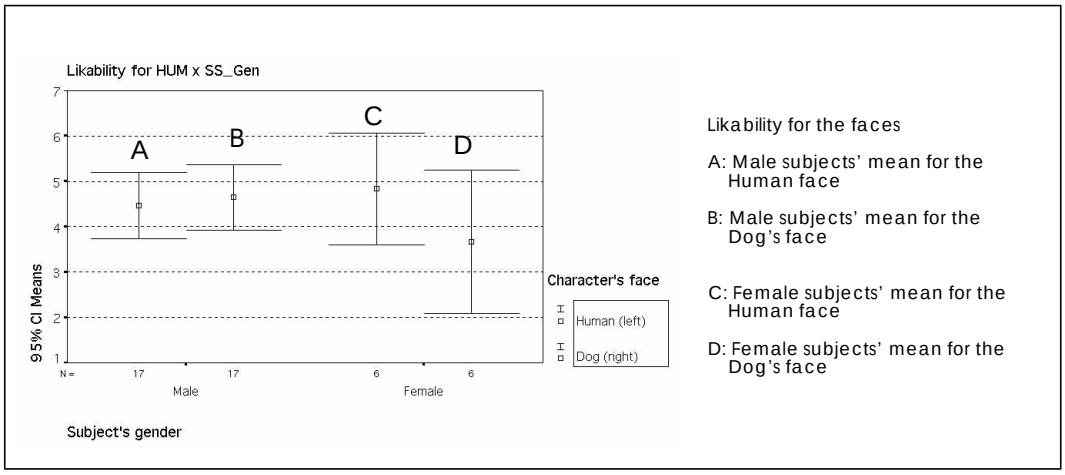

FIGURE 33. The mean value of LIKE for the Human and Dog's face, for both male and female groups.

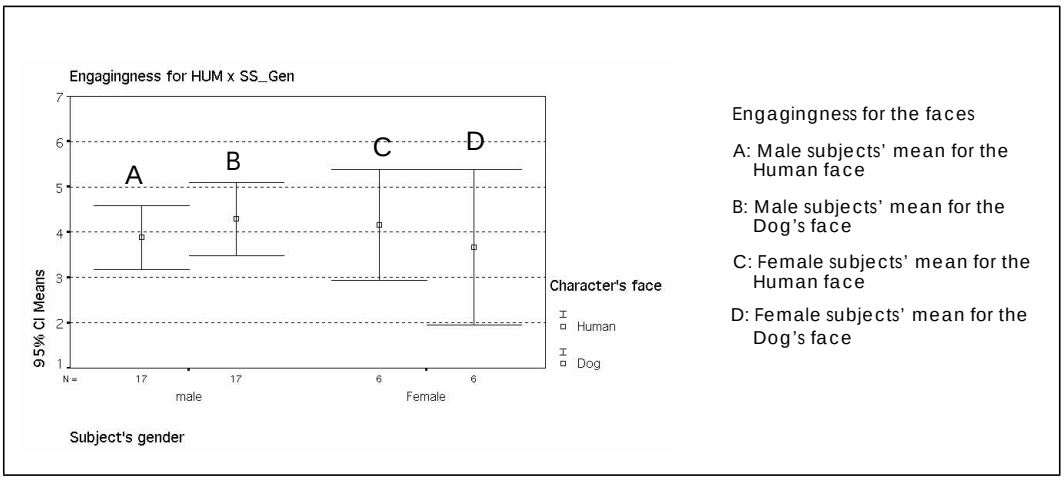

FIGURE 34. The mean value of ENG for the Human and Dog's face, for both male and female groups.

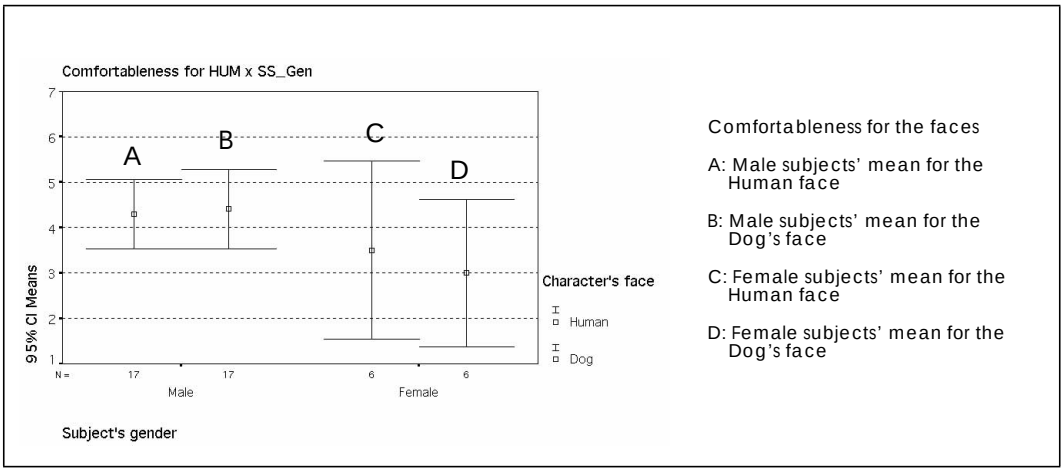

FIGURE 35. The mean value of COM for the Human and Dog's face, for both male and female groups.

