WHAT DO YOU MEAN BY BELIEVABLE CHARACTERS? : THE EFFECT OF THE

CHARACTER NEWNESS, RATING AND EVILNESS ON THE PERCEPTION OF

THE CHARACTER BELIEVABILITY

Running title: Character Believability Studies

ABSTRACT

Computer scientists working on Artificial Intelligence have recognized the importance of believable characters. Studies on the character believability have been relatively rare, and the definition of believability has different meaning in different disciplines. A multi-disciplinary literature review explores various qualities of computer characters. Different scholars and practitioners have described five believability attributes.

For the test of those five attributes, eight NPCs were chosen for the study based on game quality (high and low Metacritic scores), modernity (games released before and after 2006) and game function (good guy or bad guy). The believability of each character was surveyed with an online survey by research subjects recruited from a sophomore level introductory digital media course at a large Midwestern university. Within the survey, participants were shown a one to two minute video of each NPC interacting with a player; participants then answered believability questions about the NPC they had just seen. Role-play gaming experience and demographics were also measured. ANOVA analysis of the results suggested that NPCs from newer and/or high rated games were perceived to be more believable than characters from older or lower rated games. NPCs from newer and/or higher rated games were also found to be more enjoyable than characters from older or lower rated games.

INTRODUCTION

Characters in games are often not plausible at all. For example, the owner of a pizza house in Grand Theft Auto waits to take an order when the player character enters the restaurant. However, the owner shows no personality or emotion when he waits for the player to pay for pizza. He doesn't even try to run away or ask for help if the player beats him to steal his cash. Non Player Character (NPC)s seem to do what they are supposed to do in games, but it doesn't take long to notice that their behaviors are extremely limited. This is the moment that the "willing suspension of disbelief" is broken. The NPCs' discrepancy between their limited behavior patterns, acting like machines, and their outlook, acting like avatars that are controlled by game players, prevent from being believable participants in the game world.

Many researchers in different fields have studied various aspects of computer character believability and related areas: the importance of believable characters in the interactive drama systems (Aylett, 1999; Mateas, 1997; Szilas, 2003), a relationship between the believability of characters and the level of immersion players experienced (Bhatt, 2004; van Doorn & de Vries, 2006; Watson, 2002), linkage to enjoyment (Brown & Cairns, 2004), etc. In many cases, however, the term believability was used without clear definition. The believability studies with social science perspective, however, have been rare.

LITERATURE REVIEW

What does it mean that something is believable? This is the question that researchers have tried to answer in many contexts (Andrews, Netemeyer, & Durvasula,

1991; Beltramini, 1982; Berkos, 2003; Ewing, 1940; Hovland, Janis, & Kelley, 1953).

However, the definition of the term Believability has been made differently based on the

context of the study. The Believability of characters in a game, for example, is not

necessarily related to truthfulness or credibility (Mateas, 1997). Characters are "*artistic*

abstractions of people, whose behavior, motivations, and internal life have been

simplified and exaggerated in just such a way as to engage the audience in the artist's

vision" (Mateas, 2002, p. 8).

Recently, Lee and Heeter (2008) studied the definition and quality of the

believability through out many computer science literatures and examined it with many

theories in both the social science and communication area. Based on their research, they

defined believability as "the size and nature of the cognitive gap between the character

players experience and the character they expect. When the player's expectations exactly

match their experience, a character is fully believable. The larger the gap, the more likely

it is to interfere with suspension of disbelief."

The Lee and Heeter's analysis of believability qualities produced 5 key believability

categories: appearance, personality, goals, emotions, and social relations. Each of these

qualities is expected to contribute to overall general believability. Certain qualities may

be more important to general perceived believability than others.

1. Appearance

The character appearance category describes all the qualities that are exposed to

the human sensory input including visual hints that communicate information such as

gender, age, ethnicity, height, socioeconomic status, etc. When people look at a

character, they make guesses on not only what kind of character it is but also how they can interact with it based on context and the information perceived by sensory input.

In order to create a believable agent, the character appearance must do something more than simply reveal demographic information. Rather, character appearance should suggest other believability qualities such as emotion and personality. This leads to the following hypothesis:

H1: A character's appearance will be closely related to that character's general perceived believability.

2. Personality

Personality defines the uniqueness and peculiar qualities of computer characters that distinguish them from other computer characters. Some qualities of personality are closely related to the psychological traits.

Personality has been one of the most important factors for many computer scientists as well as early cartoon animators. Recently, personality has been suggested as one of the most critical factors in creating believable agents (Allbeck & Badler, 2002; Bates, Loyall, Reilly, Castelfranchi, & Wemer, 1994; Reilly, 1997; Romano & Wong, 2004). This leads to the following hypothesis:

H2: A character's personality will be closely related to that character's general perceived believability.

3. Goals

Goals are one of the most important fundamentals for building believable agents because goals directly affect all other qualities of believability. For instance, the outfit of an agent should be designed carefully with a consideration of its role. If it is a kind of tutoring character that teaches cooking, for example, it may be more natural for a player to expect the agent to wear a high cook's hat and white colored chef's coat. This leads to the following hypothesis:

H3: The goals of a character will be closely related to that character's general perceived believability.

4. Emotions

Emotion is one of the two most important qualities of believable agents (Bates, et al., 1994; Hayes-Roth & Doyle, 1998; Hayes-Roth, Maldonado, & Moraes, 2002; Loyall, 1997; Reilly & Bates, 1995; Romano & Wong, 2004).

The emotion category doesn't prescribe what kinds of emotions believable characters have to possess. Rather, all of the research literature about emotions suggests that it is important for believable agents to clearly exhibit emotions of their own. The emotion category requires believable agents to reveal their emotions as an outcome of unseen internal processes. At the same time, believable agents should respond to players' emotions correctly in a given context. This leads to the following hypothesis:

H4: A character's emotions will be closely related to that character's general perceived believability.

