SOCIAL RESPONSES TO VIRTUAL HUMANS: THE EFFECT OF HUMAN-LIKE CHARACTERISTICS

A Dissertation
Presented to
The Academic Faculty

by

Sung Park

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in the
School of Psychology

Georgia Institute of Technology

May 2009

Social Responses to Virtual Humans: The Effect of Human-like Characteristics

Committee Members:

Dr. Richard Catrambone, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Gregory Corso
School of Psychology
Georgia Institute of Technology

Dr. Jack Feldman
School of Psychology
Georgia Institute of Technology

Dr. Wendy A. Rogers
School of Psychology
Georgia Institute of Technology

Dr. John T. Stasko
School of Interactive Computing
Georgia Institute of Technology

Date Approved: 04/27/2009

I dedicate this thesis to my parents and my sister, for their never-ending support.

ACKNOWLEDGEMENTS

I wish to thank Dr. Richard Catrambone for his guidance. I would also like to thank the members of the Problem Solving and Educational Technology lab for their support and creativity.

TABLE OF CONTENTS

Abstract

A framework for understanding the social responses to virtual humans suggests that human-like characteristics (e.g., facial expressions, voice, expression of emotion) act as cues that lead a person to place the agent into the category "human" and thus, elicit social responses. Given this framework, this research was designed to answer two outstanding questions that had been raised in the research community (Moon & Nass, 2000): 1) If a virtual human has more human-like characteristics, will it elicit stronger social responses from people? 2) How do the human-like characteristics interact in terms of the strength of social responses? Two social psychological (social facilitation and politeness norm) experiments were conducted to answer these questions. The first experiment investigated whether virtual humans can evoke a social facilitation response and how strong that response is when participants are given different cognitive tasks (e.g., anagrams, mazes, modular arithmetic) that vary in difficulty. They did the tasks alone, in the company of another person, or in the company of a virtual human that varied in terms of features. The second experiment investigated whether people apply politeness norms to virtual humans. Participants were tutored and quizzed either by a virtual human tutor that varied in terms of features or a human tutor. Participants then evaluated the tutor's performance either directly by the tutor or indirectly via a paper and pencil questionnaire. Results indicate that virtual humans can produce social facilitation not only with facial appearance but also with voice recordings. In addition, performance in the presence of voice synced facial appearance seems to elicit stronger social facilitation (i.e., no statistical difference compared to performance in the human presence condition) than in the presence of voice only or face only. Similar findings were observed with the politeness norm experiment. Participants who evaluated their tutor directly reported the tutor's performance more favorably than participants who evaluated their

tutor indirectly. In addition, this valence toward the voice synced facial appearance had no statistical difference compared to the valence toward the human tutor condition. The results suggest that designers of virtual humans should be mindful about the social nature of virtual humans.

Chapter 1: Introduction

Interest in virtual humans or embodied conversational agents is growing in the realm of human computer interaction (HCI). Recent improvements in computation have facilitated the use of virtual humans in various applications, such as entertainment (Jeong, Hashimoto, & Makoto, 2004), engineering (Zorriassatine, Wykes, Parkin, & Gindy, 2003), clinical practice (Glantz, Rizzo, & Graap, 2003), and the military (Hill et al., 2003). Nass' work, and the work of other researchers (Lester et al., 1997; Nass, Moon, & Carney, 2000; Reeves & Nass, 1996), suggests that there is a striking similarity between how humans interact with one another and how a human and a virtual human interact. In this dissertation I intended to gain a deeper understanding of the social dimension of the interaction (i.e., social interaction) between humans and virtual humans. Specifically, I am interested in the following two questions:

1) If a virtual human has more human-like characteristics, will it elicit stronger social responses from people?
2) How do the human-like characteristics interact in terms of the strength of social responses?

Such questions have been raised in the research community (Moon & Nass, 2000) but not empirically addressed. A qualitative framework of how humans respond socially to virtual humans (Chapters 2, 3, and 4) suggests (Chapter 4) that as long as a virtual human is categorized as human, then the number of human-like characteristics does not affect the strength of social responses.

To test this hypothesis I conducted two social psychological experiments (Chapter 5): 1) a social facilitation experiment and 2) a politeness norm experiment. Details on methods and procedures (Chapter 6 and Chapter 7) and results (Chapter 8) will follow.

Chapter 2: Social Responses to Virtual Humans

Being social refers to people thinking about, influencing, and relating to one another (Myers, 2005). Being social constitutes many behaviors (e.g., perception of others and judgment of others) and psychological constructs (e.g., belief, persuasion, attraction). In the human computer interaction literature, responding socially to virtual humans means that people exhibit behavioral responses to virtual humans as if they are humans (i.e., social behaviors). Based partly on the work of Nass and his colleagues (Nass, Steuer, & Tauber, 1994; Reeves & Nass, 1996; for a review, see Moon & Nass, 2000), there is general agreement in the research community that people do respond socially to computers. Nass's research method (e.g., Nass et al., 1994) is essentially a Turing test (Turing, 1950) but assesses non-human entities' social capability (i.e., whether they can elicit social responses from humans) rather than assesses a machine's ability to demonstrate intelligence. His procedure is to take a well-established study from social psychology, replace the person in the study with a computer, and assess whether the results of the study are equivalent to the original study with a person, that is, whether participants behave in the same way when interacting with a computer as they would have with a human. In their studies, participants applied many social rules (e.g., ethnicity stereotype, emotional consistency) on their behaviors to computers that are normally observed when interacting with humans.

However, there are only a handful of studies that have tested Nass's argument or critically examined the methodology. Most researchers have accepted the claim that "interactions with computers are fundamentally social" (p. 72, Nass et al., 1994), with few caveats or limitations, and take it as the basis of their work. Few researchers have questioned external validity, not to mention ecological validity. Sundar (2004) questioned the validity of generalizing findings from a short-term, one-time interaction with a newly encountered computer, when everyday interactions with computers are typically long-term, persistent, and personalized. History and length are critical factors in any social interaction, which I will reveal by identifying the psychological processes of social behavior in this chapter.

What seems to be vital, in terms of understanding the question at hand (i.e., why people respond socially to virtual humans), is literature on how we as humans respond socially to another person and how we know about others' temporary states (e.g., emotions, intentions) and enduring dispositions (e.g., beliefs, abilities); topics in social psychology that are relevant to the discussion but have been largely omitted. Finally, key psychological concepts of social behavior such as control and automaticity have not been considered as tools for understanding social interactions with virtual humans.

In this chapter I will recast the virtual human literature by incorporating insights gained from social psychology. I will lay out a framework from the person perception literature on how we respond socially to people; this framework will lead us to reconstruct our understanding of the observed social responses to virtual humans. Given this, I will hypothesize that human-like characteristics act as cues or stimuli that lead a person to place the virtual human into the category "human" and thus, elicit social responses. These processes are practiced since birth and are strongly ingrained and have the characteristics of automaticity (e.g., generally unconscious), such that a single cue may activate the human category (automatic over-reliance) and block other cues that would activate the virtual human category.

Lack of knowledge regarding virtual humans might be one factor as to why we ignore asocial cues of virtual humans. This indicates that individuals do not have enough experience with virtual humans to form a rigid and concrete virtual human category. Experiencing a virtual human and realizing how the technology is still far from perfectly mimicking a human and identifying the asocial nature of virtual humans might contribute to forming a category of virtual human, where its disposition is non-human and non-social. The framework of how humans respond socially to virtual humans provides critical insights in answering the research questions presented in Chapter 1.

How Humans Respond Socially to Humans

Although many theories have been offered to explain why people respond socially to virtual humans, it is surprising that these theories have largely omitted the literature on how we as a person respond socially to another person. Given that virtual humans have human-like characteristics, it seems only natural to investigate the social nature of humans.

"Ordinary personology" refers to the ways in which an individual comes to know about another individual’s temporary states (emotions, intentions, and desires) and enduring dispositions (beliefs, traits, and abilities; Gilbert, 1998). However, there is no unified theory of ordinary personology that uses a standard set of language to cover various topics in social psychology, such as categorization, stereotyping, attribution, impression formation, all of which are relevant to the question at hand (i.e., social responses to virtual human).

Nevertheless, Gilbert (1998) suggested a framework for ordinary personology, which attempted to incorporate attribution theories (i.e., how a stimulus engenders identification and dispositional inferences) and social cognitive theories (i.e., how responses engender impression; See Figure 1). An example used in Gilbert’s paper explains the framework well: an actor who has an appearance (e.g., a crucifix or a Mohawk) that allows him to be classified as a member of a category (e.g., a priest or a punk), which then allows the observer to draw inferences about the actor’s dispositions (he is religious or rebellious). In addition to the discrete items of information gathered from the disposition process (e.g., he should be religious because he's a priest), we also gather information from observation to find a unifying explanation (i.e., form an impression) for those features. This sometimes strengthens categorization (e.g., he's a priest).

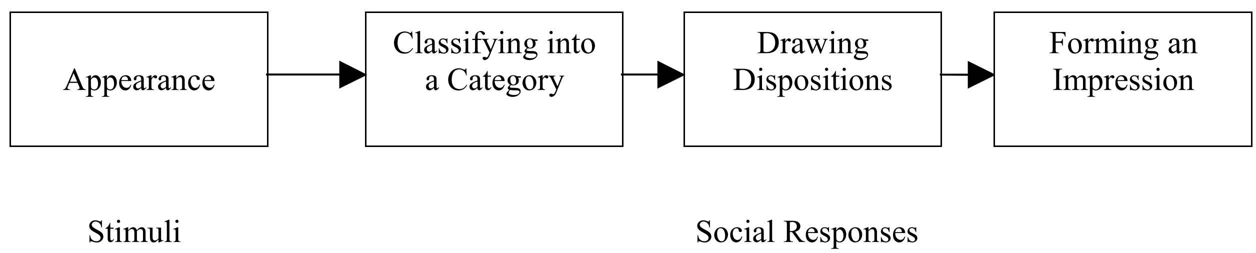


*Figure 1***.** A Framework for Ordinary Personology adapted from Gilbert (1998).

Humans inevitably categorize objects and people in their world (Allport, 1954). Categorizing an entity or an object as a human is not an exception (See Classifying in Figure 1). There are numerous studies on how individuals identify and recognize human faces from static images at the perceptual level (e.g., visual recognition of robots; for a review, see Zhao, Chellappa, Phillips, & Rosenfeld, 2003). However, there is no study on how different human-like cues (stimuli in Figure 1; e.g., voice, facial expression, showing of emotion and intelligence) play a role when individuals identify or categorize an entity or an object as a human. Perhaps it is because the process is so ingrained and natural to individuals that no one has bothered to question it. Furthermore, it may have held less value until now because individuals have not previously been faced with the problem of mis-categorizing (i.e., false negative) a human or over-categorizing (i.e., false positive) a non-human object, as can occur with virtual humans and robots. Ironically, the social interaction component of virtual human research has spurred this fundamental question (i.e., how do individuals identify an entity as a human and respond socially).

While Gilbert's aforementioned example assumes that the actor is a human and in fact involves a property of a person (i.e., his or her occupation), the framework can be equally applied to how individuals categorize any entity as a human. While a significant amount of prior research (Crocker, Major, & Steele, 1998; Fiske, 1998) has been conducted on how people draw dispositional inferences (a kind of social behavior) from appearance cues (e.g., Gilbert's example), the virtual human literature suggests that a few human-like cues (e.g., facial expressions, voice; See Chapter 3) elicit social behaviors.

Categorization and determining disposition are rapid and automatic due to their resource saving purposes (Fiske, 1998). This also applies to stereotyping, prejudice, and discrimination behaviors. Automaticity is the idea that sufficient practice with tasks maps stimuli and responses consistently and therefore produces performance that is autonomous, involuntary, unconscious, and undemanding of cognitive resources (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977; for a review, see Bargh, 1989). The automatic nature of social responses (categorization and determining disposition) is the result of constant pairing of stimuli and social behaviors and therefore inherits all of the aforementioned characteristics. However, forming an impression (See Figure 1) or making other social judgments is not necessarily an automatic process and might require both intention and attention (i.e., enough processing capacity; Bargh, 1989, 1990).

How Humans Respond Socially to Virtual Humans

The categorization and disposition processes (Figure 1) in the previous section can be equally applied when interacting with virtual humans (See Figure 2).

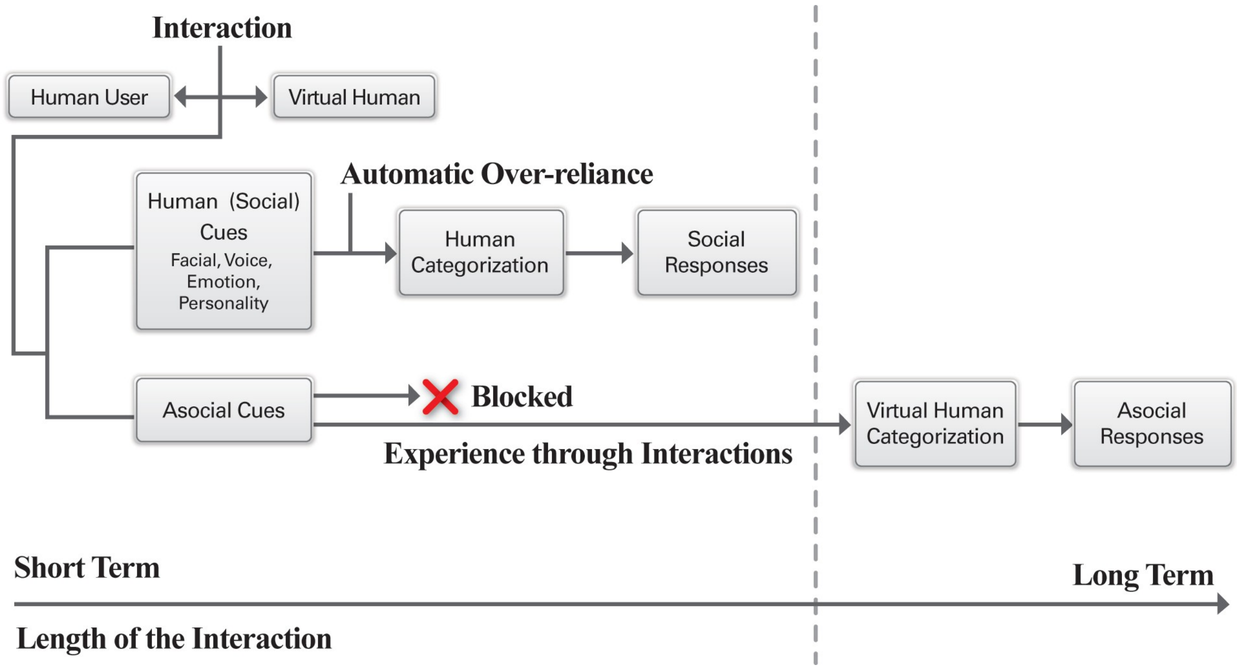


Figure 2. How humans respond socially to virtual humans. This, however, assumes a virtual human to be an imperfect human, as is the case at this time. That is, this framework might not be applicable to hypothetical androids[1], which are more human-like in terms of appearance and other human-like factors (personality, emotion).