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

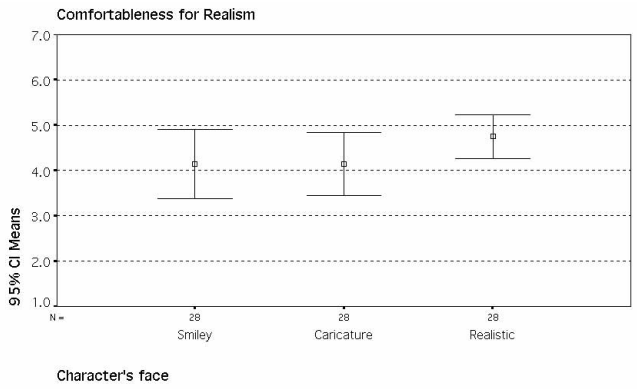

FIGURE 36. The mean value of COM for Smiley, the Caricature, and the Realistic face.

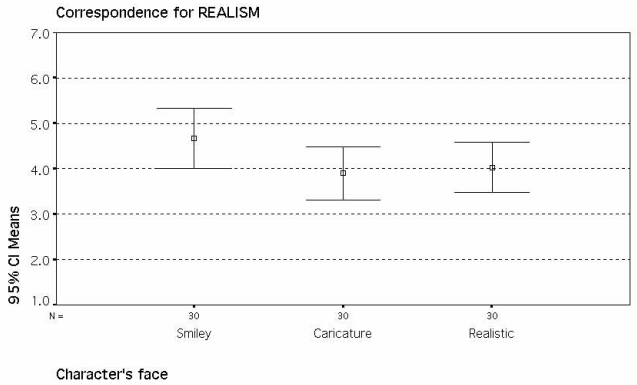

FIGURE 37. The mean value of COR for Smiley, the Caricature, and the Realistic face.

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.

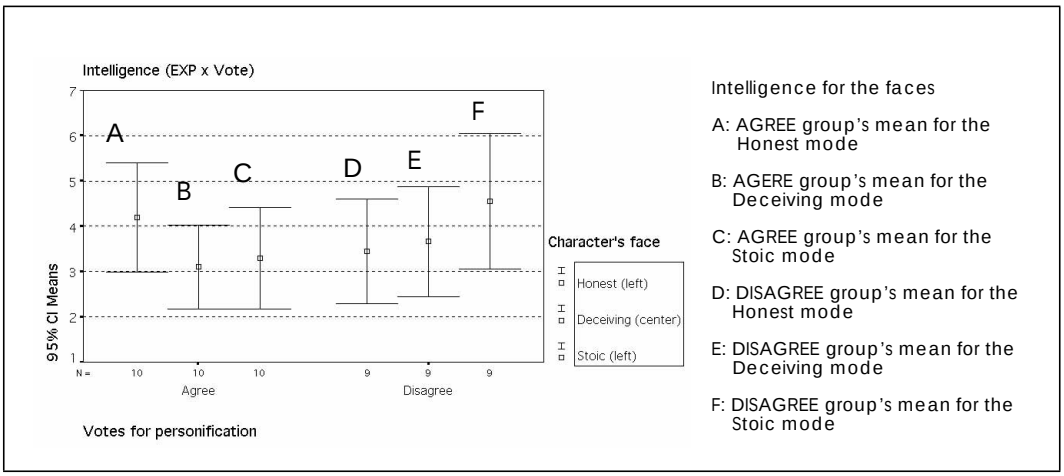

FIGURE 38. The mean value of INT for the Honest, Deceiving, and Stoic face, for both the AGREE and DISAGREE groups.

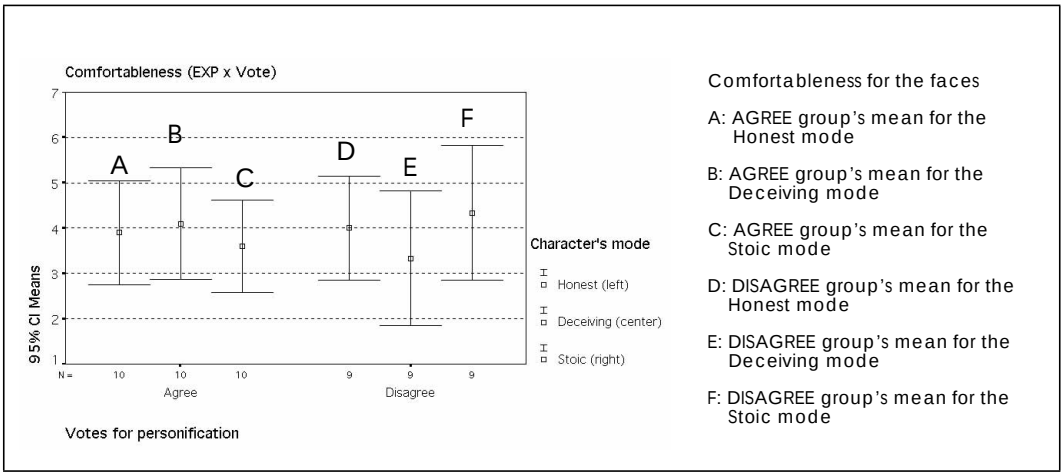

FIGURE 39. The mean value of COM for the Honest, Deceiving, and Stoic face, for both the AGREE and DISAGREE group.

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.