5. Social relations

Many studies on believable agents described social relationships among computer characters that influence an interaction and are influenced by the interaction in turn among characters (Bates, 1994; Mateas, 1997; Thomas & Johnson, 1981). Some studies described a social aspect to the interaction between computer characters and players (Hayes-Roth & Doyle, 1998; Hayes-Roth, et al., 2002).

A character that appears to have a social relationship with other characters helps players willingly suspend their disbelief and conceive of the character as real. The social relationship between a computer agent and the player can also affect the perception of believability. The more users feel a social connection with the computer characters, the more believable the character will seem. This leads to the following hypothesis:

H5: A character's social relations will be closely related to that character's general perceived believability.

RESEARCH QUESTIONS AND HYPOTHESES

The qualities insisted to create believable agents were examined and categorized into five different categories: appearance, personality, goals, emotions, and social relations. However, these believability qualities are not necessary conditions for character believability. Humans require very few cues in order to react socially to a computer (Reeves & Nass, 1996). Also researchers found that the social presence could be triggered by only minimum intelligence (Biocca, Harms, & Burgoon, 2001). One difficulty in applying the believability qualities to NPCs is that the individual believability qualities are not independent: they are inter-related to each other. This is

why two different believability qualities in one computer character don't guarantee doubled-believability. Often, character believability can be established mainly by one or two distinct individual believability qualities. Perceived character believability may be less than the sum of each individual believability quality if the believability qualities are not designed with a combing principle such as goals and/or personality. In other words, an individual believability quality may have a negative effect on the total character believability if it is in conflict with another main principle. This leads to the following research question:

>   **RQ1: How will the general believability level relate to the five believability qualities? Will the general believability be established by one or few believability qualities?**
>
>   **RQ2: If so, how will other believability qualities affect the general believability? Will lower perceived believability of specific believability qualities be associated with lower general believability?**

Gaming for entertainment is a voluntary activity. Players tend to have preferred genres they play often (for example, see 2008 Pew Foundation tables comparing the frequency of genre play among female and male teenagers). In a study of strategic difference in a computer game with 76 elementary students, Hong and Liu found that students in the expert group used more analogical approaches while the novice group repeated a pattern of trials and errors (2003). Similarly, the difference in visual attention to a first person shooter game was studied. Researchers found that experts who played more than eight hours a week showed faster and more accurate responses to the game

than the novice players who played less than 30 minutes a week (Smith, Tsai, Wong, Brooks, & Peterson, 2008). Higher sport skills and better understanding of game situations (Blomqvist, Luhtanen, & Laakso, 2000) and better problem solving skill (DeVane & Durga, 2008) were found in the expert's group. Considering the above performance difference, it is expected that the perception of believability is different between experts and novices. Returning to believability, those who often play Role Play Games are experienced with the genre. Through experience, they have developed expectations about what to do and the role of NPCs within a game. Others who never play RPGs do not have those expectations. Experts' more developed schemas about RPG NPCs may be associated with higher overall believability ratings. This leads to the following hypothesis:

H6: RPG players will rate NPC believability higher than those who are not experienced with RPGs.

Technology has been one of the important factors in making a game realistic. As technology developed, many factors in games such as character design and sound and video quality became more and more realistic/believable. Researchers found that many technological developments in game design such as sound, graphic, CPU speed, etc. have made games more realistic (Frauenfelder, 2001; Kramer, 1995; Loftus & Loftus, 1983; Newman, 2002). Moreover, studies on game environment such as sounds and graphics revealed that gamers liked a more realistic game environment, and male gamers liked these realistic settings more than female gamers did (Wood, Griffiths, Chappell, & Davies, 2004). Shapiro and Chock studied the relation between typicality and reality

9

using video taped television programs. They found that the drama that contains more realism evokes more enjoyment (2003). Also, believability was insisted as one of the prerequisites of media enjoyment (Vorderer, Klimmt, & Ritterfeld, 2004). Usually new games adapt more advanced technologies than relatively old games. Concurrent with technological advances, game designers continue to innovate and attempt to make better games. This leads to the following hypothesis:

H7a: Characters in Newer games will be more believable than characters in older games.

H7b: Characters in Newer games will be more enjoyable than characters in older games.

The quality of games varies widely, whether the criterion is game sales or critical acclaim. Many rating websites have published the quantified scores of all various media forms such as movie, video games, TV, music, etc. based on various categories. Technological improvement can also be seen in many good-rated games in general. The ingredients for good video games have revealed various aspects of the video games ranging from character to interface. Especially for RPG, deep world history design, evolution of MOBs (evil NPCs) and believability principles of MOB behavior were suggested as factors for making enhancing Massively Multiplayer Online Games (Tychsen). Also, character was reported as one of the important elements with other qualities in creating a believable world (Dormans, 2006). This leads to the following hypothesis:

H8a: Characters in games with good rating will be more believable than characters in lower rating games.

H8b: Characters in games with good rating will be more enjoyable than characters in lower rating games.

Except for several recent games, RPG generally put game players in a good guy position completing a main quest by fighting against villainous characters or going through adventures. Related to the game players' role in games, NPCs usually have two different roles: helpers or antagonists. All NPCs in First-Person-Shooter games are all villains trying to kill game players while NPCs in RPG are populated good guys, bad guys, and neutral characters. Good characters in RPG are characters that help game players by providing some hints or information that can be critical in completing quests while evil characters try to threat/kill hindering game players in finishing quests. It is assumed that evil characters appear to be someone liver posting tangible threats than good characters. On the other hand, good characters are given relatively more important roles than evil characters in general. This leads to the following research question:

RQ3a: How will the general believability level relate to the character good/evilness? Will the evil characters be more believable or less believable?

RQ3b: How will the enjoyment level relate to the character good/evilness? Will the evil characters be more enjoyable or less believable?