Human-like characteristics (the use of facial expression or voice implementation in Figure 2) act as cues or stimuli that lead a person to place the virtual human into the category "human" (human categorization in Figure 2) and thus, elicit social responses. Specifically, a single cue may activate the human category and block other cues that would activate the virtual human category (automatic over-reliance in Figure 2). Hence, neither the virtual human per se nor the computer is the main reason individuals respond socially but rather the human-like cues possessed by the virtual human that elicit our social responses.

[1] A good example is replicants introduced in the 1982 science fiction film Blade Runner. A replicant is a bioengineered being that is virtually identical to an adult human. Humans cannot discern whether it is a human or a machine without a special device.

Nass and his colleagues' finding--that participants denied that they thought the computer was a human but nevertheless responded socially--seems to be comparable to Bargh's (1989) idea of "postconscious automaticity". In both cases the initial cues are conscious (i.e., human-like cues were clearly observable), yet their effects on category associations are not reportable (i.e., participants produce social behaviors when interacting with humans, yet explicitly deny considering the computer to be human-like).

The concept of mindless behavior (Langer, 1989), which is comparable to automatic processing, is the standard explanation that Nass and his colleagues have used in their attempt to understand why people respond socially to computers (Moon & Nass, 2000). Langer (1992) explicitly stated that her conception of mindlessness is different and even orthogonal to automatic processing on the grounds that, while automatic processing involves the repeated pairing of stimulus and response and consciousness is not precluded, mindlessness does not require repetition and cannot reach conscious awareness at all. While I want to avoid this debate between the two conceptions, one cannot deny that the processes of identifying and categorizing an entity as a human through the repeated pairing (i.e., practice) of human-like stimuli and human-like responses (e.g., a repetition of hearing a voice through phone and responding socially) is an automatic process. Interestingly, Bargh (1998), a prominent figure in automaticity research, seems to use the notion of automatic processing and mindlessness interchangeably (p. 464).

While I have answered why individuals respond to virtual humans as if they are humans by illuminating our over-reliance on automatic categorization of human-like cues, one might question why individuals ignore asocial cues (non human features; See Figure 2) of virtual humans. Based on the review of the literature in social psychology, there maybe at least two reasons: 1) the ingroup advantage and 2) lack of knowledge regarding virtual humans.

First, the ingroup advantage indicates that individuals respond faster to primes (stimuli) from the same group. For example, in sorting by gender, participants classify same-gender photographs faster than other-gender photographs (Zarate & Smith, 1990). This effect holds when they sort by job as well (Zarate & Sandoval, 1995). There is a great deal of evidence for the ingroup advantage (see Fiske, 1998), which can also be applied when responding to virtual humans. That is, humans might respond faster to cues from the same group (human -- i.e., virtual human's human cues) than a different group (non-human -- i.e., a virtual human's asocial cues) and as a result, lock into categorizing the virtual human as human. Once categorization is committed, the individual is oblivious to other stimuli (i.e., non-human cues)

Second, the lack of knowledge of virtual humans indicates that individuals do not have enough experience with virtual humans to form a rigid and concrete virtual human category. There are two causes of social behavior: proximal causes and distal causes. While proximal causes include perception, cognition, and affect at the individual level, a distal cause is the history of the individual in a specific set of contexts (i.e., individual history), the effect of which is mediated by the individual's internal processes and schemata (i.e., cognitive and affective structures) developed through experience (Feldman, 1999). Experiencing a virtual human and realizing how the technology is still far from perfectly mimicking a human and identifying the asocial nature of virtual humans might contribute to forming a schemata or a category of virtual human, where its disposition is non-human and non-social. A negative emotional experience with a virtual human due to its asocial characteristic might expedite this process (e.g., a frustration at an improper emotional expression by a virtual human that does not fit).

Accessibility, which also occurs through automaticity (Bargh, 1994), to an individual's declarative and procedural knowledge (e.g., knowledge of virtual humans' limitations, such as the fact that their emotional expression is simply an attempt to mimic real human emotion) at the time of the interaction is important as well (Feldman, 1999). This also suggests that individual differences in terms of expertise, skills, and the knowledge structures on virtual humans would influence the nature of social responses to them. This becomes especially important, because behaviors in various domains map onto their respective dispositions in different ways (Reeder & Brewer, 1979). People know and use a variety of domain-specific inference rules (Gilbert, 1998). If one is an expert (e.g., a developer) of a virtual human, then one would build up a separate set of inference rules that apply to virtual humans exclusively and independently of the rules one apply to humans. For example, a lag in human conversation may mean that the person is simply pondering or choosing their words carefully, but a lag in a virtual human conversation, to an expert's eyes, may mean that the system fell in an infinite loop, or the lag is a result of inefficient code.

Inevitability, length and time of the interaction (See length of the interaction in Figure 2) become critical factors influencing social interaction with virtual humans. The more time spent with a virtual human, the greater chance there is to identify the asocial nature of a virtual human. One will be able to deliberately map the stimulus (virtual human) and the responses (asocial responses in Figure 2) one would apply to a virtual human and not to a human. Uncertainty, ambiguity, or inconsistency of behavior will become a driving force (i.e., motivation), which will preclude categorization of the stimulus object or the production of behavior. These also prompt controlled attributional (as opposed to the aforementioned automatic) processing (Feldman, 1999) and lead one to start isolating a special class of invariant properties that distinguish one (e.g., virtual humans in general; virtual human categorization in Figure 2) from the other (e.g., humans in general; Jones & Davis, 1965). For example, users may feel inconsistency between the observed characteristics of a virtual human (i.e., at one point it seems to show human-like behavior but at other times, due to imperfection in the technology, it does not).

Nevertheless, the ability to categorize an entity as human is developed since birth and is strongly ingrained, so it may require considerable time and deliberate effort to avoid over-reliance on human categorization. The importance of the length of interaction on social responses implies limitations on Nass's experiments, where the interaction was exclusively short-termed. This also provides insights into long-term interactions, where users might begin a relationship with virtual humans.

Conclusion

In this chapter I argue that it is not the virtual human nor the computer per se that elicits social responses but the human-like characteristic that the entity possesses. Therefore, Nass's statement that "computers are fundamentally social" (p. 72, Nass et al., 1994) is misleading. Many papers failed to understand Nass's experiment properly and have blindly quoted this statement, without further explanation or caveats, as if people respond socially to any computer in any circumstance. In all of Nass's experiments, voice was used as a human-like characteristic and is one of the vital factors that elicited social response, not the computer itself. In short, the mere presence of a computer will not necessarily elicit a social response. However, the mere presence of a humanoid agent may elicit a social response, but only because it has facial expressions (a human-like characteristic) and not because it is a humanoid agent per se.

This chapter has reconstructed our understanding of the social responses to virtual humans, by incorporating the insights gained from social psychology (See Figure 2). Human-like characteristics act as cues or stimuli that leads a person to place the virtual human into the category "human" and thus, elicits social responses. These processes are practiced since birth and are strongly ingrained and have the characteristics of automaticity, such that a single cue may activate the human category and block other cues that would activate the virtual human category. Automaticity is not a conception restricted to short-term interaction but also applies to relationship cognition in long-term interactions.

Many factors such as knowledge, motivation, accessibility, and the time of interaction may play a contributing role in ascribing the cue to the virtual human category rather than the human category. This may have implications for long-term interactions with virtual humans and agents especially designed to support relationships with humans (e.g., relational agents).

Chapter 3: Virtual Humans Characteristics

In the previous chapter I have reached the conclusion that human-like characteristic is the key requirement in eliciting social responses from humans. What are the particular characteristics of virtual humans that encourage social responses? Illuminating such characteristics is not only critical to understanding all aspects of humans' social responses to virtual humans but also necessary to decide which characteristics to manipulate in this dissertation. The purpose of this chapter is to organize all human-like characteristics that are known to elicit social responses in the literature.

Investigating human-like cues naturally leads us to think what constitutes a human. Some characteristics that come to mind are the ability to talk, listen, express emotion, and look like a human. Virtual humans may have any combination of these implemented. I will organize the physical characteristics of virtual humans, including facial and voice embodiment (i.e., voice-over implementation), and the cognitive characteristics of virtual humans, including expression of emotion and personality. These characteristics are selected based on valid empirical evidence. Specifically, empirical studies are included only if there are: 1) a clear social response or social behavior involved with a virtual human, and 2) an observable human-like cue embodied in a virtual human.

Facial Embodiment

Facial embodiment not only includes the embodiment of a human-like face but also non-verbal facial behaviors such as facial expressions, raising eyebrows, the movement of the head and gaze. The effect of facial embodiment has been investigated in many studies. Yee, Bailenson, and Rickertsen (2007) provided a meta-analysis of the impact of human-like faces on user experiences in interfaces. While their focus is not entirely comparable to my focus on social responses to virtual humans, nevertheless, they provided some interesting findings. Yee et al. reviewed empirical studies (12 studies with performance measures and 22 studies with subjective measures) that compared interfaces with facial embodiment to interfaces without facial embodiment and found that the presence of a facial representation produced more positive interactions (i.e., increment in task performance, increase in liking) than the absence of a facial representation. This effect was found both in performance measures and subjective measures.

However, studies reviewed by Yee et al. (2007) did not necessarily involve social behaviors. In this chapter I organized the literature (See Table 1) investigating facial embodiment only if it had a social behavior involved with a virtual human. Literature that had significant limitations has been excluded.

Social Responses or Behaviors	**Human-like Characteristics**	**Results**	**References**
Ethnicity stereotyping; Same-ethnicity person perceived the others to be more attractive, trustworthy, persuasive, and intelligent.	Virtual human (**facial embodiment**)	Social rule holds.	(Nass, Isbister, & Lee, 2000)
Persuasiveness of a speaker (See Chapter 5 for details)	Virtual human (**facial embodiment**) with **voice**	Virtual humans are as effective as humans.	(Zanbaka, Goolkasian, & Hodges, 2006)
Punitive Ostracism; a group or individual ostracizes an individual as a form of punishment by using silent treatment or excommunication.	**Emotion** manifested in **facial expression**	Virtual humans are as effective as humans in simulating ostracism.	(Selvarajah & Richards, 2005)
Social Facilitation (See Chapter 5 for details)	Virtual human (**facial embodiment**)	Social rule partially applied (i.e., decrement in task performance)	(Rickenberg & Reeves, 2000)
Social Facilitation	Virtual human (**facial embodiment**)	Social rule holds.	(Park & Catrambone, 2007)

Table 1. Studies that investigated social responses in the presence of virtual humans that had facial embodiment as a human-like characteristic.

While facial embodiment in a few studies (e.g., Selvarajah & Richards, 2005; Zanbaka, Goolkasian, & Hodges, 2006) was confounded with other human-like characteristics, facial embodiment in the other studies (e.g., Park & Catrambone, 2007) was the sole characteristic embodied in a virtual human. Some studies, mostly Nass's studies, investigated social rules that had been well known in social psychology and tested participants interacting with virtual humans to see whether the social rules or the phenomena still hold. Other studies (e.g., Park & Catrambone, 2007) tested participants in the human condition (i.e., whether humans elicit social responses from participants) as well.

In experiments with a virtual human with a facial embodiment, the social rule (e.g., ethnicity stereotyping, social facilitation) was observable or the virtual humans were as effective as humans in simulating a social phenomenon (e.g., chaining attitudes through persuasion, simulating ostracism). That is, facial embodiment or facial expression is a human-like characteristic that elicits social responses. This is understandable because the human face is a powerful signal of human identity (Sproull et al., 1996). Infants exhibit a preference for face-like patterns over other patterns (Bond, 1972); they begin to differentiate facial features by the age of two months (Morton & Johnson, 1991). Faces induce appropriate social behaviors in particular situations. For example, covering people's faces with masks can produce inappropriate behaviors (Deiner, Fraser, Beaman, & Kelem, 1976).

Voice Embodiment

Many studies employed voice as a human-like characteristic to investigate a particular social response (See Table 2). For example, almost every study by Nass had voice implemented.

Social Responses or Behaviors	**Human-like Characteristics**	**Results**	**References**
Consistency-attraction	**Personality** manifested in **voice**	Social rule holds.	(Lee & Nass, 2005)
Emotional consistency (consistency of voice emotion and content emotion; People like content more when matched.)	**Emotion** manifested in **voice**	Social rule holds.	(Nass, Foehr, Brave, & Somoza, 2001)
Gender stereotypes (e.g., praise from males is more convincing than praise from females)	Text messages with synchronized **voice**	Social rule holds.	(Nass, Steuer, & Tauber, 1994; Nass, Moon, & Green, 1997)
Notion of "self" and "other" (e.g., other-praise is more valid than self-praise, other-praise is friendlier than self-praise)	Text messages with synchronized **voice**	Social rule holds.	(Nass, Steuer, & Tauber, 1994)
Persuasiveness of a speaker	Virtual human (**facial embodiment**) with **voice**	Virtual humans are as effective as humans.	(Zanbaka, Goolkasian, & Hodges, 2006)
Persuasiveness of a speaker	**Voice** (computerized text-to-speech)	Just as effective as human voice.	(Stern, Mullennix, Dyson, & Wilson, 1999)
Persuasiveness of a speaker	**Voice** (computerized text-to-speech)	Just as effective as human voice.	(Mullennix, Stern, Wilson, & Dyson, 2003)
Politeness norms; direct, as opposed to	Text messages with	Social rule holds.	(Nass, Steuer, &

indirect, requests for evaluations elicit more positive responses and more homogeneous responses.	synchronized **voice**		Tauber, 1994; Nass, Moon, & Carney, 1999)
Similarity-attraction	**Personality** manifested in **voice**	Social rule holds.	(Lee & Nass, 2005)

Table 2. Studies that investigated social responses in the presence of virtual humans that had voice as a human-like characteristic.

