Using Social Agents to Explore Theories of Rapport and Emotional Resonance

Chapter · November 2013

DOI: 10.1093/acprof:oso/9780195387643.003.0012

Jonathan Gratch
University of Southern California

Ning Wang
University of Southern California

Social Emotions in Nature and Artifact

Edited by Jonathan Gratch and Stacy Marsella

OXFORD
UNIVERSITY PRESS

Contents

11

Using Social Agents to Explore Theories of Rapport and Emotional Resonance

Jonathan Gratch, Sin-Hwa Kang, & Ning Wang

Introduction

Emotions are often described as momentary, even discrete, reactions to some specific event. For example, we see a bear in the forest, and fear seems to emerge without bidding, fully formed. Writing around the same time as William James, George Herbert Mead employed a very different metaphor to emphasize how emotions can incrementally unfold and resonate between social actors. He asks us to imagine two angry dogs preparing to fight (Mead, 1934, pp. 42–43):

> The act of each dog becomes the stimulus of the other dog for his response. There is then a relationship between these two; and as the act is responded to by the other dog, it, in turn, undergoes change. The very fact that the dog is ready to attack another becomes a stimulus to the other dog to change his own position of his own attitude. He has no sooner done this than the change of attitude in the second dog in turn causes the first dog to change his attitude.… [More generally, referring to human combatants,] If the individual is successful a great deal of his attack and defense must be not considered, it must take place immediately. He must adjust himself "instinctively" to the attitude of the other individual. He may, of course, think it out. He may deliberately feint in order to open up a place of attack. But a great deal has to be without deliberation.

Of course, such "instinctive" unfolding of emotions is not limited to battles. Imagine instead two close friends engrossed in conversation. The speakers seem tightly enmeshed in something like a dance. They rapidly resonate to each other's gestures, facial expressions, gaze, and patterns of speech. Tickle-Degnen and Rosenthal (1990) referred to this seemingly automatic attunement of positive emotional displays, gaze, and gestures as *rapport* and documented a wide range of beneficial interpersonal consequences that result from its establishment.

As these two examples suggest, such unfolding and resonating patterns of emotional display can arise without conscious deliberation (Bargh, Chen, & Burrows, 1996; Lakin, Jefferis, Cheng, & Chartrand, 2003), but they are hardly fixed responses to external stimuli. The same individual that smiles reflexively to the smile of a friend may frown to the smiles of his opponent (Lanzetta & Englis, 1989). The way such interpersonal patterns unfold depends on a host of intra- and interpersonal factors including the relative power of individuals in an interaction—with weaker partners tending to mimic the more powerful but not vice versa (Tiedens & Fragale, 2003); prior expectations—with unexpectedly positive behaviors producing more favorable outcomes than expected ones (Burgoon, 1983); and conformity to social norms such as reciprocity (Roloff & Campion, 1987) or the appropriateness of expressing negative emotions (Adam, Shirako, & Maddux, 2010). More generally, as discussed by Parkinson (Chapter 2, this volume), these patterns both arise from and help to redefine an evolving affective relationship between individuals.

Building computer programs that can engage people in this unfolding emotional dance is a fascinating prospect with potentially profound practical and scientific consequences. In humans, such "emotional resonance" has been argued to play a crucial role in the establishment of social bonds (Tickle-Degnen & Rosenthal, 1990), in promoting or diffusing conflict (Lanzetta & Englis, 1989), in determining the persuasiveness of speech (Bailenson & Yee, 2005; Fuchs, 1987), and in the establishment of identity Mead (1934). Computer agents that engage people in this manner, on the one hand, could enhance our understanding of this fundamental social process by systematically manipulating nonverbal patterns in ways that are difficult or impossible for human confederates in social science experiments, allowing the possibility of studies that definitively establish whether such patterns *cause* these effects or merely reflect them (see Bailenson, Beall, Loomis, Blascovich, & Turk, 2004; Bente, Kraemer, Petersen, & de Ruiter, 2001; Forgas, 2007). Indeed, in this chapter we show that synthesized patterns of nonverbal behavior do seem to cause predicted changes in the impressions and behaviors of human subjects in laboratory studies, and give some insight into the factors that promote these effects. On the other hand, assuming theses effects can be demonstrated outside the laboratory, such emotionally resonant agents could have dramatic implications for computer-based applications to health, commerce, and entertainment.