METHODS

To test the hypotheses, four RPG games were selected (two new and two old; two with high Metacritic scores and two with low Metacritic scores). Within each game, one antagonist NPC and one protagonist NPC were selected. One minute digital videos of the NPC's interaction with a player were recorded. Study participants viewed each video, answered a series of believability questions about that character, and repeated this process for all eight characters.

NPC Selection

Different game genres include particular NPC roles and NPC-player interactions. For the character believability study with NPCs, a game genre was needed that met some basic requirements of the study. First, the game should guarantee that game players' interaction time with NPCs is long and various enough to experience believability qualities that go beyond physical appearance, such as emotion and personality. Also, it is important for the game to be a single player game. In a multiplayer game, live humans represented by avatars would co-exist in the game with NPCs. In those games, the believability of the NPCs may suffer by comparison to real humans.

For the actual screening of game genres, the list of computer game genres in Metacritic was used. The Metacritic website (www.metacritic.com) uses a unique scoring system called Metascore to evaluate the quality of various electronic medium such as games, television, movie, music, etc. All the game genres in Metacritic were scrutinized with the above screening rules. RPGs satisfy all of the above basic requirements of this study. Specifically, RPGs generally provide a game environment in which game players are able to interact with NPCs without any time limitation. Also, most RPGs have rich

and diverse narrative structures in which NPCs take diverse roles ranging from primary (main quest) to minor (side quest).

Games for this study were selected based on two criteria -- published year and game quality score. Four games would be selected – one newer game with a high Metacritic score, one newer game with a low Metacritic score, one older game with a high Metacritic score, and one older game with a low Metacritic score. Within each of the chosen games, two NPCs would be selected based on the NPC's role (antagonist or protagonist).

RPGs published earlier than year 2006 were grouped in "old games" while games published later than or in the year 2006 were grouped in "new games." For the published year category, year 2006 was selected as the determining year considering the duration of game developments (one to three years)[1]. The Metacritic website listed 63 RPGs initially. The Metascore categorization was decided by the distribution of all RPGs in the Metacritic website considering the lowest score (44) and the highest score (95). For the game quality score, the group with higher than 90 Metascore was labeled as "good games," and games with Metascore lower than 70 were labeled as "bad games."

With the two between-game criteria, all games in the Metacritic were grouped in four different groups: high Metascore-new games (n=1), high Metascore-old games (n= 7), low Metascore-new games (n= 6), and low Metascore-old games (n=13). The four games for the study were randomly selected from those four game groups.

The actual NPC recruitment was conducted within the four games. One protagonist or helpful NPC and one antagonist NPC was recruited from each game. The

---

[1] http://en.wikipedia.org/wiki/Game_development

final NPCs were recruited along with the storyline of the game based on first-timeness and importance of character's role. Deciding the importance of a character's role was dependent on whether the character was involved in main quest.

All final NPCs were important characters staged around events that are related to a main quest. The final eight NPCs were the first good or evil characters encountered along with a main quest. For this, many cheating websites providing step-by-step information along with a main quest were used.

Procedures

The five believability qualities (personality, emotion, appearance and behavior, goal, and social relation) were revealed through extensive literature reviews. The believability of each NPC was measured in general believability questions asking about subjects' general believability perception, as well as in specific believability questions asking about subjects' believability perception of each five believability quality.

Eight video clips were recorded from the very game environment of the eight recruited NPCs. The video clips included almost all behavior patterns of each individual NPC in their game environments. The NPC videos varied in length from one minute thirty seconds to two minutes depending on the behavior patterns.

Subjects in the study were required to access a website which contains all the materials and questionnaires. The order in which the eight NPCs were presented was varied randomly in order to avoid the practice effect in repeated measure. Subjects were able to control the video clips in terms of play, pause, stop, rewind, and fast forward so that they could replay the video clips if they wanted. After each NPC video clip, subjects were asked to answer questions regarding the believability of the NPC they just watched.

The survey consisted of two parts. The first set of questions measured game players' perception on believability of each character they just watched. Four general believability questions asked about general believability of the NPC, and six believability questions asked about their perception on five different NPC qualities (two for appearance and one each for behavior, emotion, goal, social relation, and personality). A final question asked about subjects' general enjoyment of the NPC. All of the operationalizations of these concepts were developed for this study due to the lack of previous research on believability. The response categories used a five level Likert Scale with 5 representing strong agreement and 1 representing strong disagreement. In the last part of the survey, subjects were asked about demographic information such as age, gender, college year, etc.

RESULTS

Subjects in this study were recruited from a sophomore level introductory digital media course at a large Midwestern university. A total of 161 subjects participated. Participants were given extra credit in exchange for their participation. Thirty-eight were female, 119 were male, and 4 left gender blank on the survey. There were 38 subjects in their freshman year, 51 sophomores, 39 juniors, and 28 seniors. The subjects' average age was roughly 21 (20.97) years old. Age of 19 and 20 occupied 49 % of the whole participant population.

General Believability

Four general believability questions measured the participants' general perception on general believability including "In general, this character is believable within the

15

game (Q1)," "The character in the game is what I would expect it to be (Q2)," "I think this character is realistic within the game (Q3)," and "I think I could guess how this character might respond under different circumstances (Q4)." The four general believability questions were combined into a single additive scale. Cronbach's alpha for the GB scale was 0.897.

Specific Believability Qualities

The specific believability (SB) qualities measured the participants' perception on five specific believability qualities that were derived from the believability literature review. The specific believability qualities included "The personality of this character is convincing to me (personality)," "The appearance of this character makes sense to me (appearance 1)," "The way this character behaves makes sense to me (appearance 2)," "The way this character responds to me resembles how humans respond to each other (social relation)," "I can clearly understand this character's motivations (goal)," and "The emotional expression of this character is very clear and convincing to me (emotion)." The five specific believability questions were combined into a single additive scale. Cronbach's alpha for the SB scale was 0.917.

Enjoyment

The enjoyment question measured the participants' general perception on how much they expected to enjoy playing with the NPC: "It would be enjoyable to play with this character in a game (enjoyment)."