While some studies (e.g., Stern, Mullennix, Dyson, & Wilson, 1999; Mullennix, Stern, Wilson, & Dyson, 2003) investigated the effect of voice exclusively, other studies investigated characteristics such as emotion or personality manifested in voice (Lee & Nass, 2005; Nass, Foehr, Brave & Somoza, 2001). Regardless, in experiments with a voice embodied in a virtual human, the social rule (e.g., gender stereotypes, politeness norm) was observable or the virtual humans were as effective as humans in engendering a social phenomenon (e.g., changing attitudes through persuasion). That is, a human-like cue via voice is sufficient to encourage users to apply social rules. Individuals unconsciously applied social rules: 1) despite consciously knowing that virtual humans do not have personality (Lee & Nass, 2003), and 2) despite being presented with a voice that was not human (e.g., synthesized voice; Nass et al., 2001).

Emotion

Emotion plays a powerful role in people's lives. Emotion impacts our beliefs, informs our decision-making, and affects our social relationships. Humans have developed a range of behaviors that express emotional information as well as an ability to recognize other's emotions.

However, there is no universally accepted definition of emotion, nor is there a general agreement as to which phenomena are accepted as the manifestation of emotion (Gratch & Marsella, 2005). Some theories support a set of distinct emotions with neurological correlates (Ekman, Friesen, & Ellsworth, 1972) whereas others argue that emotions are epiphenomenal, simply reflecting the interaction of underlying processes (LaDoux, 1996). The relationship between emotion and other psychological constructs such as feeling, mood, and personality is also unclear (Gratch & Marsella, 2005). Nevertheless, emotion, as a psychological construct, plays a major role in the social interaction with virtual humans because emotions are crucial for effective communication (Brave & Nass, 2002) and social reasoning (Forgas, 1991; Frank, 1999).

There is a an explosion of interest in the role of emotion in the design of virtual humans where most research is directed at making virtual humans seem more convincing, believable, intelligent, and enjoyable (Bartneck, Reichenbach, & Van Breemen, 2004). However, a handful of studies have investigated social responses or social behaviors of users when emotion is manifested through virtual humans (See Table 3).

Social Responses or Behaviors	**Human-like Characteristics**	**Results**	**References**
Emotional consistency (consistency of voice emotion and content emotion;	**Emotion** manifested in **voice**	Social rule holds.	(Nass, Foehr, & Somoza, 2001)

People like content more when matched.)			
Punitive Ostracism; a group or individual ostracizes an individual as a form of punishment by using silent treatment or excommunication.	**Emotion** manifested in **facial expression**	Virtual humans are effective in simulating ostracism.	(Selvarajah & Richards, 2005)

Table 3. Studies that investigated social responses in the presence of virtual humans that had emotion as a human-like characteristic.

In the two studies, emotion was manifested either by voice (Nass et al., 2001) or facial expression (Selvarajah & Richards, 2005). Individuals also express emotion through other channels such as gestures or text (i.e., language; Gratch & Marsella, 2005) so further studies might manifest emotion by these channels.

There are a few papers on how to design a believable agent through implementing emotion (Bates, 1994; Grates & Marsella, 2005; Thomas & Johnston, 1981). For example, Thomas and Johnston (1981) claimed that to increase the believability of virtual humans, the emotional state of the virtual humans must be clearly defined and accentuated. However, believability is an unclear concept. No research has clearly conceptualized or operationalized such a construct. It is also difficult to consider believability as a synonym for people's social responses to virtual humans. As I have addressed in the previous section, a human-like cue will elicit social responses to virtual humans, which are automatic and generally unconscious (Bargh, 1989). That is, people might not consciously believe (i.e., believability) that the virtual human is human-like when showing social behaviors.

Personality

Personality is defined by behavioral consistency within a class of situations that is defined by the person (Feldman, 1999). A few studies employed personality as a human-like characteristic to investigate a particular social response (See Table 4). In the present context, of interest is whether people respond to virtual humans as if virtual humans have a personality. In experiments with a virtual human with a personality employed, the social rule (e.g., consistency-attraction, similarity-attraction) was observable.

A number of studies suggest that there are at least three factors that play a role in the social interaction with virtual humans involving personality, including characteristics of textual content (Moon & Nass, 1996; Nass, Moon, Fogg & Reeves, 1995), character appearance (Isbister & Nass, 2000), and character behavior (Ball & Breese, 2000; Reeves & Nass, 1996). Nevertheless, we do not have empirical evidence that the latter two elicit social responses (See Table 4).

Social Responses or Behaviors	**Human-like Characteristics**	**Results**	**References**
Consistency-attraction	**Personality** manifested in **voice**	Social rule holds.	(Lee & Nass, 2005; Moon, 1998)
Gain-loss theory; Changes in the direction of similarity had a more positive effect on attraction than a consistently similar personality.	**Personality** manifested in text messages	Social rule holds.	(Moon & Nass, 1996)
Self-serving bias; When personalities matches, individuals are more likely to give the other credit for success and less likely to blame for failure.	**Personality** manifested in text messages	Social rule holds.	(Moon & Nass, 1998)
Similarity-attraction hypothesis; People	**Personality** manifested	Social rule holds.	(Nass, Moon, Fogg,

prefer to interact with others who are similar in personality.	in text messages		Reeves, & Dryer, 1995)
Similarity-attraction	**Personality** manifested in **voice**	Social rule holds.	(Lee & Nass, 2005)

Table 4. Studies that investigated social responses in the presence of virtual humans that had personality as a human-like characteristic.

Integration

In this section I have organized all human-like characteristics that are known to elicit social responses. Four characteristics are identified through empirical evidence as human-like characteristics: personality, emotion, facial embodiment, and voice (See Figure 3). While facial embodiment and voice implementation can each directly elicit social responses (e.g., personality-neutral content through voice elicited social responses; Lee & Nass, 2003), personality and emotion presumably work indirectly through facial expression, voice, or specific word choices.

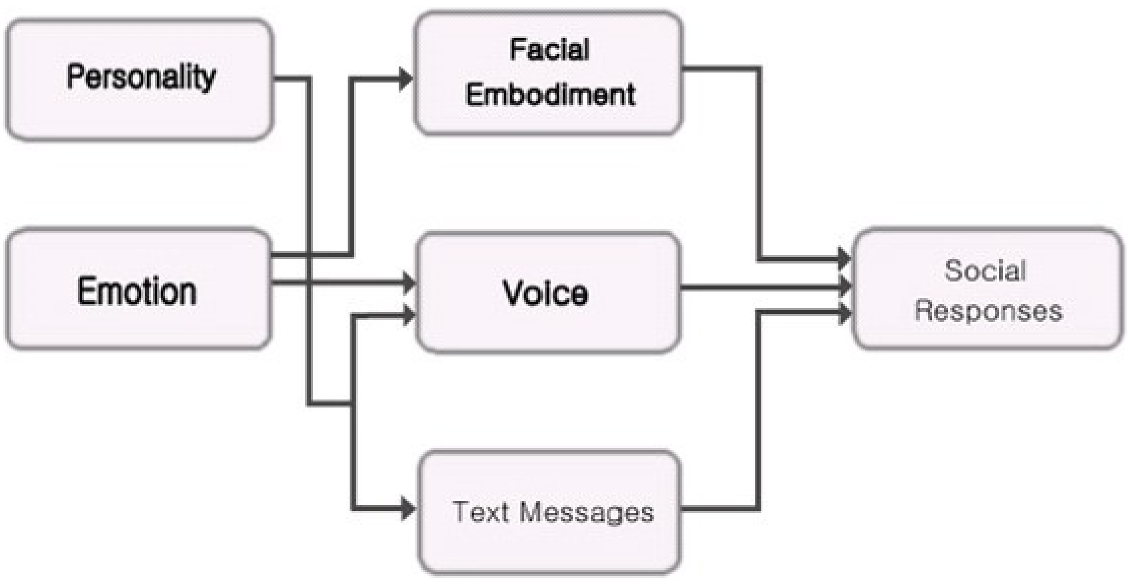


Figure 3. Human-like characteristics that are known to elicit social responses. Text messages or language work only as a channel toward social responses.

There are other variables such as intelligence or autonomy that need further investigation. King and Ohya (1996) found that human forms presented on a computer screen are assessed as more intelligent than simple geometric forms. Paiva et al. (2004) suggested that autonomy is one of the key features for the achievement of believability. Nevertheless, none of the studies have provided empirical evidence that virtual humans embodied with such constructs elicit social responses.

Apparently, there is a close relationship between emotion and personality. Emotion is short-lived, whereas personality remains stable over a long period of time (Moffat, 1997). Emotion focuses on particular events or objects, whereas personality is more diffuse and indirect. Due to this relationship, several projects have built an integrated model of emotion and personality. For example, Ball and Breese (2000) modeled dependencies between emotion and personality in a Bayesian network. Prendinger and Ishizuka (2001) modeled the relationship between a virtual human's social role and the associated constraints on emotional expression. Emotion is also closely related to time. In my framework, emotion plays a stronger role in long-term interactions due to the history of interaction between users and virtual humans. For example, a negative emotional experience (a frustration at an improper emotional expression by a virtual human that does not fit) with a virtual human due to its asocial characteristic may influence subsequent interactions.

Chapter 4: Framework of Social Responses and Hypotheses

The previous chapters reconstructed our understanding of the social responses to virtual humans, by integrating empirical evidence gained from HCI and the insights gained from social psychology. Figure 4 presents the complete framework.

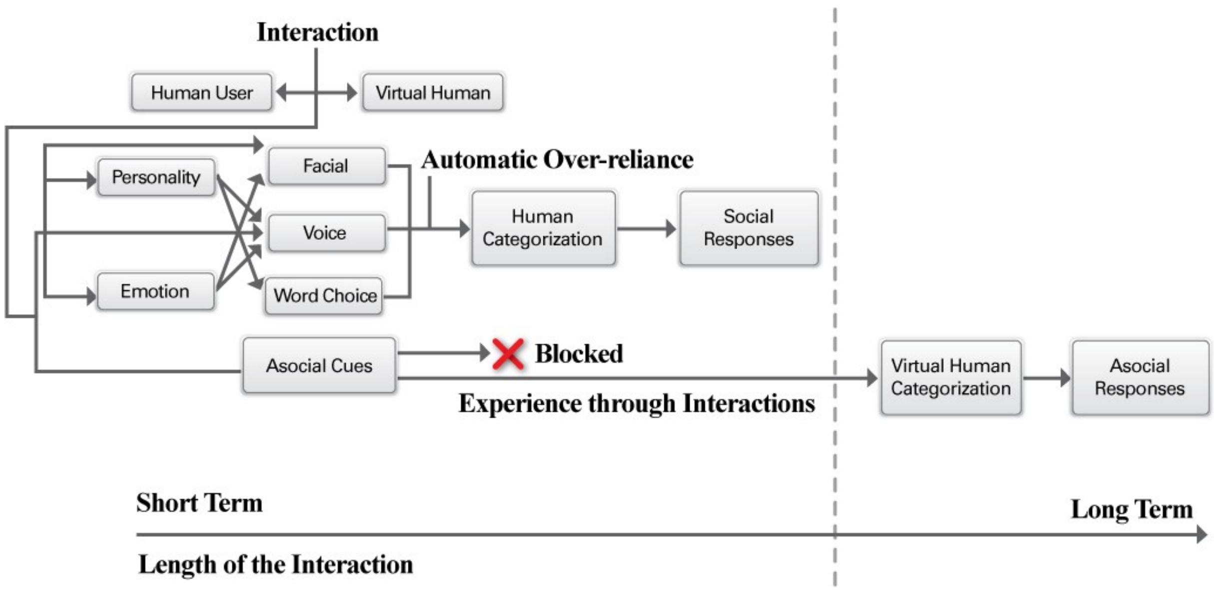


Figure 4. The complete framework of how humans respond socially to virtual humans. This, however, assumes a virtual human to be an imperfect human, as is the case at this time. That is, this framework might not be applicable to hypothetical androids, which are more human-like in terms of appearance and other human-like factors (personality, emotion).

Human-like characteristics of virtual humans act as a cue that leads a person to place the agent into the category "human" and thus, elicits social responses. Once categorization has occurred, the user is less likely to notice asocial cues (e.g., awkward behavior of a virtual human). Four characteristics are identified as human-like: personality, emotion, facial embodiment, and voice. While facial embodiment or voice implementation is enough to elicit social responses (e.g., personality-neutral content through voice elicited social responses; Lee &

Nass, 2003), personality and emotion should be manifested through facial expression, voice, or text messages. The processes of social responses are so ingrained through practice that a human-like cue will set off social responses, which are automatic and therefore generally unconscious (Bargh, 1989). However, many factors such as knowledge, motivation, accessibility, and the time of interaction may play a contributing role in ascribing the cue to the virtual human category rather than the human category. For example, the more time spent with a virtual human, the greater chance there is to identify the asocial nature of a virtual human. That is, one will deliberately map the stimulus (virtual human) and the responses one would apply to a virtual human and not to a human (Virtual Human Categorization in Figure 4).

Social Responses as a Function of the Number of Human-like Characteristics

The framework of social responses to virtual humans provides the basis for insights to outstanding questions that have been raised in the research community. One such question is whether humans are more likely to elicit social responses if virtual humans present more human-like characteristics (Moon & Nass, 2000). No study has empirically answered this question. In fact, as depicted in chapter 3, most studies had only one human-like characteristic (emotion, personality, voice, facial embodiment) implemented. A few studies that had a combination of human-like characteristics were not designed to test the additive effect of multiple human-like characteristics.

Paiva et al. (2004) suggested that it is not so much one factor that matters, but instead the combination of all factors together provides ingredients for the believability. However, social responses are not necessarily a result of believability. As outlined in chapter 2, social responses to virtual humans can occur regardless of consciousness (e.g., believing virtual humans are not humans).

In this dissertation I hypothesize that as long as a clearly observable human-like characteristic of a virtual human exists, then the number of human-like characteristics does not affect the strength of social response. The framework of how humans respond socially to virtual humans (Figure 4) states that social response is a result of a human placing the agent into the category “human”. The majority of psychological studies of categorization have used cues that vary on a few (2-4) dimensions (e.g., shapes, colors; Berretty, Todd, & Blythe, 1997) with most target items (i.e., items to be classified by participants) ranging from living entities (e.g., cats, dogs) to nonliving entities (e.g., chair, sofa). Given that all prominent categorization models (exemplar, decision bound, and neural network models) assume that the cues may be presented as points in a *multidimensional* space, the number of cues matters especially when classifying an entity that has a competing entity (e.g., distinguishing an alligator from a crocodile). That is, more distinguishing cues (e.g., alligators are smaller, lighter, and have more teeth) make classification more accurate.