We argue that several challenges must be overcome before realizing this vision: the computer must approach a social interaction with certain goals and expectations about the social context, it must recognize and interpret the verbal and nonverbal cues, in real-time, of its human interaction partner, it must rapidly generate responses contingent on its current goals and expectations, and it must evolve its goals and expectations of self and other, based on the unfolding interaction. More importantly, success depends not on simply overcoming these challenges, but demonstrating that establishing such interactivity has measurable and desirable consequences for human–computer interaction (see also, Gratch, 2008, 2010).

In this chapter, we describe one line of research in our laboratory that builds toward this vision of emotionally resonant agents. We describe the Rapport Agent, an interactive agent and methodological tool designed to investigate the role of nonverbal patterning in human–computer and computer-mediated interaction. The Rapport Agent recognizes and reflexively responds to certain human nonverbal cues with the aim of establishing a sense of rapport between the human and computer. In this sense, the system is solving a much simpler problem than the one outlined previously: we explore a fixed mapping between nonverbal perception and action, and ignore how these actions and reactions alter the agent's expectations and perception of the social interaction. Nevertheless, as we will outline, even this simple system yields some important insights into the social processes that govern human–agent interaction and is a stepping stone to more comprehensive models of human social behavior.

As the focus of the Rapport Agent is to establish a subjective sense of rapport in a human interaction partner, we first outline the concept of rapport. We next describe the basic mechanics of the system. We then outline a series of laboratory studies and resulting findings that give insight into how nonverbal patterns of behavior can influence both subjective perceptions (such as feelings of rapport or embarrassment) and behavioral outcomes (such as speech fluency or intimate self-disclosure). Finally we end with some speculation on the future of such human–computer systems.

Interpersonal Rapport Between People and Machine

Rapport is studied across a range of scientific disciplines for its role in fostering emotional bonds and prosocial behavior. In their seminal article, Tickle-

Degnen and Rosenthal (1990) equate rapport with behaviors indicating positive emotions (e.g., head nods or smiles), mutual attentiveness (e.g., mutual gaze), and coordination (e.g., postural mimicry or synchronized movements).[1] The role of rapport in fostering effective social interaction is well established in the field of social psychological research. Rapport is easy to assess—observers can reliably predict whether people will report feelings of rapport by watching their nonverbal behavior (Ambady, Bernieri, & Richeson, 2000)—and these judgments and feelings predict success in a surprising range of social activities (Bernieri, Gillis, Davis, & Grahe, 1996). For example, rapport is argued to underlie social engagement (Tatar, 1997), success in teacher–student interactions (Bernieri, 1988), success in negotiations (Drolet & Morris, 2000), improving worker compliance (Cogger, 1982), psychotherapeutic effectiveness (Tsui & Schultz, 1985), improved test performance in classrooms (Fuchs, 1987), and improved quality of child care (Burns, 1984).

The power of rapport in face-to-face interaction has served as an inspiration to researchers in human–computer interaction, and a number of projects have sought to endow computers with the ability to sense and respond to human nonverbal behavior. For example, Neuro Baby is a computer-generated baby that analyzes speech intonation and uses the extracted features to trigger emotional displays (Tosa, 1993). Similarly, Breazeal's Kismet robot extracts emotional qualities in the user's speech (Breazeal & Aryananda, 2002). Whenever the speech recognizer detects a pause in the speech, the previous utterance is classified (within 1 or 2 seconds) as indicating approval, an attentional bid, or a prohibition, soothing or neutral. This recognition feature is combined with Kismet's current emotional state to determine facial expression and head posture. Only a few systems can interject meaningful nonverbal feedback *during* another's speech, and these methods usually rely on simple acoustic cues. For example, REA will execute a head nod or paraverbal (e.g., "mm-hum") if the user pauses in midutterance (Cassell et al., 1999). Some work has attempted to extract extralinguistic features of a speakers' behavior, but not for the purpose of informing listening behaviors. For example, Brand's voice puppetry work attempts to learn a mapping between acoustic features and facial configurations, inciting a virtual puppet to react to the speaker's voice (Brand, 1999). Although there is considerable research showing the benefit of such feedback on human-to-human interaction, there has been almost no research on their impact on human-to-virtual human rapport (cf. Bailenson & Yee, 2005; Cassell & Thórisson, 1999; Schröder et al., 2008).