H1 to H5: General and specific Believability

The general believability score and specific believability qualities were subjected

to a Pearson correlation. For an alpha level of .01, all correlations were found to be statistically significant.

In general, personality and appearance (both appearance and behavior) showed high correlation with general believability. The highest correlation with the general believability among believability qualities was personality ($r(159)= 0.805$, $p < 0.01$) while the second and third highest correlation was behavior ($r(159)=0.789$, $p < 0.01$) and appearance ($r(159) = 0.769$, $p < 0.01$) in turn. Among five believability qualities, appearance 1(behavior), appearance 2, goals, emotion, and personality were strongly correlated with general believability while the correlation between social relation and general believability was significant but weak. **Hypothesis 1, 2, 3, 4 were supported by the data while hypothesis 5 (general believability and personal relations) was not supported.**

The five believability qualities were subjected to a Pearson correlation. For an alpha level of .01, all correlations were found to be statistically significant. The strongest correlation was found between personality and appearance ($r(159) = 0.826$, $p < 0.01$) while the weakest correlation was between goal and social relation ($r(159) = 0.498$, $p < 0.01$).

RQ1 and RQ2: Which specific believability qualities are most related to general believability?

Regression was used to analyze the relationship between the general believability and the specific believability qualities. Consistent with the result from correlation analysis, the regression result showed higher coefficients in personality, appearance, and

behavior with statistical significance. In all NPCs, six believability qualities explained 60 to 78 percent of the whole variation in general. (Table 1 goes here)

Personality, appearance 1 (outfit), and appearance 2 (behavior) were significant predictors for general believability with similar level. General believability seemed to reflect surface level of believability. Table 2 shows the five believability quality scores of eight different NPCs along with their general believability scores as well as the average of five believability quality scores. (Table 2 goes here)

Individual Differences and Believability

H6: RPG gaming hours will result in higher general believability.

Regression was used to analyze the relationship between hours of RPG playing and general believability. The hours of RPG playing did not significantly predict the general believability (b = .011, $t(151) = 1.663$, $p = 0.098$). Hypothesis H6 was not supported.

Believability Factors

Factor analysis was conducted on the four general believability and six specific believability quality questions which were averaged with eight NPCs to look for underlying dimensions within each character. Principal components analysis with Varimax rotation was used.

Two factors emerged accounting for 79.68% of the variance. Three of the general believability questions (general believability question 1, 2, and 3) and three of the specific believability qualities (personality, appearance1 and appearance2) loaded .6 or higher on factor 1. General believability question 4 and four of the specific believability

qualities (personality, social relation, goal and emotion) loaded .6 or higher on factor 2. Three questions -- "the personality of this character is convincing to me (personality)," "the way this character behaves makes sense to me" (appearance2), and "I can clearly understand this character's motivations (goal)," loaded higher than .5 on both factors. The items loading high on both factors have something in common conceptually-- they relate to how well the respondent feels they could predict how the NPC might behave.

Dimension 1, Surface Believability, combines four questions that had higher loadings than 0.8 in only component 1. Surface Believability relates to judgments about the NPC which could be made quickly, at first glance. Average responses (3.53) combining all characters were higher than neutral (3) on the five-point scale of Surface Believability ranging from 1.59 to 5.

Dimension 2, Personal Believability, combined three believability qualities (the appearance 2 behavior patterns, personality and goal). Average responses combining all characters were 3.33 on a five-point scale of Personal Believability ranging from 1.71 to 4.75. Personality Believability measured a more intermediate type of believability attributes that need a little more cognitive capability to process than Surface Believability. However, these believability attributes didn't need heavy cognitive processing compared to the Predictable Believability.

Predictable Believability, the special third scale, combined two believability qualities (emotion and social relation) and one general believability quality. Average responses were slightly higher than neutral (3) on the five-point scale of Predictable Believability ranging from 1.58 to 4.75. Predictable Believability consisted of attributes that need heavy cognitive capability to process and time interacting with the character.

Predictable Believability is probably the hardest for designers to incorporate and might be expected to evidence lower believability ratings.

The three believability scales were constructed by summing items that loaded .6 or higher on the factor with their relative contribution on the other factor. The summed scales were then divided by the number of items so that the resulting means could be interpreted as roughly corresponding to the 5 point Likert scale used for the individual items. Table 3 shows the correlations between three new believability dimensions. All the correlations were statistically significant at the 0.01 level. (Table 3 goes here)

Cronbach's Alpha was calculated to check the reliability of these scales. All reliabilities were above .78, with Surface Believability at .954, Personal Believability at .9, and Predictable Believability at 0.781.

Comparing Believability of High and Low Rated Games, New and Old Games, and Good and Evil Characters

The Surface Believability, the Personal Believability, and the Predictable Believability scores were subjected to a four-way ANOVA with three levels of NPC evilness (good, bad), game newness (new, old), game rate (high, low), and gender (male, female).

H7a: Characters in newer games will be more believable than characters in older games.

There was a significant main effect for game newness, $F(1, 155) = 32.296$, $p < 0.001$, indicating that the mean **Surface Believability** score was significantly greater for new games ($M = 3.66$, $SD = 0.058$) than for old games ($M = 3.42$, $SD = 0.060$). Participants showed higher Surface Believability on NPCs recruited from newer games than NPCs recruited from older games.

There was a significant main effect for game newness, $F(1, 155) = 23.722$, $p <$ 0.001, indicating that the mean **Personal Believability** score was significantly greater for new games ($M = 3.48$, $SD = 0.056$) than for old games ($M = 3.27$, $SD = 0.058$). Participants showed higher Personal Believability on NPCs recruited from newer games than NPCs recruited from older games.