However, the category “human” does not have a competing category when a clear observable characteristic (e.g., facial expression) is presented. For example, the human face is a powerful, unique, and exclusive signal of human identity (Sproull et al., 1996) -- a clearly distinguishing cue from other identities. Therefore, once users respond to a single human-like cue of a virtual human and lock into categorizing the virtual human as human, it might no longer require additional human-like characteristic to strengthen this commitment to categorization.

Chapter 5: Social Facilitation

I conducted two social psychological experiments to test the hypothesis described in the previous section: a social facilitation experiment and a politeness norm experiment. Both experiments involved tasks that were known to elicit social responses to virtual humans in literature as described in this chapter.

Social Facilitation

Humans behave differently and presumably process information differently, when there is someone else nearby versus when they are alone. This is referred to as the social facilitation phenomena in social psychology. More specifically, social facilitation is generally referred to as performance enhancement on a simple or well-learned task and as performance impairment on a complex or novel task.

Only a handful of studies have investigated whether social facilitation can be produced by virtual humans. Walker, Sproull, and Subramani (1994) investigated participants' responses to a synthesized talking face displayed on a computer screen in the context of a questionnaire study. Compared with participants who answered questions presented via a text display on a screen, participants who answered the same questions spoken by a talking face spent more time, made fewer mistakes, and wrote more comments.

Walker et al. (1994) claimed that this enhancement in task performance was attributable to social facilitation. However, one premise of the social facilitation effect is that performance is facilitated only if the task is simple or well learned. The researchers never explicitly stated or demonstrated whether the questionnaire task in their study was a simple task. In addition, spending more time with the talking face did not necessarily enhance task performance. This might simply mean that it took longer to listen to a question than to read it, and the study did not address this issue.

Rickenberg and Reeves (2000) investigated the effects of different presentations of virtual humans on user anxiety, task performance, and subjective evaluations in the context of web sites. Users felt more anxious when virtual humans monitored their web site work, and this effect was the strongest for users with an external control orientation. The presence of a monitoring character decreased task performance.

Zanbaka, Ulinski, Goolkasian, and Hodges (2004) attempted to investigate social facilitation attributable to virtual humans. Participants first learned a task and were then randomly assigned to perform the same task or a novel task alone, in the presence of a human, or a virtual human. In general, Zanbaka et al. (2004) were unable to replicate the social facilitation effect. As they noted in their paper, a ceiling effect may have been one reason their research failed. Participants were able to learn the correct pattern in the learning stage, which left little room for improvement. This is also a common problem in social facilitation research in social psychology (Bond & Titus, 1983).

Recently, two studies (Park & Catrambone, 2007; Zanbaka, Ulinski, Goolkasian, & Hodges, 2007) used simple and complex tasks rather than learned and novel tasks for their experiments. The results were inconsistent. While Park and Catrambone (2007) were successful in producing social facilitation (both enhancement and impairment) in the presence of a virtual human as with a human, Zanbaka et al. (2007) were able to show only impairment in the presence of a virtual human or a human but not facilitation in both conditions.

One shortcoming in Zanbaka et al.'s (2007) study is the tasks. While all three tasks (anagram, maze, and modular arithmetic) used in Park and Catrambone's (2007) study had a strong grounding in social facilitation research with firm evidence of social facilitation with humans (Davidson & Henderson, 2000 for anagram tasks; Rajecki, Ickes, Corcoran, & Lenerz, 1977 for maze tasks; Beilock, Kulp, Holt, & Carr, 2004 for modular arithmetic tasks), this was not the case with Zanbaka et al.'s study (2007). This is a potential reason why they were not able to produce social facilitation, even in a human condition. In other words, the use of unproven social tasks with a virtual human is problematic, in contrast to Nass's research paradigm (Nass et al., 1994), where he used only well established tasks from social psychology.

It is possible to speculate that the social facilitation of virtual humans reported by Park and Catrambone (2007) was due to the virtual human being an animated graphical object rather than due to the facial characteristics of the virtual human. Therefore, it is necessary to add an animated graphical object that is not a virtual human (i.e., lacks human-like characteristics) as a condition in future experiments regarding social facilitation with virtual humans.

Chapter 6: Social Facilitation Experiment

In Chapter 5 I reviewed the literature on social facilitation in the context of virtual humans. Virtual humans do produce the social facilitation effect (Park and Catrambone, 2007). For easy tasks, performance in the virtual human condition was better than in the alone condition, and for difficult tasks, performance in the virtual human condition was worse than in the alone condition. In this study, I was interested in the strength of social facilitation as a function of varying kinds of human-like characteristics. Such characteristics of virtual humans that encourage social responses were reviewed in Chapter 3. I was also interested in whether the number of human-like characteristics affects the strength of social facilitation. This chapter will detail the methods and procedures of an experiment designed to answer this question.

Tasks

Social facilitation experiments must involve cognitive tasks for the participants to do. This section describes how such tasks were selected and how each task was executed in the experiment. The experimental tasks needed both breadth and depth to test the social facilitation effect but, at the same time, needed to be applicable to the realm of virtual humans. Hence, the following two criteria were used for selecting experimental tasks: (a) Is the task something that a user might do with the assistance of a virtual human, and (b) Is the task scalable in terms of difficulty?

With respect to the first criterion, virtual humans can assist users in many different tasks, but based on prior research it seems likely that virtual humans would be most helpful with high level cognitive tasks (Catrambone, Stasko, & Xiao, 2004). Some tasks can be opinion-like (e.g., choosing what to bring on a trip), and others can be more objective (e.g., implementing edits in a document). With respect to the second criterion, the present study examined differences in task performance between simple and complex tasks so it should be possible to produce both easy and difficult instances of the tasks.

Using these two criteria, the present study used the following three cognitive tasks: anagrams, mazes, and modular arithmetic. These three tasks provided a good mixture of verbal, spatial, mathematical, and high-level problem solving skills. All three tasks were cognitive tasks and had an objective, and therefore they were within the range of tasks with which a virtual human might assist. It was also possible to produce both easy and difficult instances of these tasks.

Anagram task

Social facilitation (attributable to a human being present) in anagram tasks has been studied in the context of electronic performance monitoring (EPM), a system whereby every task performed through an electronic device may be analyzed by a remotely located person (Davidson & Henderson, 2000). The social facilitation effect was clearly observed in the presence of EPM, easy anagrams being performed with greater proficiency and difficult anagrams being performed with less proficiency. In the present study, anagrams were divided into two categories (easy or difficult) using normative solution times from Tresselt and Mayzner's (1966) anagram research (see also Davidson & Henderson, 2000).

Maze task

Research has suggested that participants tend to perform better in the presence of a human on simple maze tasks (Rajecki et al., 1977). In the present study, simple mazes included wide paths and few blind alleys so that the correct route was readily perceivable, whereas difficult mazes included narrow paths with many blind alleys. Materials for the maze task were similar to the ones of Jackson and Williams (1985). Participants were given a maze and a cursor on the screen and were asked to draw a path to the exit.

Modular arithmetic task

The object of Gauss's modular arithmetic is to judge if a problem statement, such as "50 ≡ 22 (mod 4)," is true. In this case, the statement's middle number is subtracted from the first number (i.e., 50 – 22) and the result of this (i.e., 28) is divided by the last number (i.e., 28 ÷ 4). If the quotient is a whole number (as here, 7), then the statement is true. Difficulty of the task was manipulated by controlling the number of digits presented to participants for the first two numbers of a given problem; one for an easy task (7 ≡ 2) and two for a difficult task (51 ≡ 19).

Beilock et al. (2004) claimed that modular arithmetic is advantageous as a laboratory task because it is unusual and, therefore, its learning history can be controlled. In the modular arithmetic tasks, problem statements were given to the participants. Easy problems consisted of single-digit no-borrow subtractions, such as "7 ≡ 2 (mod 5)"; hard problems consisted of double-digit borrow subtraction operations, such as "51 ≡ 19 (mod 4)." These were similar to the ones of Beilock et al. (2004).

Participants

One hundred eight participants were recruited from the Georgia Institute of Technology. Participants were compensated with course credit.

Materials

Participants did all tasks (anagrams, maze, and modular arithmetic) on a computer. Java application and Java script were used to implement the tasks on the computer. An additional computer was used to present the virtual human and the graphical shape. For the facial virtual human condition, Haptek Corporation's 3-D character was loaded on this computer (see Figure

5); the appearance of the virtual human was held constant. The character displayed lifelike behaviors, such as breathing, blinking, and other subtle facial movements. The graphical shape rotated slowly during the experiment (see Figure 6).

Figure 5. Virtual human in the present study

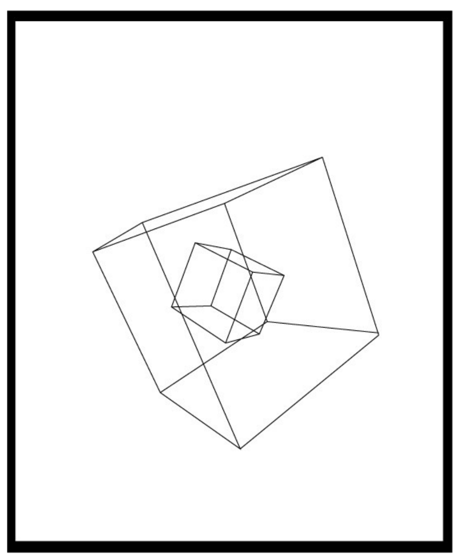


Figure 6. Graphical shape in the present study

The monitor was positioned so that the virtual human and the graphical shape was oriented to the task screen, not to the participant, and was located about 4 feet (~1.2 m) from the task monitor and about 3.5 feet (~1 m) from the participant. This is also the location where the human observer would sit.

Design and Procedure

The present study was a 6 (virtual human type: facial appearance, voice recordings, voice synced facial appearance, facial appearance conveying emotion, voice recordings conveying emotion, voice synced facial appearance conveying emotion) × 3 (task type: anagram task, maze task, modular arithmetic task) × 2 (task complexity: easy, hard) × 4 (presence: alone, presence of a human, presence of a virtual human, and presence of a graphical shape) mixed design. The presence of a graphical shape depicts the minimal visual interaction with the participant to make sure that the effect of virtual human is due to its human-like characteristics and not due to the mere presence of a graphical component.

All participants did both easy and hard versions of all task types in all presence conditions. Only the appearance of a virtual human was a between-subject factor. Each participant experienced one of the virtual human types in Table 5.

The type of virtual human	**Facial appearance**	**Voice implementation**	**Emotion conveyed**	**Deliverance of cover story**
A: Facial appearance only	Yes	No	Neutral	By text
B: Voice recordings only	No	Yes	Neutral	By voice
C: Voice synced facial appearance	Yes	Yes	Neutral	By voice

D: Facial appearance conveying emotion	Yes	No	Happy	By text
E: Voice recordings conveying emotion	No	Yes	Happy	By voice
F: Voice synced facial appearance conveying emotion	Yes	Yes	Happy	By voice

Table 5. The types of virtual humans manipulated in the social facilitation experiment.

The two within subjects factors (task complexity, presence) were crossed to produce eight types of trials, in which participants did:

- a simple task alone
- a simple task in the presence of a human
- a simple task in the presence of a virtual human
- a simple task in the presence of a graphical shape
- a complex task alone
- a complex task in the presence of a human
- a complex task in the presence of a virtual human
- a complex task in the presence of a graphical shape.

Every participant experienced multiple instances of each of the eight trial types.

The order of the presence factor was varied across participants using a Latin square. That is, some participants did the first set of tasks in the presence of a human (H), the next set in the presence of the graphical shape (G), the third set in the presence of the virtual human (VH), and the fourth set alone (A). The other three orders were:

- G → V → A → H
- V → A → H → G
- A → H → G → V

Within a particular presence situation (e.g., virtual human), participants did a block of anagrams, a block of mazes, and a block of modular arithmetic problems. Task order was manipulated using a Latin square resulting in three possible orders:

- anagram → maze → modular arithmetic
- maze → modular arithmetic → anagram
- modular arithmetic → anagram → maze

Within each task block, participants conducted a combination of easy and hard trials for that particular task (e.g., anagrams). The number of easy and hard trials was the same in each block; however, the order of easy and hard trials was one of the three predetermined pseudo-randomized orders.

In the anagram tasks, a five-letter anagram appeared on the screen, and the participants were asked to solve the anagram quickly and accurately by typing in the answer using the keyboard and then pressing the Enter key (See Appendix A). Completion time and error rates were measured. Feedback was not given.

In the maze tasks, a maze appeared on the screen. Participants were asked to move the cursor by dragging the mouse through each maze and to find the exit as fast as possible (See Appendix B). Completion time was measured.

In modular arithmetic tasks, a problem statement, such as "50 ≡ 20 (mod 4)," appeared on the screen. Participants were asked to decide whether the statement was true or false by pressing the corresponding button (*Y* for "true," *N* for "false") on the keyboard (See Appendix C). Completion time and error rates were measured. Feedback was not given.

Each participant was briefed on each task. Briefing consisted of a demonstration by the experimenter and four hands-on practice trials for the participants so that they could familiarize themselves with the computer and the task.

For conditions involving a human, a virtual human, or a graphical shape, the participants were told that a human, a virtual human, or a graphical shape is there to "observe" the task, not the participant. Specifically, when a human was present, participants were told, "An observer will be sitting near you to observe the tasks you will doing. The observer will be present to learn more about the tasks and try to catch any mistakes we made in creating the tasks. The observer is not trying to learn how you go about working on the tasks and, in fact, will not be allowed to communicate with you while it is sitting here."

When a virtual human with only facial appearance (type A in Table 5) or with only facial appearance conveying emotion (type D) or a graphical shape was present, participants were told by the experimenter, "A virtual human (graphical shape) will observe the task. The virtual human (graphical shape) is an artificial intelligence that attempts to analyze events that happen on the computer screen. The virtual human (graphical shape) will be present to learn more about the tasks and try to catch any mistakes we made in creating the tasks. The virtual human (graphical shape) is not trying to learn how you go about working on the tasks and, in fact, will not be allowed to communicate with you while it is present."