The success of such technological approaches remains an empirical question, but a more fundamental challenge lies in finding appropriate theoretical constructs to inform their design. Research by Nass and colleagues on the "Media Equation" (Reeves & Nass, 1996)—which illustrates that people will often respond to computer agents as if they are people—suggests that human–computer interactions are governed by many of the same social processes and norms that apply in human face-to-face interactions and, by inference, that social theories could be applied directly to the design of computational systems. However, agents are not people. They are far more limited in their intellectual capabilities. They can look and behave in ways that are decidedly unreal. And (currently) you can be certain you won't bump into them in the street or worry about them besmirching your reputation to your friends. Thus, a central question in designing emotionally resonant agents is whether social theories such as Rapport can be applied directly to their design, or whether we must develop unique theories to explain this new and increasingly common form of social interaction.

Rapport Agent

The Rapport Agent (see Figure 11.1) is designed to evoke subjective feelings and behavioral consequences of rapport with human participants in a restricted set of social interactions we refer to as *quasi-monologs*. In a quasi-monolog, the human does most of the talking and the agent primarily prompts human speech and provides attentive listening feedback. The Rapport Agent was originally developed to replicate findings by Duncan and colleagues on listening behaviors (Welji & Duncan, 2004). In their studies, a speaker (the narrator) retells some previously observed series of events (e.g., the events in a recently-watched video) to some listener. In our case, the listener is a computer-generated character that has been programmed to produce (or fail to produce) the type of nonverbal feedback we seen in rapportful conversations. More recently, we have explored interview settings where

FIGURE 11.1 A speaker interacting with the Rapport Agent.

the agent asks a person a series of questions and listens attentively until it is time to ask the next question (Kang, Gratch, Wang, & Watt, 2009). By focusing on quasi monologs, rather than richer two-way conversational forms, we sidestep the current limitations in automatic natural language processing technology: the system produces what seems to be meaningful feedback without having to deeply process the meaning of the monolog. By changing these behaviors and the cover story surrounding this interaction (e.g., is the listener controlled by a computer or the behavior of an actual listener?), we hope to elicit the same correlations seen in face-to-face interaction between nonverbal behaviors, self, and other reports of rapport, and social outcomes such as liking, persuasion, and conversational fluency, as well as to unpack the factors that govern these effects.

The central challenge for the Rapport Agent is to generate the nonverbal listening feedback associated with rapportful interactions. Although our recent work has focused on how to learn to produce these behaviors automatically from examples (Huang, Morency, & Gratch, 2010; Morency, de Kok, & Gratch, 2008), the system we describe here was crafted by hand from findings in the psycholinguistic literature showing that feelings of rapport are correlated with simple contingent behaviors between speaker and listener. Such feedback includes the use of backchannel continuers (nods, elicited by speaker prosodic cues, that signify the communication is working), postural mirroring, and mimicry of certain head gestures, for example, gaze shifts and head nods (Chartrand & Bargh, 1999; Ward & Tsukahara, 2000; Yngve, 1970). The Rapport Agent generates such feedback by real-time analysis of acoustic properties of speech (detecting backchannel opportunity points, disfluencies, questions, and loudness) and speaker gestures (detecting head nods, shakes, gaze shifts, and posture shifts).

Behavior of Human Listeners

The psycholinguistic literature has identified a variety of behaviors that human listeners perform when in face-to-face conversations. Even though many listener behaviors provide feedback about the semantic content the speaker's speech, a large class of behaviors appear unrelated to specific meaning. Rather, these behaviors seem to trigger off of nonsemantic features of the speaker's presentation, may precede complete understanding of the speech content, and are often generated without the listener or speaker's conscious awareness. Nonetheless, such behaviors can significantly influence the flow of a conversation and the impressions and feelings of the participants. Here we review some of these behaviors, the circumstances that trigger their production, and their hypothesized influence on the interaction.

Backchannel Continuers

Listeners frequently nod and utter paraverbals such as "uh-huh" and "mm-hmm" as someone is speaking. Within the psycholinguistic literature, such behaviors are referred to as backchannel continuers and are considered as a signal to the speaker that the communication is working and that they should continue speaking (Yngve, 1970). Several researchers have developed models to predict when such feedback occurs. Cathcart et al. (2003) propose a model based on pause duration and trigram part-of-speech frequency. According to the model of Ward and Tsukahara (2000), backchannel continuers are associated with a lowering of pitch over some interval. Cassell (2000) argues that head nods could result from the raised voice of the speaker. The approaches of Ward and Cassell are more amenable to a real-time treatment as they are based purely on simple properties of the audio signal.