Also, there was a significant main effect for game newness, $F(1, 155) = 22.531$, $p < 0.001$, indicating that the mean **Predicable Believability** score was significantly greater for new games ($M = 3.21$, $SD = 0.052$) than for old games ($M = 2.98$, $SD = 0.054$). Participants showed higher Predictable Believability on NPCs recruited from newer games than NPCs recruited from older games. The hypothesis H7a was supported.

H8a: Characters in games with good rating will be more believable than characters in lower rating games.

There was a significant main effect for game rate, $F(1, 155) = 4.477$, $p = 0.035$, indicating that the mean **Surface Believability** score was significantly greater for high Metascore games ($M = 3.59$, $SD = 0.06$) than for low Metascore games ($M = 3.49$, $SD = 0.058$). Participants showed higher Surface Believability on NPCs recruited from games with a higher rating than NPCs recruited from games with a lower rating.

There was a significant main effect for game rate, $F(1, 155) = 47.236$, $p < 0.000$, indicating that the mean **Personal Believability** score was significantly greater for high Metascore games ($M = 3.53$, $SD = 0.058$) than for low Metascore games ($M = 3.22$, $SD = 0.057$). Participants showed higher Personal Believability on NPCs recruited from games

with a higher rating than NPCs recruited from games with a lower rating. The hypothesis 3 was supported.

There was a significant main effect for game rate, $F(1, 155) = 19.115$, $p = 0.000$, indicating that the mean **Predicable Interaction** score was significantly greater for high Metascore games ($M = 3.19$, $SD = 0.053$) than for low Metascore games ($M = 3.00$, $SD = 0.051$). Participants showed higher Predictable Believability on NPCs recruited from games with a higher rating than NPCs recruited from games with a lower rating. The hypothesis H8a was supported.

RQ3a: How will the general believability level relate to the character good/evilness? Will the evil characters be more believable or less believable?

There was no significant main effect for character evilness in Surface Believability, $F(1,155)= 1.181$, $p=0.279$. The **Surface Believability** was not significantly different between good characters (3.57) and evil characters (3.52). Participants showed no Surface Believability difference between good characters and evil characters.

There was no significant main effect for character evilness in Personal Believability, $F(1,155)= 0.606$, $p=0.437$. The **Personal Believability** was not significantly different between a good character (3.4) and evil character (3.35). Participants showed no Personal Believability difference between good characters and evil characters.

There was a significant main effect for character evilness in **Predictable Interaction**, $F(1,155)= 16.122$, $p=0.000$. The Predicable Interaction was significantly

different between good characters and evil characters. Good characters showed significantly higher Predictable Believability (M = 3.4, SD =.054) than that of evil characters (M = 2.98, SD = .057). Participants showed higher Predictable Believability on good characters than evil characters.

Comparing Enjoyment of High and Low Rated Games, New and Old Games, and Good and Evil Characters

ANOVAs were used to compare the enjoyment of each character among good/bad NPC, high/low Metascore, new/old game and between genders. The item for enjoyment was "It would be enjoyable to play with this character in a game." Average responses (3.04) were about neutral (3) on the five-point scale of enjoyment. The Surface Believability scores were subjected to a four-way ANOVA with three levels of NPC evilness (good, bad), game newness (new, old), game rate (high, low), and gender (male, female).

H7b: Characters in newer games will be more enjoyable than characters in older games.

There was a significant main effect for game newness, $F(1, 155) = 4.844$, $p = 0.029$, indicating that the mean enjoyment score was significantly greater for new games ($M = 3.10$, $SD = 0.067$) than for old games ($M = 2.97$, $SD = 0.067$). Participants showed higher enjoyment on NPCs recruited from newer games than NPCs recruited from older games. The hypothesis 7b was supported.

H8b: Characters in games with good rating will be more enjoyable than characters in lower rating games.

There was a significant main effect for game rate, $F(1, 155) = 12.824$, $p = 0.000$, indicating that the **enjoyment** was significantly greater for high Metascore games ($M =$

3.14, $SD = 0.065$) than for low Metascore games ($M = 2.93$, $SD = 0.067$). Participants showed higher enjoyment on NPCs recruited from games with a higher rating than NPCs recruited from games with a lower rating. The hypothesis 8b was supported.

RQ3b: How will the enjoyment level relate to the character good/evilness? Will the evil characters be more enjoyable or less enjoyable?

There was no significant main effect for game rate, $F(1,155)= 2.205$, p=0.14. The enjoyment was not significantly different between good characters (3.09) and evil characters (2.98). Participants showed no enjoyment difference between good characters and evil characters.

Interaction Effects

There was a significant main effect between game rate and game newness ($F(1,155)=7.865$, $p=0.006$), game rate and game newness (F(1,155)= 5.867, p= 0.017). The Predictable Believability difference between new games and old games was higher in low Metascore games (0.23) than high Metascore games (0.11). Also the possible-enjoyment of characters across game newness and game rate was about the same except for characters from low rated and old games.

A Pearson correlation addressed the relationship among enjoyment ($M = 3.04$, $SD = 0.64$), Surface Believability ($M = 3.53$, $SD = 0.58$), Personal Believability ($M = 3.34$, $SD = 0.56$), and Predictable Believability ($M = 3.09$, $SD = 0.51$). For an alpha level of .01, all correlations were found to be statistically significant. The possible enjoyment was related a little more with Personal Believability than Surface Believability or Predictable Believability.

DISCUSSION

General believability was found to be closely related to specific believability qualities. Four of five specific believability qualities (personality, emotion, goal, and appearance) were significantly contributed to build the general believability while social relation was not related to the general believability. It may be due to the fact that many NPCs in RPGs appear by themselves without having any social relation with other NPCs. Also, the NPC's social relationship with the game player can be revealed only by the interaction with game players. Perhaps having research participants watch video of the NPC in the study instead of actually playing with the characters limited their experience of interaction. Furthermore, NPCs don't engage in social relations with game players until the player approaches the boundary which triggers the NPC's interaction.