The virtual human presented the passage with his own voice for the rest of the virtual human types (type B, C, E, and F in Table 5). The CSLU (Center for Spoken Language and Understanding) Toolkit was used to create a computer generated baseline which the human recorder of the voice can reference. The setting of F_0 = 184 Hz, pitch range = 55 Hz, and word rate = 198 words per minute for the happy voice (type E and F) and F_0 = 137 Hz, pitch range = 30 Hz, and word rate = 177 words per minute for the neutral voice (type B and C) were used. These criteria are from the literature on the markers of emotion in speech (Cahn, 1990). Manipulation checks indicated that the happy recording was recognized as happy voice and the neutral recording was recognized as neutral voice.

Emotion was also conveyed in facial appearances (type D and F; see Figure 6 for comparison in terms of appearance). Although Haptek's agent software provided a mean to manipulate basic emotion (e.g., happy, sad, anger), it was not based on any emotional model or literature, thus manipulation checks were conducted. Participants indicated that the happy facial expression was recognized as happy and the neutral facial expression was recognized as neutral.

Figure 6. Screenshot of the virtual human with a happy and a neutral expression.

Results

A four-way repeated measures ANOVA (virtual human factor, task factor, complexity factor, and presence factor) was initially conducted and was followed by post hoc analyses. Analysis was conducted on completion time because it has been the most frequently measured dependent variable in social facilitation research (Bond & Titus, 1983). The pseudo order and Latin square factors were tested to examine whether they had an effect on performance. Pseudo order and the Latin square orders had no effect on task performance, and all results were collapsed over these variables.

Data were transformed into *z* scores for each task to perform the analysis involving complexity, presence, and task. The results (summarized in Figures 7 to 9) show that the effect of presence on completion time was conditional upon the combination of the task, task complexity, and virtual human type, resulting in a significant four-way interaction of Presence x Task Type x Complexity x Virtual Human Type, $F(30, 612) = 1.54$, $MSE = .26$, $p < .05$. Of particular importance, the results show that combined across task types (anagram, maze, and modular arithmetic), if the task was easy, completion times in the presence of the virtual human and the real human tended to be faster than in their absence or the graphical shape, whereas if the task was hard, the mean completion times were slower in the presence of the virtual human and the real human than in their absence. This observation is supported by a Presence x Complexity interaction that is consistent with the social facilitation effect, $F(3, 306) = 75.65$, $MSE = .27$, $p < .001$.

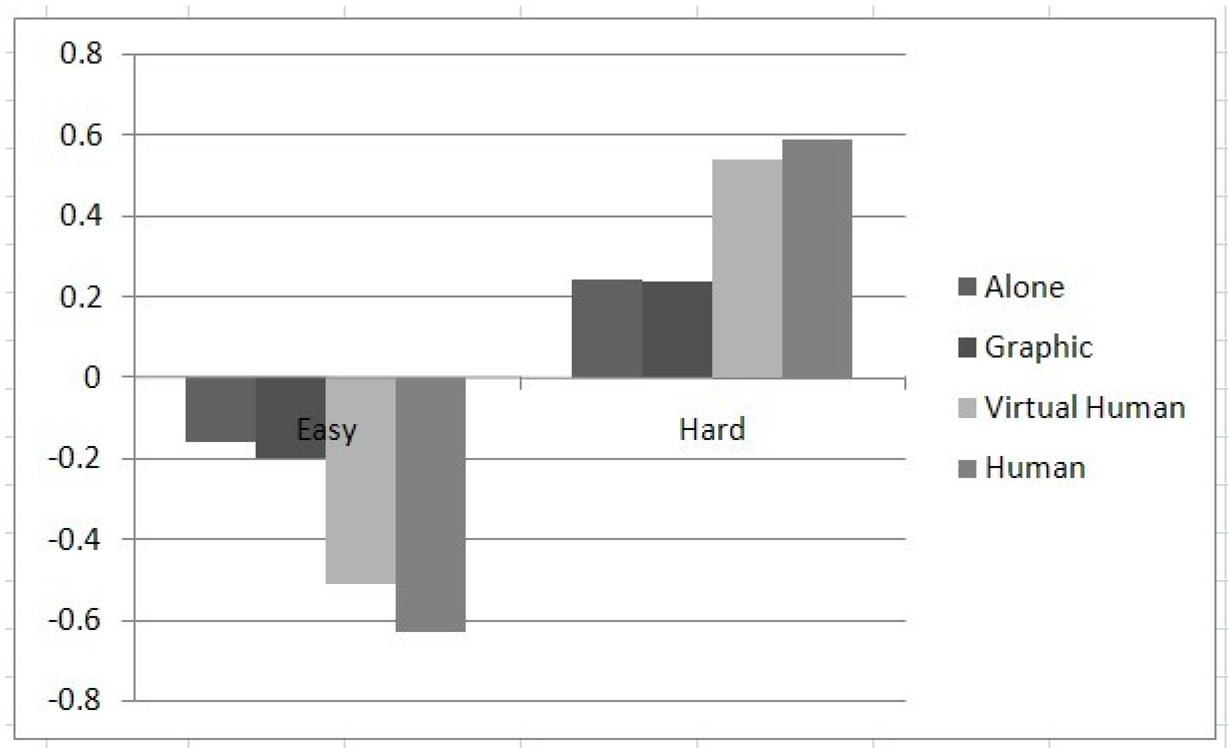


Figure 7. Anagram Task. Mean completion time in seconds for the task (n = 108).

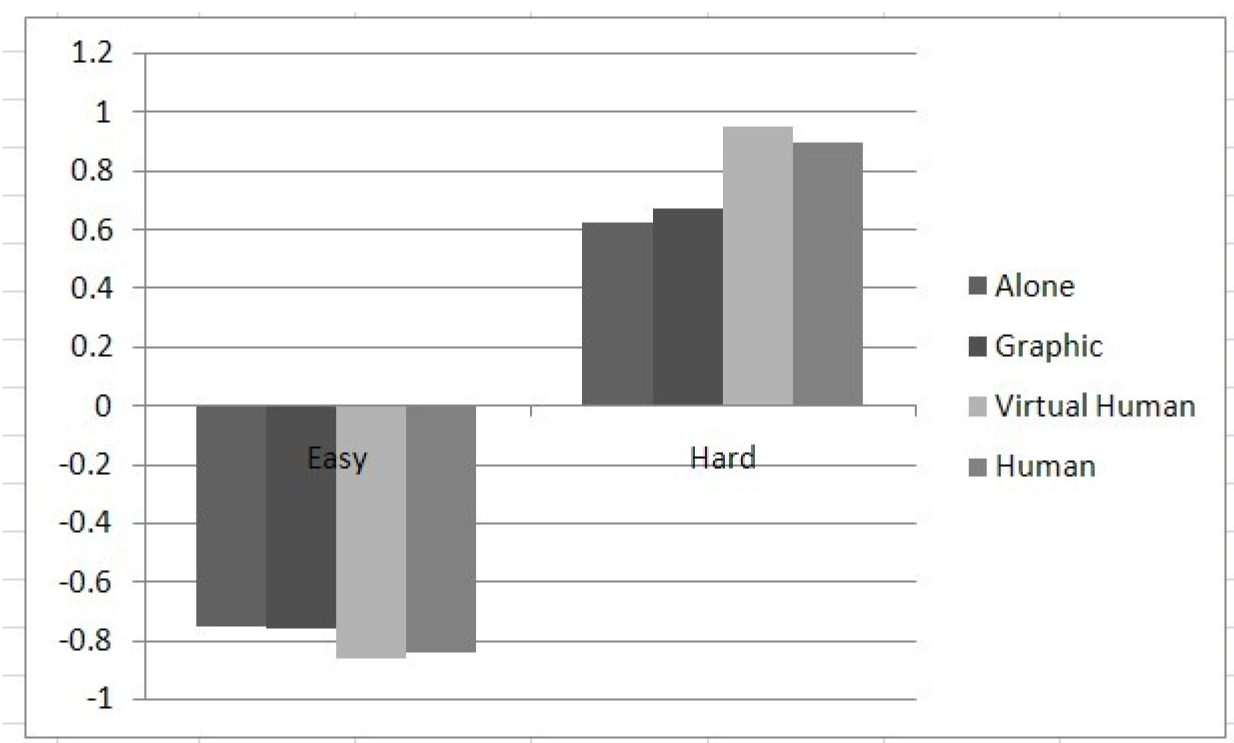


Figure 8. Maze Task. Mean completion time in seconds for the task (n = 108).

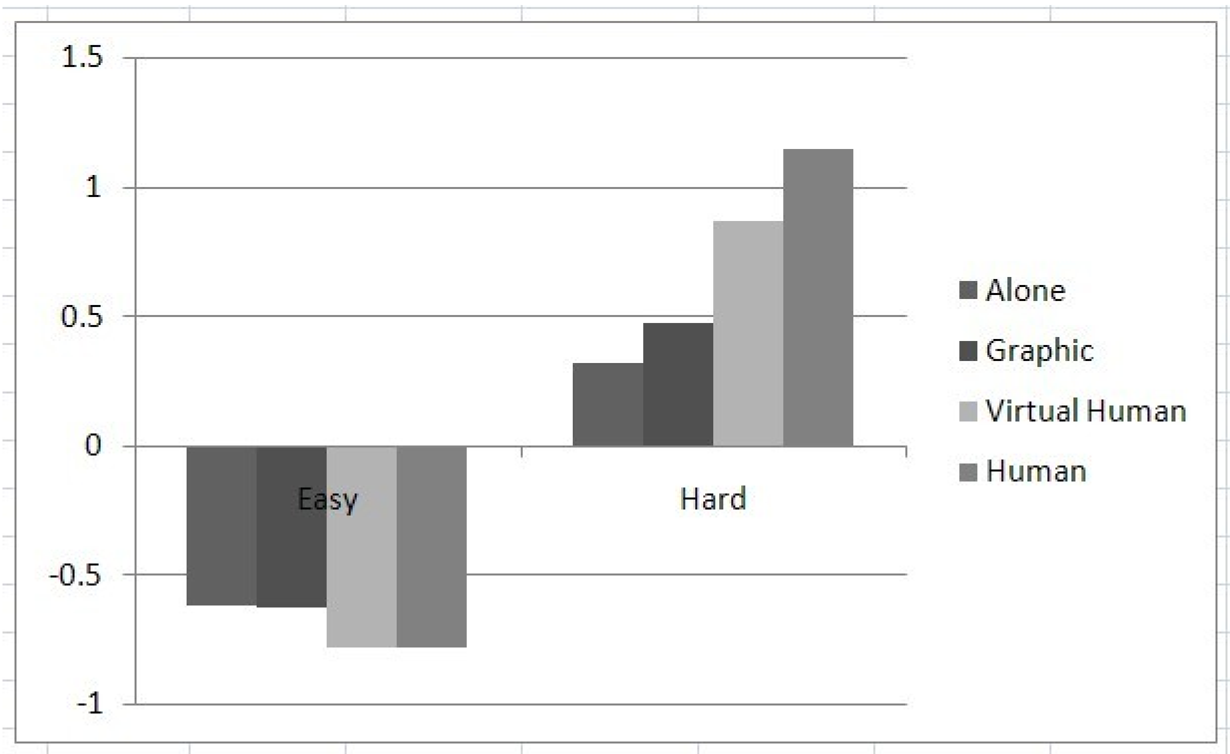


Figure 9. Modular Arithmetic Task. Mean completion time in seconds for the task (n = 108).

There was also a main effect of complexity (easy, hard), $F(1, 102) = 872.36$, $MSE = 1.18$, $p < .001$, and presence (alone, virtual human, human, graphical shape), $F(3, 306) = 10.32$, $MSE = .25$, $p < .001$ but no main effect of task type, $F < 1$.

Post Hoc Analyses for Each Task Type

A three-way Presence x Task x Complexity interaction, $F(6, 612) = 3.89$, $MSE = .26$, $p < .005$, suggests that each task type should be further analyzed for the relationship between presence and complexity. I conducted a post hoc Dunnett's test to compare each presence condition to the alone condition separately for each task type. For each task type, the social facilitation effect for virtual humans was demonstrated and, for each task type, the social facilitation effect for humans was demonstrated. The analyses for each of these observations are presented next.

Virtual human verses alone. For all three tasks (anagram, maze, and modular arithmetic), pairwise comparisons show that completion time for easy tasks was shorter in the virtual human condition than in the alone condition, and that completion time for hard tasks was longer in the virtual human condition than in the alone condition (see Figure 7 to 9); anagram easy: $t(214) = 5.07$, $p < 0.001$; anagram hard: $t(214) = 4.86$, $p < 0.001$; maze easy: $t(214) = 2.00$, $p < .05$; maze hard: $t(214) = 4.14$, $p < 0.001$; modular arithmetic easy: $t(214) = 2.43$, $p < 0.05$; modular arithmetic hard: $t(214) = 7.86$, $p < 0.001$.

Human versus alone. For all three tasks, pair-wise comparisons show that completion time for easy tasks was shorter in the human condition than in the alone condition, and completion time for hard tasks was longer in the human condition than in the alone condition (see Figure 7 to 9); anagram easy: $t(214) = 7$, $p < 0.001$; anagram hard: 4.86, p < 0.001; maze easy: $t(214) = 2.23$, $p < .05$; maze hard: $t(214) = 4.68$, $p < 0.001$; modular arithmetic easy: $t(214) = 2.42$, $p < 0.05$; modular arithmetic hard: $t(214) = 11.72$, $p < 0.001$.

Post Hoc Analyses for Each Virtual Human Type

Voice synced facial appearance versus human. Of interest was whether social response to a voice synced facial appearance is similar to social response to a human. For all three tasks, pair-wise comparisons show that completion time for both easy and hard tasks was statistically the same with the human condition *with an exception of maze easy task*; anagram easy: $t(34) < 1$; anagram hard: t(34) < 1; maze easy: $t(34) < 1$; maze hard: $t(34) = 4.17$, $p < 0.001$; modular arithmetic easy: $t(34) < 1$; modular arithmetic hard: $t(34) < 1$.

Facial appearance versus human. Of interest was whether social response to a face only virtual human is similar to social response to a human. For all three tasks, pair-wise comparisons show that completion time for both easy and hard tasks was statistically different from the human condition *with the exception of the modular easy task and the hard task*; anagram easy: $t(34) = 2.5$, $p < 0.05$; anagram hard: $t(34) = 3.0$, $p < 0.01$; maze easy: $t(34) = 2.47$, $p < .05$; maze hard: $t(34) = 2.0$, $p < 0.05$; modular arithmetic easy: $t(34) > 1$; modular arithmetic hard: $t(34) > 1$.