Disfluency

Spoken language often contains repetition, spurious words, pauses, and filled pauses (e.g., ehm, um, un). Such disfluency is viewed as a signal to the listener that the speaker is experiencing processing problems or experiencing high cognitive load (Clark & Wasow, 1998) and frequently elicit "take your time" feedback from the listener (Ward & Tsukahara, 2000). According to our own video analysis, rather than nod or utter sounds as in backchannel continuers, listeners tended to perform posture shifts, gaze shifts, or frowns in response to disfluency. The presumed meaning of such a posture shift is that the listener is telling the speaker to take his time (Cassell, 2000).

Mimicry

Listeners often mimic behavior of a speaker during a conversation. Although they are not necessarily aware of doing it, people in a conversation will adjust the rhythm of speech, their body posture, and even their breathing to each other (Lakin et al., 2003; McFarland, 2001; Warner, 1996). Mimicry, when not exaggerated to the point of mocking, has a variety of positive influences on the interactants. Speakers who are mimicked are more helpful and generous toward the listener (Van baaren, Holland, Kawakami, & Knippenberg, 2004). Mimicry can result in the perception of a pleasant, natural conversation (Warner, Malloy, Schneider, Kooth, & Wilder, 1987). It may also be important in synchronizing conversational flow, for example, by providing expectations on when a speaker can be interrupted. Given such influences, many of the agent's listening behaviors should mimic aspects of the speaker's behavior.

One salient speaker behavior is shifts in posture. When a speaker shifts her posture, for example, by changing her weight distribution from one leg to another, or by folding her arms, this is often mirrored by the listener. Such posture shifts, both for speakers and listeners, tend to occur at discourse segment boundaries and may function to help manage such transitions (Cassell, Nakano, Bickmore, Sidner, & Rich, 2001). When present, such mimicry has been shown to positively influence the emotional state of the speaker (Van baaren et al., 2004). Gaze is also an important aspect of a speaker's behavior. Speakers will often gaze away from the listener; for example, when mentioning a concrete object within his vicinity, he will often look at it. Listeners will frequently mimic the head gestures of a speaker also. If a speaker shakes or nods his head, listeners may repeat this gesture. Although this may simply reflect an understanding and agreement with the speaker's utterance, many of us have probably been in conversations where such gestures were produced without any real understanding. In any event, an agent can easily mimic such gestures without explicit understanding.

Other External Influences

Many factors may influence the occurrence of gestures during a conversation. For example, listeners frequently mimic the facial expression of speakers, and this apparently plays an important role in the perception of empathy (Sonnby-Borgstrom, Jonsson, & Svensson, 2003). Individuals also differ in their response to the same speech based on a variety of dispositional and situational factors. There are people who almost do not gesture at all, and there are people who gesture like it is a workout. Often, this is related to the speaker's emotions during the conversation. For example, people tend to gesture more when excited and less when sad. Also, the relation of the two people is of importance. People tend to gesture remarkably more when they talk to a friend than when they are talking to a complete stranger (Welji & Duncan, 2004).

Behavior of the Rapport Agent

From this literature we extract a small number of simple rules that the Rapport Agent could possibly utilize to drive its behavior. The mapping presented here (summarized in Table 11.1) is not a complete coverage of all gestures that at all times are accompanied by the certain speech features, but could be sufficient to increase the perceived authenticity of the conversation.

TABLE 11.1 Rapport Agent Behavior Mapping

Human speaker behavior	*Rapport Agent response*
Lowering of pitch	Head nod
Raised loudness	Head nod
Speech disfluency	Posture/gaze shift
Shift posture	Mimic
Gaze away	Mimic
Head nod/head shake	Mimic

Real-Time Classification of Speaker Behavior

To produce listening behaviors, the Rapport Agent first collects and analyzes the speaker's upper-body movements and voice to detect the features listed in Table 11.2.

We implement the behavior rules listed in the previous section by detecting the various implicated aspects of the speaker's behavior. As such listening behaviors occur within utterances, this imposes strong real-time requirements on what features can be reasonably extracted given the limits of current technology. Here we describe the implementation of feature detectors that support the behavioral rules.

Participant physical movements are recognized and tracked using computer vision techniques. The current version of the Rapport Agent is limited to tracking the motion of the speaker's head (see Huang, Morency, & Gratch, 2011, for more recent developments). Specifically, we use Watson, developed by Louis-Philippe Morency, an image-based tracking library that uses stereo images to track the participants' head position and orientation (Morency, Sidner, Lee, & Darrell, 2005). Watson also incorporates learned motion classifiers that detect head nods and shakes from a vector of head velocities. Other features are derived from the position and orientation of participant's head (filtered to reduce the impact of noise). For example, from the head position, given the participant is seated in a fixed chair, we can infer the posture of the spine.