The believability perception game players construct about a character through game playing seems to be driven by one or a few qualities. In this study, the general believability perception of the characters was driven by the appearance of each character regardless of games' newness, game ratings, or characters' good/bad role. The general believability score was almost the same as the highest believability score among five believability qualities. It seemed that the general believability perception was constructed mainly by one or few well-designed believability qualities.

The results clearly show that characters from high rated games are more believable than characters from low rated games. Also, characters from newer games were found to be more believable than characters from older games. Expected NPC enjoyment was very consistent with NPC believability. Characters from newer games or high rated games were expected to be more enjoyable. The hypotheses regarding game rating and game newness were strongly supported. Newer games are able to draw upon

newly developed technologies such as better 3D graphics and sound as well as newer AI and other game algorithms. Also, highly rated games are probably highly rated in part because they incorporate well-designed, believable NPCs. Highly rated new games were particularly strong in NPC Predicable Believability.

Protagonist and antagonist NPCs were most different in terms of Predictable Believability. Good characters showed significantly higher Predictable Believability than evil characters. No significant difference was found between good character and bad character design for Surface Believability or Personal Believability or for expected NPC enjoyment. The result may be due to the fact that Surface and Personal Believability are mostly based on visual cues, whereas Predictable Believability requires inferring intent and extrapolating about behavior beyond the immediate gaming situation. Perhaps we expect good people to be consistent, whereas bad guys are more of a mystery. Or perhaps antagonist NPCs are not as carefully or fully developed in a game. People seem to perceive the good character as more believable than the evil character only after a certain amount of interaction with it because the character intention such as goodness and evilness of character can only be revealed in the game storyline. In other words, people can feel a good character or an evil character more believable when they can make some prediction on that character.

The effect of improvement in good characters was found within game newness. The improvement of newer games seems to be found more on good character design than evil character design. This is where improvements were found for NPC Personal Believability, Predictable Believability and enjoyment. Good characters were more believable than bad characters from newer games in the Personal Believability and

Predictable Believability which need some interaction time to be revealed. Also good characters were more possibly enjoyable than evil characters in newer games. The improvement from new game design seems to have more impact on the believability which needs more interaction time compared to immediate, snap judgment Surface Believability.

The improvement of Predictable Believability could be found in characters from either new games or high rated games or both. Characters from old and low rated games both showed significantly lower Predictable Believability.

Surprisingly, RPG experience, represented by RPG playing time, was not a significant predictor of any of the believability scales. There was no difference in how believable the NPCs were for study participants who frequently played RPGs and those who never played. Those two groups of subjects would be expected to have vastly different schemas for RPG NPCs. Yet both types of respondents made similar judgments of the NPCs.

As Uncertainty Reduction Theory described, the three believability subscales are dependent on time. Surface Believability is a kind of perception people construct in the first place when they encounter strange characters in a game setting. Surface Believability is the first perception people build based on character. Generally, Surface Believability is built on the visual information such as appearance. Once people build an initial perception on character, they tend to seek further information which can be interpreted personally to them. It will take a longer time for game players to construct Personal Believability than Surface believability. In Surface Believability phase, game players construct more perception on a character's personality side on top of Surface

Believability. That's why Personal Believability is slower than Surface Believability in terms of the speed of recognition. Personal Believability is situated in the middle between Surface Believability and Predictable Believability in terms of its source of information. On one hand, Personal Believability is constructed on the very low level perception and believability quality such as appearance. On the other hand, Personal Believability shares a lot of common cognitive grounds with Predictable believability. Predictable Believability will be the last perception game players construct on a character due to its cognitive nature. In this phase, game players try to create a schema for a character and make a prediction on it.

The Uncertainty Reduction Theory, however, cannot explain the whole interaction between game players and game character. Different with Uncertainty Reduction Theory, game players are more active and generous in receiving and interpreting the information they get in the course of interaction with a game character. In a game environment, game players much more willingly suspend their disbelief than people in the interaction with strangers. Just because of the definition of believability: a willing suspension of disbelief, the process of building believability from Surface Believability, through Personal Believability, to Predictable Believability is much easier for game players than people who come across strangers in a real life setting.

The mean value of the three believability scales showed a snap shot of time variation among the scales. The materials for this study are around two minutes long, and it may not be enough for the participants to construct perceptions down to Predictable Believability level. For the two minutes, participants showed the highest mean value on Surface Believability while Predictable Believability is the lowest among the scales.

Personal Believability is in the middle. The variation of believability values among the three scales shows that Surface Believability contributed the most in terms of building general believability on a character while Predictable Believability contributed the least in the first two minutes of interaction. It cannot directly apply to the real game setting because this study used a game video rather than a real game. However, it seems that game players depend more on Surface Believability quality than the qualities that need more cognitive power in order to interpret a character as believable.

FUTURE STUDY

Character believability includes three different stages with different response time. Due to the fact that each believability stage has a distinct mode of interaction, investigating each individual stage in detail will be necessary in order to answer some questions: which stage is more important than others in terms of making character more believable? How does each stage affect some important concepts such as enjoyment, like, satisfaction, etc? How are these stages different with those in Uncertainty Reduction Theory explained? It is especially important to define what kind of information is delivered to game players in each different stage. Fundamentally, human beings' interaction with a computer (game characters in this case) is expected to be different than that with other human beings (as URT expected) in the precedent assumption about their interaction partner. Research on how this precedent assumption on the interaction partner can change the mode and attitude toward the partner will be needed.

In a bigger picture, the causal relationship between character believability and the quality of game will be interesting even though it was not studied in this research. Good role playing games include many other factors including character believability. This

study found empirical evidence that high rated and relatively newer games are populated with more believable characters than low rated and older games in general. However, it is not clear how the character believability is related to the general game assessment. If there are other factors affecting the game assessment, what are they? Especially in a Role Playing Game, the character believability may be more important than other factors due to the nature of the game. Studies on factors that affect the Role Playing Game evaluation not only from experts' perception but also causal game players' perception will be important in terms of providing some tips to character designers.