Voice recordings versus human. Of interest was whether the social response to a voice only virtual human is similar to the social response to a human. For all three tasks, comparisons show that completion time for both easy and hard tasks was statistically different from the human condition *with the exception of the maze easy task*; anagram easy: $t(34) = 3.25$, $p < 0.05$; anagram hard: $t(34) = 2.25$, $p < 0.05$; maze easy: $t(34) = 1.17$, $p > .05$; maze hard: $t(34) = 2.50$, $p < 0.05$; modular arithmetic easy: $t(34) = 2.13$, $p < 0.05$; modular arithmetic hard: $t(34) = 3.00$, $p < 0.01$.

Facial appearance conveying emotion versus human. Of interest was whether the social response to a facial appearance conveying emotion is similar to the social response to a human. For all three tasks, comparisons show that completion time for both easy and hard tasks was statistically different from the human condition *with the exception of the anagram easy and the modular arithmetic easy*; anagram easy: $t(34) > 1$; anagram hard: $t(34) = 2.25$, $p < 0.05$; maze easy: $t(34) = 2.41$, $p < .05$; maze hard: $t(34) = 2.17$, $p < 0.05$; modular arithmetic easy: $t(34) < 1$; modular arithmetic hard: $t(34) = 2.10$, $p < 0.05$.

Voice recordings conveying emotion versus human. Of interest was whether social response to a voice recording conveying emotion is similar to social response to a human. For all three tasks, pair-wise comparisons show that completion time for both easy and hard tasks was statistically different from the human condition with *an exception of anagram easy and modular arithmetic easy*; anagram easy: $t(34) > 1$; anagram hard: $t(34) = 2.17$, $p < 0.05$; maze easy: $t(34) = 2.23$, $p < .05$; maze hard: $t(34) = 2.41$, $p < 0.05$; modular arithmetic easy: $t(34) > 1$; modular arithmetic hard: $t(34) = 2.10$, $p < 0.05$

Voice synced facial appearance conveying emotion versus human. Of interest was whether social response to a voice synced facial appearance conveying emotion is similar to social response to a human. For all three tasks, pair-wise comparisons show that completion time for both easy and hard tasks was statistically the same from the human condition; anagram easy: $t(34) > 1$; anagram hard: $t(34) > 1$; maze easy: $t(34) > 1$; maze hard: $t(34) > 1$; modular arithmetic easy: $t(34) > 1$; modular arithmetic hard: $t(34) > 1$.

In general, the social facilitation experiment showed that as with a human, virtual humans do produce social facilitation with all six virtual human types. In addition, performance in the presence of voice synced facial appearance and voice synced facial appearance conveying emotion seems to elicit stronger social facilitation (i.e., no statistical difference compared to performance in the human presence condition) than in the presence of voice only or face only. Conveying emotion on voice only or face only did not produce this additional finding.

Chapter 7: Politeness Norm Experiment

Overview

Thus far data from the social facilitation experiment showed that as with a human, virtual humans can produce social facilitation with all virtual human type conditions. In addition, combing voice with face seems to elicit stronger social facilitation (i.e., no statistical difference compared to performance in the human presence condition) than in the presence of voice only or face only. Conveying emotion did not produce this additional finding.

The purpose of the second experiment was first, to expand this results to a different social behavior (or social response) known in the literature; this behavior will be the "politeness norm". The second purpose was to expand the results in a more ecologically valid setting where some amount of interaction might exist.

In the social psychology literature, socially desirable response bias in interview situations has been well documented (Jones, 1964; Singer, Frankel & Glassman, 1983; Sudman & Bradburn, 1974). One type of social desirability effect occurs when people bias their responses according to the perceived preference of the interviewer. For example, when person A is asked by person B to evaluate person B (direct evaluation), person A generally applies politeness norm and give a positive response to avoid offending Person B. On the other hand, if person C were to ask person A to evaluate person B (indirect evaluation), then person A tend to give a more honest (likely more negative) response.

Studies by Nass (Nass, Moon, & Carney, 1999; Nass, Steuer, & Tauber, 1994) on whether people apply the politeness norm to computers provided some insights in designing this study. In this experiment, of interest is whether direct, as opposed to indirect, requests for evaluations of virtual humans elicit more positive responses (i.e., politeness norm). In addition, this experiment collected data of people's responses to a human tutor as a basis for comparison.

To accomplish this, participants were tutored and quizzed either by a virtual human tutor that varied in terms of features or a human tutor. Participants then evaluated the tutor's performance either directly by the tutor or indirectly via a paper and pencil questionnaire.

Participants

Eighty participants were recruited from the Georgia Institute of Technology. Participants were compensated with course credit.

Materials

Participants did the experiment on a computer. HTML was used to implement the tasks on the computer. Audio (voice only condition) or video clips (facial appearance only and face with voice conditions) of Haptek Corporation's 3-D character were imbedded in the HTML code.

Design

The present study was a 4 (tutor: virtual human with facial appearance only, virtual human with voice only, virtual human with voice synced facial appearance, human) x 2 (evaluation: direct, indirect) between-subject design. The virtual human conditions are identical to the social facilitation experiment except conditions that conveyed emotion (see Table 6). Such conditions were eliminated because conveying emotion did not produce any clear effects in the first experiment.

The type of virtual human	**Facial appearance**	**Voice implementation**

A: Facial appearance only	Yes	No
B: Voice recordings only	No	Yes
C: Voice synced facial appearance	Yes	Yes

Table 6. The types of virtual humans manipulated in the politeness norm experiment

Procedure

The experiment consisted of four sessions: tutoring session, testing session, scoring session, and interview session. When arrived, participants were briefed with the general procedure of the experiment with a prepared script (see Appendix D).

Tutoring session. In this session participants were told that they will learn facts about national geography. The tutor was either a virtual human with varying characteristics (see Table 6) or a human tutor based on the condition each participant was assigned to. The tutor presented each participant with 20 facts (see Appendix E). The participant then answered to a question about whether they knew about the fact by clicking a button labeled with "I knew the fact" or a button labeled with "I didn't know the fact". Participants were told that they would receive 20 out of 1,000 facts and that facts will be chosen based on their familiarity of a subject as assessed by their performance on earlier facts. This is to provide a sense of interactivity with the virtual humans (Nass & Steuer, 1993). In reality, all participants received the same 20 facts in the same order.

Testing session. The tutor then administered a 12-item multiple choice test with 5 options per item. The participants were told that 12 questions would be randomly drawn from a total of 5,000 questions. In reality, all participants received the same 12 questions in the same order.

Scoring session. During this session, the tutor provided the participant's score as well as evaluation on their own performance. All participants were informed that they had answered the same 10 of 12 questions correctly and received identical and positive evaluation of the tutor ("The tutor performed well."; see Appendix G for the exact script). A constant and fixed score (10 out of 12) was presented to prevent individual scores from affecting the evaluation and become a confounding variable. However, it was also important to not let participants realize this fake scoring. Facts and questions were designed so that the score is acceptable by the participants. Manipulation checks indicated participants were not surprised with the score and believed in it.

Interview session. Following the completion of the scoring session, participants were asked questions (e.g., competence, likable, fair) about the performance of the tutor (see Appendix H). If the tutor was a virtual human then this interview was conducted by either the same virtual human (direct evaluation) or through a paper-and-pencil questionnaire (indirect evaluation). If the tutor was a human then the interview was conducted by the same human tutor (direct evaluation) or through a paper-and-pencil questionnaire (indirect evaluation).

Results

I averaged all 13 items in the interview questionnaire to a single unit for each participant to capture each participant's overall valence toward the tutor. Consistent with the interview questionnaire (see Appendix H), higher scores on this unit indicated more positive responses toward the tutor.

A two-way between subjects ANOVA (tutor and evaluation factor) was conducted and was followed by a post hoc Dunnett's test. Analysis was conducted on the overall valence toward the tutor (i.e., the average of all 13 items). The results (summarized in Figure 8) show that combined across tutor types, consistent with the politeness prediction, participants gave more positive responses when the tutor asked about its own performance (direct evaluation) as compared to when participants answered on paper-and-pencil (indirect evaluation). This observation is supported by a main effect of Evaluation (direct, evaluation), $F(1, 79) = 10.27$, $MSE = 12.74$, $p < .01$. There was also a main effect of Tutor (virtual human with facial appearance only, virtual human with voice only, virtual human with voice synced facial appearance, human), $F(3, 237) = 11.94$, $MSE = 14.81$, $p < .001$, but no Evaluation by Tutor interaction, $F < 1$.

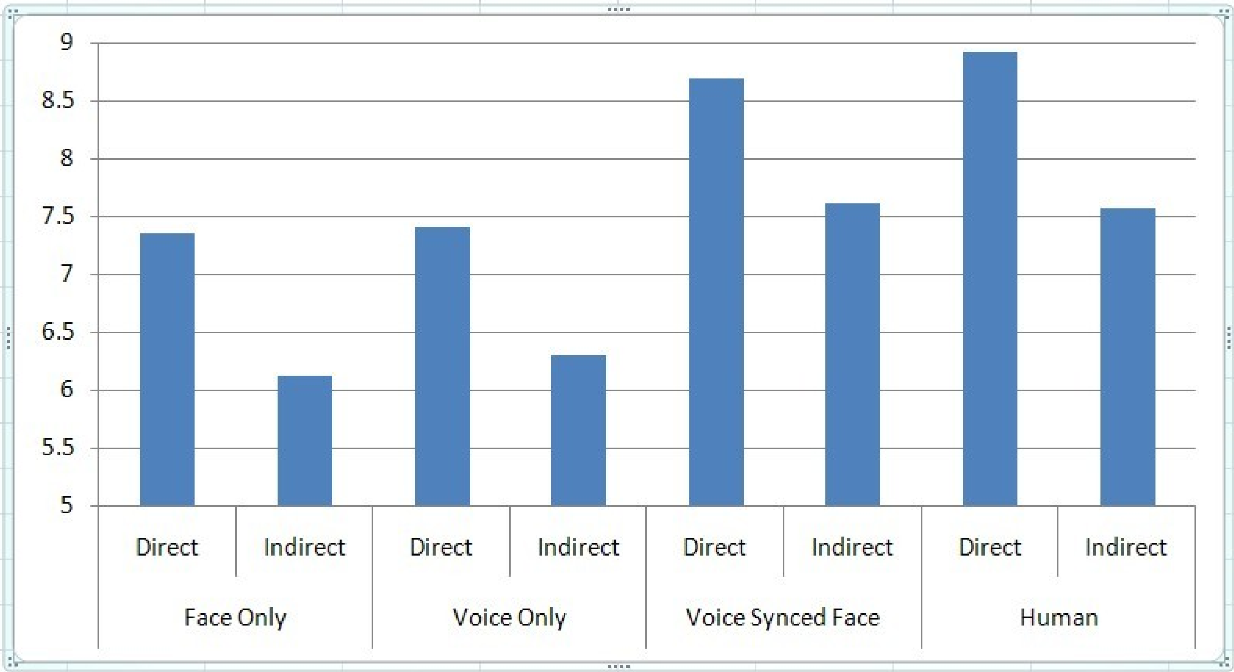


Figure 8. Mean valence toward the tutor for each valuation type (direct, indirect) for each tutor type (face only, voice only, voice synced face, human) (n = 80).

Post Hoc Analyses

Direct versus indirect. The main effects of Tutor and Evaluation suggest that each tutor type should be further analyzed to see whether participants applied politeness norm to all tutor types. I conducted a post hoc Dunnett's test to compare each direct condition to the indirect condition separately for each tutor type. For each tutor type, the politeness norm was demonstrated; face only: $t(72) = 2.46$, $p < 0.05$; voice only: $t(72) = 2.22$, $p < 0.05$; voice synced face: $t(72) = 2.16$, $p < 0.05$; human: $t(72) = 2.72$, $p < 0.01$.

Virtual human versus human. Of interest was whether each social response to the virtual human is similar to that of social response to human. The valence toward the voice synced face condition had no statistical difference to the valence toward the human tutor condition for both direct and indirect conditions, $t < 1$. However, this valence toward the voice synced face voice was significantly higher than the valence toward the voice only condition for both direct and indirect conditions, $t(72) = 2.6$, $p < 0.05$. There was no significant difference between the voice only condition and the face only condition for both direct and indirect conditions, $t < 1$.

Results from the politeness experiment indicate that, consistent with politeness prediction, participants who evaluated their tutor directly reported the tutor's performance more favorably than participants who evaluated their tutor indirectly. This was observed not only with the human tutor but with virtual human tutors with all four conditions (virtual human with facial appearance only, virtual human with voice only, virtual human with voice synced facial appearance). In addition, consistent with the social facilitation experiment, the valence toward the voice synced facial appearance had no statistical difference compared to the valence toward the human tutor condition. In other words, when voice is combined with face (i.e., voice synced face) an addictive effect occurs where the social responses to virtual humans with such features are indistinguishable from the social responses to humans.

Chapter 8: Conclusions

The purpose of this dissertation was to gain a deeper understanding of the social dimension of the interaction (i.e., social interaction) between users and virtual humans. This dissertation illuminated the meaning of "being social" when interacting with virtual humans. That is, I have reconstructed our understanding of the social responses to virtual humans, by incorporating the insights gained from the person perception or ordinary personology literature. Human-like characteristics of virtual humans act as a cue that leads a person to place the agent into the category "human" and thus, elicits social responses. Once categorization has occurred, the user is less likely to notice asocial cues (e.g., awkward behavior of a virtual human). Four characteristics are identified as human-like: personality, emotion, facial embodiment, and voice. The processes of social responses are so ingrained through practice that a human-like cue will set off social responses, which are automatic and generally unconscious (Bargh, 1989).

Given this framework, this dissertation was designed to answer the following questions that have implications for the design of avatars or virtual humans: 1) If a virtual human has more human-like characteristics, will it elicit stronger social responses from people? 2) How do the human-like characteristics interact in terms of the strength of social responses?

Given the frequent high cost and limited time to design a virtual human, it is important to understand the necessary and minimal requirements to elicit social response from the user. I conducted two social psychological experiments to answer the aforementioned questions. The first experiment investigated whether virtual humans can evoke a social facilitation response and how strong that response is when participants are given different cognitive tasks (e.g., anagrams, mazes, modular arithmetic) that varied in difficulty. The second experiment investigated whether people apply politeness norms to virtual humans; participants were tutored and quizzed either by a virtual human tutor that varied in terms of features or a human tutor. Participants then evaluated the tutor's performance either directly to the tutor or indirectly via a paper and pencil questionnaire.