Character believability studies aim to make a distinction between avatar and Non-Player Character in an environment in which both types of characters co-exist affecting each other. For human beings, it is very important to ask whether the characters we are interacting with are avatars (characters controlled by human beings) or Non Player Characters (character controlled by computer). Many virtual worlds such as Second Life or MMORPGs (Massively Multiplayer Online Role Playing Game) are populated with both avatars and NPCs. It is relatively easy for human beings to recognize NPCs because many NPCs are not believable. A kind of safe line between avatars and NPCs is clear by the less believable design of NPCs, not by any intentional intervention from the designer's side. It is critical to study how human beings recognize the existence of NPCs and what those factors are that make human beings guess who is what. Sooner or later, the NPCs will be more believable as the design and technology develops.

Regarding character believability, an interesting question raised from the literatures was about source credibility. A fundamental assumption of character believability is that users know that the computer characters they are interacting with are

not real people. They willingly suspend their disbelief (ignoring the fact that characters are not real) for the joy of interacting with them. However, the credibility that players assign to the characters seems to have a different mechanism. People may not be as willing to suspend their disbelief when the scope of interaction moves outside of entertainment. For example, it may not be a good idea to have a clearly fictional computer agent offer medical advice or sports rehabilitation coaching. Humans may not like the idea of being evaluated or advised by artificial computer characters. Credibility and believability are not synonymous. Believability contributes to enjoyment of games for entertainment. Credibility becomes important in "serious games." Research on how to create or increase the credibility of a computer character will be needed as patterns of human computer interaction become more diverse and serious.

Table 1. Regression between General Believability and Specific Believability Qualities

	Coefficients	b	p	R^2	F/sig
Dremora	Personality	0.158	0.005	0.628	43.368 / 0.000
	Appearance	0.302	0.000		
	Behavior	0.066	0.276		
	Social relation	-0.20	0.547		
	Goal	0.178	0.000		
	Emotion	0.032	0.441		
Jauffre	Personality	0.205	0.000	0.778	90.068 / 0.000
	Appearance	0.198	0.000		
	Behavior	0.246	0.000		
	Social relation	0.094	0.023		
	Goal	-0.17	0.703		
	Emotion	0.153	0.000		
Terrorist	Personality	0.136	0.014	0.607	39.613 / 0.000
	Appearance	0.174	0.000		
	Behavior	0.164	0.004		
	Social relation	0.003	0.940		
	Goal	0.128	0.003		
	Emotion	0.138	0.004		
Paul	Personality	0.138	0.011	0.715	64.545 / 0.000
	Appearance	0.213	0.000		
	Behavior	0.215	0.000		
	Social relation	0.092	0.025		
	Goal	0.181	0.000		
	Emotion	0.028	0.437		
Skeleton	Personality	0.215	0.000	0.719	65.718 / 0.000
	Appearance	0.229	0.000		
	Behavior	0.193	0.000		
	Social relation	0.009	0.746		
	Goal	0.023	0.522		
	Emotion	0.078	0.048		
Vesit	Personality	0.303	0.000	0.631	43.972 / 0.000
	Appearance	0.152	0.001		
	Behavior	0.122	0.035		
	Social relation	-0.004	0.943		
	Goal	0.155	0.001		
	Emotion	-0.012	0.800		
Grom	Personality	0.191	0.000	0.661	50.100 / 0.000
	Appearance	0.345	0.000		
	Behavior	0.131	0.008		
	Social relation	0.010	0.795		
	Goal	0.005	0.897		
	Emotion	0.067	0.111		
Yrsa	Personality	0.293	0.000	0.693	57.890 / 0.000
	Appearance	0.259	0.000		
	Behavior	0.159	0.000		
	Social relation	-0.049	0.331		
	Goal	0.120	0.019		
	Emotion	0.011	0.814		

Table 2. The Five Believability Quality Scores Among 8 NPCs

	New				Old			
	High Score		Low Score		High Score		Low Score	
	Evil	Good	Evil	Good	Evil	Good	Evil	Good
	Dremora	Jauffre	Terrorist	Paul	Skeleton	Vesit	Grom	Yrsa
Personality	3.43	3.73	3.14	3.52	3.27	3.25	3.38	2.82
Emotion	3.24	3.41[b]	2.81	3.19[b]	3.22	3.11[b]	3.27	2.59
Social relation	2.59[b]	3.56	2.62[b]	3.5	2.53[b]	3.21	2.50[b]	2.387[b]
Goals	3.34	3.74	3.22	3.69[a]	3.34	3.31	3.22	2.53
Appearance	3.58[a]	3.91[a]	3.28[a]	3.59	3.51[a]	3.56[a]	3.56[a]	3.10[a]
Behavior	3.51	3.79	3.24	3.69[a]	3.40	3.39	3.32	2.80
General B	3.52	3.74	3.33	3.63	3.48	3.53	3.50	3.09
Correlation with General B	R=0.731 P=0.000 N=161	R=0.866 P=0.000 N=161.5	R=0.764 P=0.000 N=161	R=0.828 P=0.000 N=161	R=0.803 P=0.000 N=161	R=0.768 P=0.000 N=161	R=0.733 P=0.000 N=161	R=0.779 P=0.000 N=161

< a=the highest believability among five believability quality>
< b=the lowest believability among five believability quality>

Table 3. The Result of Correlations between Three New Believability Dimensions

	Surface Believability	Personal Believability
Personal Believability	.854	
Predictable Believability	.619	.827

Reference

Allbeck, J., & Badler, N. (2002). Toward Representing Agent Behaviors Modified by Personality and Emotion. *Proceedings of the First Annual Joint Conference on Autonomous Agents and Multiagent Systems*, 202-207.

Andrews, J. C., Netemeyer, R. G., & Durvasula, S. (1991). Effects of consumption frequency on believability and attitudes toward alcohol warning labels. *Journal of Consumer Affairs, 25*(2), 323-338.

Aylett, R. (1999). Narrative in Virtual Environments-Towards Emergent Narrative. *Working notes of the Narrative Intelligence Symposium*.

Bates, J. (1994). The Role of Emotion in Believable Agents. *Communications of the ACM, 37*(7), 122-125.

Bates, J., Loyall, A. B., Reilly, W. S., Castelfranchi, C., & Wemer, E. (1994). An architecture for action, emotion, and social behaviour. *Artificial Social Systems Selected Papers from the Fourth European Workshop on Modelling Autonomous Agents in a Multi-Agent World,MAAMAW-92 (LNAI, 830*, 55-68.

Beltramini, R. F. (1982). Advertising perceived believability scale. *Proceedings of the Southwestern Marketing Association*, 1-3.

Berkos, K. M. (2003). The effects of message direction and sex differences on the interpretation of workplace gossip *The Department of Communication Studies Doctor of Philosophy*.

Bhatt, K. (2004). *Believability in Computer Games*. Paper presented at the Australian Workshop on Interactive Entertainment. from http://research.it.uts.edu.au/creative/ie/04/proceedings/081%20bhatt.pdf

Biocca, F., Harms, C., & Burgoon, J. K. (2001). Criteria for a theory and measure of social presence. *Retrieved February, 6*, 2003.

Blomqvist, M., Luhtanen, P., & Laakso, L. (2000). Expert-novice differences in game performance and game understanding of youth badminton players. *Physical Education & Sport Pedagogy, 5*(2), 208-219.

Brown, E., & Cairns, P. (2004). *A grounded investigation of game immersion.* Paper presented at the ACM Conf. on Human Factors in Computing.

DeVane, B., & Durga, S. (2008). *Problem-solving in history: strategy games and schema.* Paper presented at the Proceedings of the 8th international conference on International conference for the learning sciences.

Dormans, J. (2006). Creating a believable gameworld. Retrieved from http://www.judithdormans.com/essays/Creating_a_believable_gameworld.pdf

Ewing, T. N. (1940). *A study of certain factors involved in changes of opinion.* Duke University.

Frauenfelder, M. (2001). Smash Hits. *Wired, 9.08, August.*

Hayes-Roth, B., & Doyle, P. (1998). Animate Characters. *Autonomous Agents and Multi-Agent Systems, 1*(2), 195-230.

Hayes-Roth, B., Maldonado, H., & Moraes, M. (2002). *Designing for diversity: Multi-cultural characters for a multi-cultural world.* Paper presented at the IMAGINA 2002, Monte Carlo, Monaco.

Hong, J., & Liu, M. (2003). A study on thinking strategy between experts and novices of computer games. *Computers in Human Behavior, 19*(2), 245-258.

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and Persuasion: Psychological Studies of Opinion Change*: Yale University Press.

Kramer, G. (1995). Sound and communication in virtual reality. *Lea'S Communication Series*, 259-276.

Lee, M. S., & Heeter, C. (2008). *Computer Science and Communication Perspectives on character believability in games.* Paper presented at the Annual meeting of the International Communcation Association.

Loftus, G. R., & Loftus, E. F. (1983). *Mind at play: the psychology of video games*: Basic Books.

Loyall, A. B. (1997). *Believable Agents: Building interactive personalities.*, Carnegie Mellon University, Pittsburgh, PA.

Mateas, M. (1997). *An OZ-centric review of interactive drama and believable agents.* (No. CMU-CS-97-156): School of Computer Science, Carnegie Mellon University.

Mateas, M. (2002). *Interactive Drama, Art and Artificial Intelligence.* University of California.

Newman, J. (2002). The Myth of the Ergodic Videogame. *Game Studies, 2*(1), 1-17.

Reeves, B., & Nass, C. (1996). *The media equation: how people treat computers, television, and new media like real people and places*: Cambridge University Press New York, NY, USA.

Reilly, W. S., & Bates, J. (1995). Natural Negotiation for Believable Agents. *Technical Report CMU-CS-95-164, Carnegie Mellon University*.

Reilly, W. S. N. (1997). A methodology for building believable social agents. *Proceedings of the first international conference on Autonomous agents*, 114-121.

Romano, D. M., & Wong, A. K. L. (2004). *Personality Model of a Social Character.* Paper presented at the Workshop on Empathy in Human Computer Interaction, 7th of September at HCI 2004, the 18th British HCI Group Annual Conference.

Shapiro, M. A., & Chock, T. M. (2003). Psychological Processes in Perceiving Reality. *Media Psychology, 5*(2), 163-198.

Smith, C. F., Tsai, Y.-F. D., Wong, J. H., Brooks, D. T., & Peterson, M. S. (2008). More than meets the eye: Investigating expert and novice differences in action video games. *Journal of Vision, 8*(6), 105-105.

Szilas, N. (2003). IDtension: a narrative engine for Interactive Drama. *Technologies for Interactive Digital Storytelling and Entertainment Conference*, 187–203.

Thomas, F., & Johnson, O. (1981). *The Illusion of Life:Disney Animation*: Hyperion Books.

Tychsen, A. (July 2 2007). Playing Stories - understanding role playing games and their players Retrieved Sep 23, 2009, from http://web.science.mq.edu.au/~atychsen/html2/believability.htm

van Doorn, M., & de Vries, A. P. (2006). Co-creation in Ambient Narratives: Ambient Intelligence for Everyday Life (AmI-Life'05), Lecture Notes in Computer Science.

Vorderer, P., Klimmt, C., & Ritterfeld, U. (2004). Enjoyment: At the heart of media entertainment. *Communication Theory, 14*(4), 388-408.

Watson, M. (2002). Constructing Virtual Space.

Wood, R. T. A., Griffiths, M. D., Chappell, D., & Davies, M. N. O. (2004). The Structural Characteristics of Video Games: A Psycho-Structural Analysis. *CyberPsychology & Behavior, 7*(1), 1-10.