Results from both studies showed that as with a human, virtual humans can elicit social behaviors (social facilitation and politeness norm) with only facial appearance but also with voice recordings. In addition, performance in the presence of voice synced facial appearance seems to elicit stronger social facilitation (i.e., no statistical difference compared to performance in the human presence condition) than in the presence of voice only or face only. Conveying emotion did not produce this additional finding.

The results reconfirmed the framework for understanding the social responses to virtual humans that facial expression and voice act as cues that lead a person to place the agent into the category "human" and thus, elicit social responses. In addition, the results indicate that such social responses were engendered as a function of the type of virtual human. Social responses to virtual humans with voice or face only were significantly different from the social responses to humans. On the other hand, social responses to virtual humans with voice synced face were not significantly different from the social responses to humans.

This suggests that there are minimal requirements for humans to classify a non-biological entity to a human and response socially (See Figure 9). In this experiment we learned that voice and facial expression are such characteristics. There are other potential human-like factors (e.g., sign of intelligence, language) that I did not investigate in this study because of the lack of literature (as described in Chapter 3).

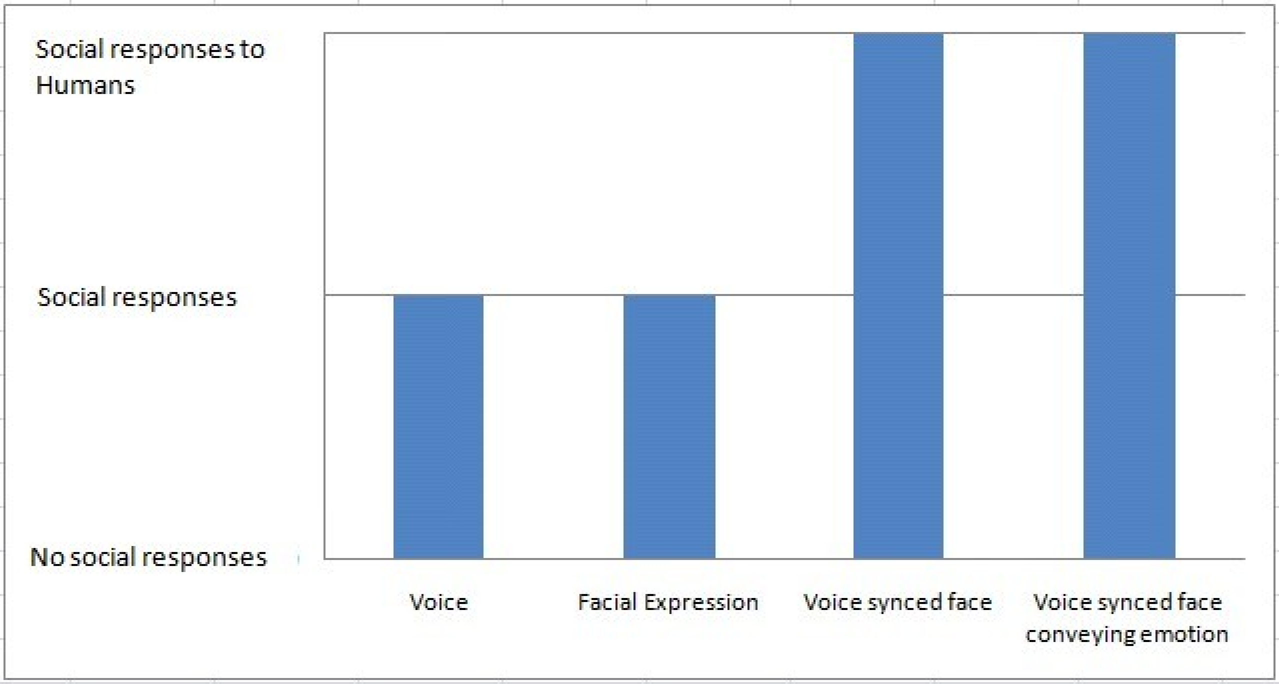


Figure 9. Social responses as a function of human-like characteristics.

A combination of voice and face (i.e., voice synced face) in both studies proved that not only they elicited social behaviors but also elicited such behaviors that were not significantly different from how humans behave to humans. Does this mean that humans behave equally to any virtual humans with voice synced face? There are many different social norms and social behaviors so this reasoning is questionable and should be further studied. As described in Chapter 2, history and length are critical factors in any social interaction so further analysis on the long-term effects of social interactions with virtual humans is required.

Nevertheless, the results have implications for the design of instructional systems as well as for other systems involving human-computer interaction. They suggest that designers of such systems should consider that users behave differently in the presence of virtual humans, as compared with when they are alone.

The study also provides insights to some practical question in terms of implementation especially when a designer intends to build a social component into a virtual human and expects social responses from users. For example, given the limited resources (e.g., time, cost), which characteristic should designers implement first? A design decision to present a virtual human should be a deliberate and thoughtful one.

References

Allport, G. W. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.

Bargh, J. A. (1989). Conditional automaticity: varieties of automatic influence in social perception and cognition. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 3-51). NY: Guilford.

Bargh, J. A. (1990). Goal is not intent: goal-directed thought and behavior are often unintentional, *Psychological Inquiry*, 1(3), 248-251.

Bartneck, C., Reichenbach, J., & Van Breemen, A. (2004). In your face, robot! The influence of A character's embodiment on how users perceive its emotional expressions, *Proceedings of the design and emotion conference*.

Ball, G. & Breese, J. (2000). Emotion and personality in conversational agents. In J. Cassell, J. Sullivan, S. Prevost, & E. Churchill (Eds.), *Embodied Conversational Agents*, Cambridge, MA: MIT Press, 189-219.

Bates. (1994). The role of emotion in believable agents. *Communications of the ACM*, *37(7)*, 122-125.

Beilock, S. L., Kulp, C. A., Holt, L. E., & Carr, T. H. (2004). More on the fragility of performance: Choking under pressure in mathematical problem solving. *Journal of* Experimental Psychology: General, 133, 584–600.

Berretty, P., Todd, P., & Blythe, P. W. (1997). Categorization by Elimination: A fact and frugal approach to Categorization. *Proceedings of the nineteenth annual conference of the cognitive science society.*

Bickmore, T. & Picard, R. (2004). Towards Caring Machines, *Proceedings of the ACM Computer Human Interaction 2004 Conference*.

Bond, E. K. (1972). Perceptions of form by the human infant. *Psychological Bulletin*, *77*, 225-245.

Brave, S. & Nass, C. (2002). Emotion in human-computer interaction. In J. Jacko & A. Sears, *Handbook of Human-Computer Interaction*, 251-271. NY: Lawrence Erlbaum Associates.

Cahn, J. (1990). The generation of affect in synthesized speech. *Journal of American voice I/O Society,* 8, 1-19.

Catrambone, R., Stasko, J., & Xiao, J. (2004). ECA user interface paradigm: Experimental findings within a framework for research. In C. Pelachaud & Z. Ruttkay (Ed.), *From brows to trust: Evaluating embodied conversational agents*, 239-267. NY: Kluwer Academic/Plenum Publishers.

Crocker, J., Major, B., & Steele, C. (1998). Social stigma. In D. Gilbert, S. Fiske & G. Lindzey (Ed.), *The Handbook of Social Psychology (Volume Two)* (pp. 89-150). NY: Oxford University Press.

Davidson, R., & Henderson, R. (2000). Electronic performance monitoring: a laboratory investigation of the influence of monitoring and difficulty on task performance, mood state, and self-reported stress levels. *Journal of Applied Social Psychology, 30,* 906–920.

Deiner, E., Fraser, S. C., Beaman, A. L., & Kelem, R. T. (1976). Effects of deindividuating variables on stealing by Halloween trick-or-treaters. *Journal of Personality and Social Psychology*, *33*, 178-183.

Eagly, A. H. (1978). Sex differences in influenceability. *Psychological Bulletin*, *85*, 86-116.

Eagly, A. H. & Wood, W. (1982). Inferred sex differences in status as a determinant of gender stereotypes about social influence, *Journal of Personality and Social*

Psychology.

Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). *Emotion in the human face: Guidelines for Research and an integration of findings*. New York: Pergamon Press.

Feldman, J. M. (1999). Four questions about human social behavior. In J. Adamopoulos and Y. Kashima (Eds.) *Social Psychology and Cultural Context: Essays in Honor of Harry C. Triandis*. NY: Sage.

Fiske, S. (1998). Stereotyping, prejudice, and discrimination. In D. Gilbert, S. Fiske & G. Lindzey (Ed.), *The Handbook of Social Psychology (Volume Two)* (pp. 89-150). NY: Oxford University Press.

Fogg, B. J. & Tseng, H. (1999). The elements of computer credibility. *Proceedings of the ACM Computer Human Interaction 1999 Conference*.

Forgas, J. P. (1991). Affect and social judgment: an introductory review. In J. P. Forgas (Ed.), *Emotion and Social Judgments*, 3-29. Oxford: Pergamon Press.

Frank, R. (1988). *Passions With Reasons: The Strategic Role of the Emotions*, NY: W. W. Norton.

Gilbert, D. (1998). Ordinary Personology. In D. Gilbert, S. Fiske & G. Lindzey (Ed.), *The Handbook of Social Psychology (Volume Two)* (pp. 89-150). NY: Oxford University Press.

Glantz, K., Rizzo, A., & Graap, K. (2003). Virtual reality for psychotherapy: Current reality and future possibilities. *Psychotherapy: Theory, Research, Practice, Training*, *40,* 55–67.

Gratch, J. & Marsella. (2005). Lessons from Emotion Psychology for the Design of Likelife Characters, *Applied Artificial Intelligence*, 19, 215-233.

Hill, R. W., Gratch, J., Marsella, S., Rickel, J., Swartout, W., & Traum, D. (2003). Virtual humans in the mission rehearsal exercise system. *Kynstlich Intelligenz, 17*(4), 32–38.

Jackson, J. M., & Williams, K. D. (1985). Social loafing on difficult tasks: Working collectively can improve performance. *Journal of Personality and Social Psychology, 49,* 937–942.

Jeong, S., Hashimoto, N., & Makoto, S. (2004). A novel interaction system with force feedback between real – and virtual human: An entertainment system: "Virtual catch ball." In *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology* (pp. 61–66). New York: Association for Computing Machinery Press.

Jones, E. E. (1964). *Ingratiation: A social psychological analysis*. New York, NY: Meredith.

Jones, E. & Davis, K. (1965). From acts to dispositions: the attribution process in person perception. In L. Berkowitz (eds), Advances in Experimental Social Psychology Vol. 2, 225, NY: Academic Press.

King, W. J. & Ohya, J. (1996). The representation of agents: Anthropomorphism, agency and intelligence. *Proceedings of CHI 1996 Conference*, 289-290.

LaDoux, J. (1996). *The emotional brain: The mysterious underpinnings of emotional life*. New York: Simon & Schuster.

Langer, E. J. (1989). *Mindfulness*. Reading, MA: Addison-Wesley.

Langer, E. J. (1992). Matters of mind: Mindfulness/mindlessness in perspective. *Consciousness and cognition*, 1, 289-305.

Lee, K. M. & Nass, C. (2005). Social-psychological origins of feelings of presence: Creating social presence with machine-generated voices. *Media Psychology*, 7, 31-45.

Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., & Bhogal, R. S. (1997).

The Persona Effect: Affective Impact of Animated Pedagogical Agents, *ACM Computer Human Interaction 1997 Conference*, 359-366.

Mullennix, J. W., Stern, S., Wilson, S. & Dyson, C. (2003). Social perception of male and female computer synthesized speech. *Computers in Human Behavior*, 19, 407-424.

Moffat, D. (1997). Personality Parameters and Programs. *Creating Personalities for Synthetic Actors*, NY: Springer Verlag, 120-165.

Moon, Y. & Nass, C. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, 56(1), 81-103.

Moon, Y. (1998). Intimate self-disclosure exchanges: Using computers to build reciprocal relationships with consumers. Cambridge, MA: Harvard Business School. Report: 99-059.

Morton, J. & Johnson, M. H. (1991). CONSPEC and CONLERN: A two-process theory of Infant face recognition. *Psychological Review*, *98*, 164-181.

Mullennix, J. W., Stern, S. E., Wilson, S. J., & Dyson, C. (2003). Social perception of male and female computer synthesized speech, *Computers in Human Behavior, 19(4)*, 407-424.

Nass, C. & Brave, S. (2005). *Wired for speech: How voice activates and advances the Human-computer relationship*. Cambridge, MA: MIT Press.

Nass, C., Foehr, U., Brave, S., & Somoza, M. (2001). The effects of emotion of voice in synthesized and recorded speech. *Proceedings of the AAAI Symposium Emotional and Intelligent II: The Tangled Knot of Social Cognition*. North Falmouth, MA.

Nass, C. & Lee, K. M. (2000). Does Computer-Generated Speech Manifest Personality? An Experimental Test of Similarity-Attraction. *In proceedings of CHI 2000*, 329-336.

Nass, C., Moon, Y., & Carney, P. (1999). Are respondents polite to computers? Social desirability and direct responses to computers. *Journal of Applied Social Psychology*, 29(5), 1093-1110.

Nass, C., Moon, Y., & Green, N. (1997). Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of applied social psychology*, *27(10)*, 864-876.

Nass, C., Isbister, K., & Lee, E. J. (2000). Truth is beauty: researching embodied conversational agent.

Nass, C., Steuer, J., & Tauber, E. (1994). Computers are social actors. *Proceedings of the ACM Computer Human Interaction 1994 Conference*, 72-78.

Paiva, A., Dias, J., Sobral, D., Aylett, R., Sobreperez, P., Woods, S., & Zoll, C. (2004). Caring for Agents and Agents that Care: Building Empathic Relations with Synthetic Agents. Proceedings of the AAMAS'04 Conference, 194-201.

Park, S. & Catrambone, R. (2007). Social Facilitation Effects of Virtual Humans. *Human Factors,* 49 (6), 1054-1060.

Park, S. & Catrambone, R. (in press). Social Responses to Virtual Humans. *Proceedings of the 8th International Conference on Intelligent Virtual Agents*.

Petty, R. & Cacioppo, J. (1986). *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*. NY: Springer-Verlag.

Petty & Cacioppo. (1986). Persuasion and Communication.

Predinger, H. & Ishizuka, M. (2001). Social Role Awareness in Animated Agents. *Proceedings of the fifth international conference on Autonomous agents*, 270-277.

Rajecki, D. W., Ickes, W., Corcoran, C., & Lenerz, K. (1977). Social facilitation of human performance: Mere presence effects. *Journal of Social Psychology, 102,* 297–310.

Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review*, *86*, 61-79.

Reeves, B. & Nass, C. (1996). *The Media Equation*. Cambridge University Press, Cambridge.

Rickenberg, R. & Reeves, B. (2000). The effects of animated characters on anxiety, task performance, and evaluations of user interfaces. *Proceedings of the ACM Computer Human Interaction 2000 Conference*, 49-56.

Rousseau, D. & Hayes-Roth, B. (1997). A Social-Psychological Model for Synthetic Actors. Report No. KSL 97-07. Knowledge Systems Laboratory, Department of Computer Science, Stanford University.

Selvarajah, K. & Richards, D. (2005). The use of emotions to create believable agents in a virtual environment, *Proceedings of the fourth international joint conference on autonomous agents and multiagent systems*, 13-20.

Schneider, W. & Shiffrin, R. (1977). Controlled and automatic human information processing. *Psychological Review*, 84(1), 1-66.

Selvarajah, K. & Richards, D. (2005). The Use of Emotions to create Believable Agents in a Virtual Environment. *Proceedings of the AAMAS'05 Conference*, 13-20.

Shiffrin, R. & Schneider, W. (1977). Controlled and automatic human information processing. *Psychological Review*, 84(1), 1-66.

Singer, E., Frankel, M. R., & Glassman, M. B. (1983). The effect of Black and White interviewers on Black responses in 1968. *Public Opinion Quarterly*, 35, 44-68.

Sproull, L., Subramani, M., Kiesler, S., Walker, J., & Waters, K. (1996). When the Interface is a Face. *Human-Computer Interaction*, 11, 97-124.

Stern, S. E., Mullennix, J. W., Dyson, C., & Wilson, S. J. (1999). The persuasiveness of synthetic speech versus human speech. *Human Factors, 41*, 588-595.

Sudman, S. & Bradburn, N. M. (1974). *Response effects in surveys*. Chicago, IL: Aldine.

Sundar, S. S. (2004). Loyalty to computer terminals: is it anthropomorphism or consistency? *Behaviour & Information Technology*, 23(2), 107-118.

Thomas, F. & Johnston, O. (1981). Disney Animation: The Illusion of Life.

Tresselt, M. E., & Mayzner, M. S. (1966). Normative solution times for a sample of 134 solution words and 378 associated anagrams. *Psychonomic Monograph Supplements, 1,* 293–298.

Turning, A. M. (1950). Computing and machinery and intelligence. *Mind*, 59, 433-460.

Walker, J., Sproull, L., & Subramani, R. (1994). Using a human face in an interface. In *Proceedings of the Conference on Human Factors in Computers 1994* (pp. 85–91). New York: Association for Computing Machinery Press.

Yee, N., Bailenson, J. N., Rickertsen, K. (2007). A meta-analysis of the impact of the inclusion and realism of human-like faces on user experiences in interfaces. *Proceedings of the ACM Computer Human Interaction 2007 Conference*, 1-10.

Zanbaka, C., Goolkasian, P., & Hodges, L. (2006). Can a virtual cat persuade you?: The role of gender and realism in speaker persuasiveness. *Proceedings of the ACM Computer Human Interaction 2006 Conference*, 1153-1162.

Zanbaka, C., Ulinski, A., Goolkasian, P., & Hodges, L. F. (2004). Effects of virtual human presence on task performance. In *Proceedings of the 14th International Conference on Artificial Reality and Telexistence* (ICAT 2004; pp. 174–181). Tokyo: Virtual Reality Society of Japan.

Zanbaka, C., Ulinski, A., Goolkasian, P., & Hodges, L. (2007). Social Responses to Virtual

Humans: Implications for Future Interface Design. *Proceedings of the ACM Computer Human Interaction 2007 Conference*, 1561-1570.

Zarate, M. A. & Sandoval, P. (1995). The effects of contextual cues on making occupational and gender categorizations. *British Journal of Social Psychology, 34*, 353-362.

Zarate, M. A. & Smith, E. R. (1990). Person categorization and stereotyping. *Social Cognition, 8(2)*, 161-185.

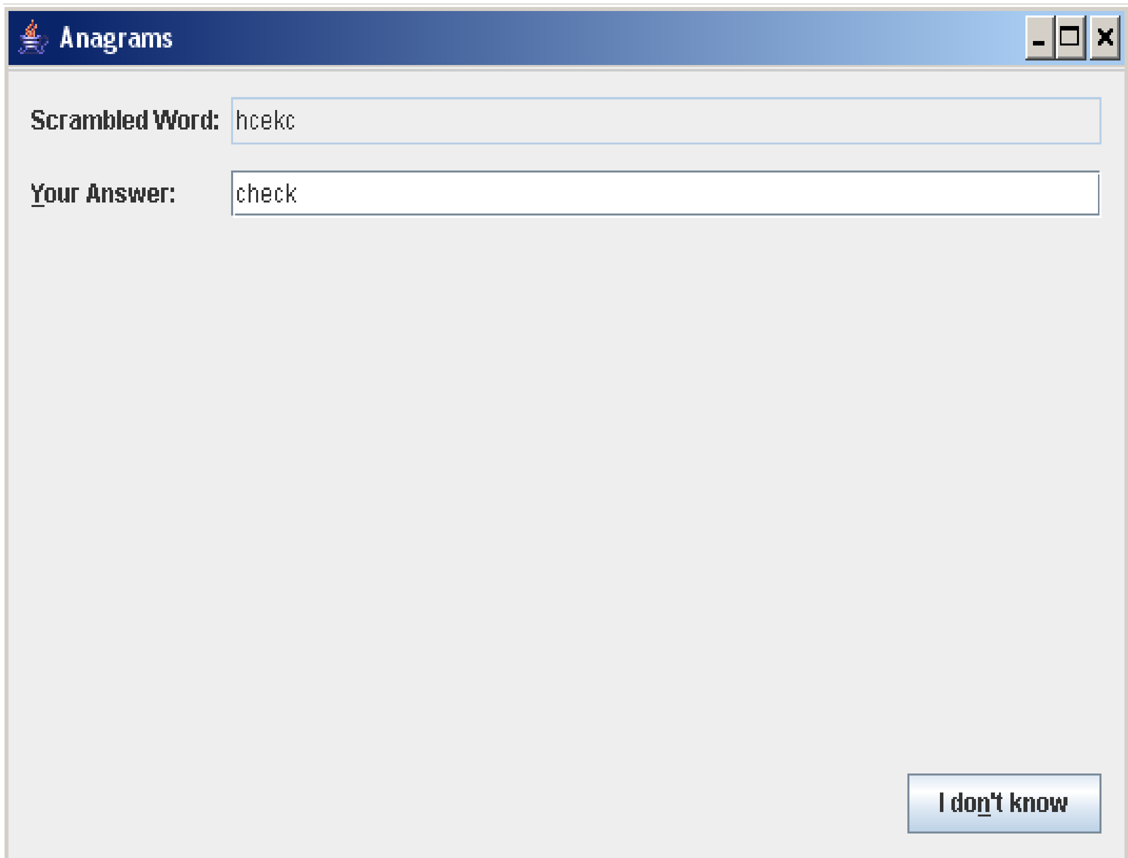
Zhao, W., Chellappa, R., Phillips, P., & Rosenfeld, A. (2003). Face recognition: a literature survey. *ACM Computing surveys*, 399-458.

Zorriassatine, F., Wykes, C., Parkin, R., & Gindy, N. (2003). A survey of virtual prototyping techniques for mechanical product development. In *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 217,* 513–530.

Zarate, M. A. & Sandoval, P. (1995). The effects of contextual cues on making occupational and gender categorizations. *British Journal of Social Psychology, 34*, 353-362.

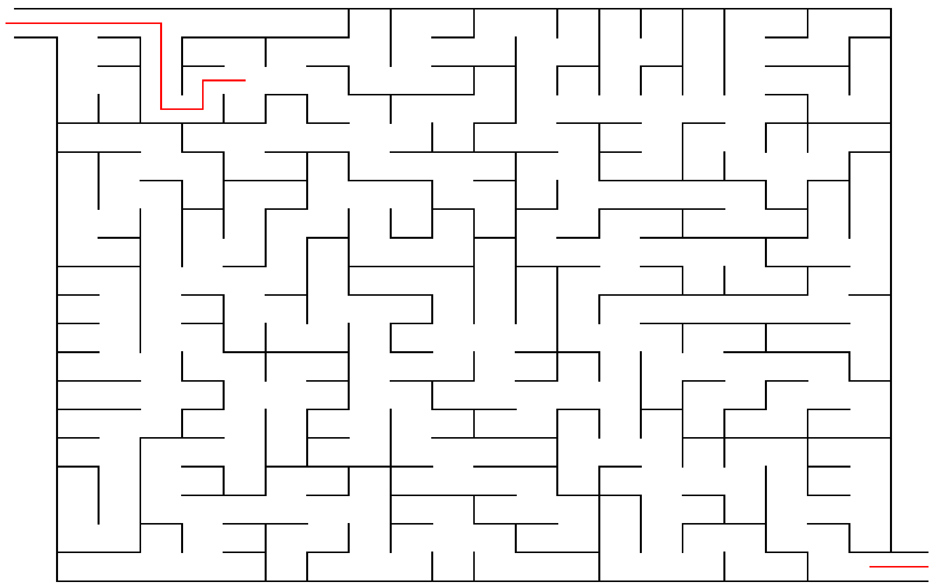
Zarate, M. A. & Smith, E. R. (1990). Person categorization and stereotyping. *Social Cognition, 8,* 161-185.

Appendix A: Anagram Task



Appendix B: Maze Task

START



EXIT

Appendix C: Modular Arithmetic Task

Appendix D: Introduction Script for Politeness Norm Experiment

Thank you for participating in this study. In this experiment you will be working with the virtual human (the experimenter). The virtual human (the experimenter) is a tutor and will provide you with facts on world geography. The tutor will present 20 out of a database of 1000 facts. After learning each fact, you will respond as either “you did know the fact” or “you did not know the fact” by selecting such options in the screen, using the mouse (by telling the experimenter). Based on your response, the tutor will adjust the topic so that you can learn something you are not familiar with.

After completing the tutoring session, you will take a test. Specifically, the tutor will ask 12 multiple choice questions out of a database 5000 questions. The tutor will then provide you with your score as well as evaluate its own performance as a tutor.

Appendix E: 20 Facts on World Geography

1. Asia is a larger continent than Africa.

2. The Amazon rainforest produces more than 20% of the world's oxygen supply.

3. The largest desert in Africa is the Sahara desert.

4. There are more than roughly 200 million people living in South America.

5. Australia is the only country that is also a continent.

6. The tallest waterfall is Angel Falls in Venezuela.

7. More than 25% of the world's forests are in Siberia.

8. The largest country in South America is Brazil. Brazil is almost half the size of South America.

9. Europe produces just over 18 percent of all the oil in the world.

10. Asia has the largest population with over 3 billion people.

11. Europe is the second smallest continent with roughly 4 million square miles.

12. The four largest nations are Russia, China, USA, and Canada.

13. The largest country in Asia by population is China with more than 1 billion people.

14. The Atlantic Ocean is saltier than the Pacific Ocean.

15. The longest river in North America is roughly 3,000 feet and it is the Mississippi River in the United States.

16. Poland is located in central Europe and borders with Germany, Czech Republic, Russia, Belarus, Ukraine and Lithuania.

17. Australia has more than 28 times the land area of New Zealand, but its coastline is not even twice as long.

18. The coldest place in the Earth's lower atmosphere is usually not over the North or South Poles.

19. The largest country in Asia by area is Russia.

20. Brazil is so large that it shares a border with all South American countries except Chile and Ecuador.

Appendix F: 12 Questions on World Geography

1. The nation that has the largest forests is ____.

1) Brazil 2) Russia 3) Germany 4) USA

2. Which country **does not** border with Poland?

1) Germany 2) Ukraine 3) Slovakia 4) France

3. China takes almost ____% of Asia in terms of population.

1) 15% 2) 33% 3) 42% 4) 52%

4. What is the third largest nation?

1) China 2) USA 3) Canada 4) Russia

5. The coldest place in the Earth's lower atmosphere is ____.

1) North Poles 2) South Poles 3) Equator 4) Siberia

6. Which continent is the largest in the world?

1) Asia 2) Africa 3) Europe 4) Australia

7. Brazil **does** share a border with ____?

1) Chile 2) Colombia 3) Mexico 4) Ecuador

8. The smallest continent is roughly ____million square miles.

1) 3 2) 4 3) 8 4) 25

9. The tallest waterfall is in ____.

1) Europe 2) Asia 3) South America 4) North America

10. Which continent is the smallest in the world?

1) Europe 2) Australia 3) Africa 4) Asia

11. Which country **does not** have a coastline?

1) New Zealand 2) Australia 3) Switzerland 4) Austria

12. Brazil is roughly 3,300,000 square miles. Given this, South America is roughly _____ square miles.

1) 1,000,000 2) 4,000,000 3) 6,000,000 4) 9,000,000

Appendix G: The Script used in the Scoring Session

Thank you for answering all 12 questions. We will now conclude the testing session and begin the scoring session. (pause 3 seconds). You have answered 10 out of 12 questions. Your score is 84 out of 100.

The tutor provided very useful facts for answering the questions. The tutor performed well (pause 3 seconds).

We will now conclude the scoring session and begin the interview session.

Appendix H: Interview Questionnaire

INSTRUCTIONS: You were tutored by the virtual human (the experimenter) in this experiment. Please think about what you had done with the virtual human (the experimenter) when you answer the items on this questionnaire. (pause 3 seconds if recorded with voice)

For each of the following adjectives, please indicate how well it describes your overall experience of the tutoring, the testing, and the scoring session with the virtual human (the experimenter) by providing the appropriate response.

Adjectives	(describes very poorly)									(describes very well)
Analytical	1	2	3	4	5	6	7	8	9	10
Competent	1	2	3	4	5	6	7	8	9	10
Enjoyable	1	2	3	4	5	6	7	8	9	10
Fair	1	2	3	4	5	6	7	8	9	10
Friendly	1	2	3	4	5	6	7	8	9	10
Fun	1	2	3	4	5	6	7	8	9	10
Helpful	1	2	3	4	5	6	7	8	9	10
Informative	1	2	3	4	5	6	7	8	9	10
Knowledgeable	1	2	3	4	5	6	7	8	9	10
Likable	1	2	3	4	5	6	7	8	9	10
Polite	1	2	3	4	5	6	7	8	9	10
Useful	1	2	3	4	5	6	7	8	9	10
Warm	1	2	3	4	5	6	7	8	9	10