From the speech signal, we extract a variety of "paralinguistic" features (i.e., features of the voice signal other than the meaning of the words). This includes such features as vocal pitch and then intensity of the speech signal (the Rapport Agent ignores the semantic content of the speaker's speech), using a signal processing package, Laun, developed by Mathieu Morales (Lamothe & Morales, 2006). Features include silent/normal/loud speech (derived from signal intensity) and "backchannel opportunity points" (pitch contours that are associated with listener feedback), derived using the approach of Ward and Tsukahara (2000). In addition, we approximately segment the audio signal into separate utterances based on silences and attempt to detect some features that hold across the utterance including pitch range (positive affect is often associated with wider pitch range).

Behavior Mapping

Recognized speaker features are mapped into listening behaviors through a set of authorable mapping rules. The language is based on five primitives:

- Each participant in the interaction is described by an *agent*. Agents consist of a set of actions, states, animations, and reactions. For the discussion that follows, we will assume two agents: the agent that represents the human speaker and the agent that represent the Rapport Agent listener.
- *Actions* represent discrete behavioral events that can be generated by an agent. These can consist of the detectable features of human behavior (Table 11.2) or arbitrary behavior outputs of the Rapport Agent.
- *States* describe characteristics of an agent that can persist over time. Typically, states are asserted as consequences of actions (e.g., after detecting LeanLeft, the speaker is in the state of LeaningLeft). States can be constrained logically (e.g., the speaker cannot be simultaneously speaking and silent) and temporally (e.g., to ensure an agent stays in some state for some period of time).
- *Animations* are physical behaviors described in the Behavior Markup Language (BML) (Kopp

TABLE 11.2 Rapport Agent Detected Speaker Features

Motion features		*Vocal features*	
Gestures	Nod, shake	Intensity	Silent, normal, loud
Head roll	Upright, lean left, lean right	Range	Wide, narrow
Gaze	Straight, up, down, left, right	Others	Backchannel opportunity

et al., 2006) that can be associated with agent actions. For example, a backchannel continuer might be associated with a nod animation.
- *Reactions* map from an action in one agent to an action in another agent. The mapping is conditional on the current state of one or more agents and can map, probabilistically, to one of a set of other actions.

Typically, reactions map actions of the speaker to actions by the Rapport Agent. For example, if Laun detects a backchannel opportunity point in the speaker, this could cause the Rapport Agent to react with a Nod with probability 0.6 or GazeUp with probability 0.2, assuming the Rapport Agent is in the state of GazingForward. The framework, however, can support more general settings. For example, one could define mapping rules for multiparty settings (e.g., multiple speakers or multiple listening agents). Alternatively, one could transform the behavior of a human listener into some, perhaps altered, animated behavior (Bailenson et al., 2004).

Animation

Rapport Agent animation commands are passed to the SmartBody animation system (Kallmann & Marsella, 2005; Thiebaux, Marshall, Marsella, & Kallmann, 2008). This is a virtual human animation system designed to seamlessly blend animations and procedural behaviors. These animations are rendered in the Unreal Tournament™ game engine and displayed to the Speaker.

System Architecture

The Rapport Agent has an open modular architecture that facilitates the incorporation of different feature detectors and animation systems, and has an easily authored mapping between features and behavior. The behavior mapping language incorporates contextual features, probabilistic responses, and some control over the temporal dynamics of behavior. We used a vision-based tracker and set the setting to a seated interaction with a life-sized image of a character's head displayed on a computer monitor. Figure 11.2 illustrates the basic outlines of the Rapport Agent architecture.

As we are simultaneously recognizing features from multiple channels (head, body and speech), and listening behaviors have some duration, it is possible that a listening behavior will be proposed that overlaps with a currently executing behavior. This could easily result in unnatural behavior.

We use a synchronization module to solve this problem. This module parses messages on the communication bus and determines if the message is intended for the agent and which type of gesture is contained in the command. Once this parsing has been done, the function accompanying that type of gesture can be called. This function determines whether a previous gesture is still performing, and when this is not the case, a message is created, which is to be sent to the agent. The module also incorporates a flexible behavior mapping allowing designers to easily experiment with alternative mappings between classes of listening behaviors and their physical manifestation.

Empirical Findings

The Rapport Agent exhibits some of the nonverbal resonances that occurs in natural conversation, but are people influenced by computer-generated behaviors? In a series of empirical studies (which we will discuss below), we have used the Rapport Agent as methodological tool to explore such questions about the role of nonverbal behaviors in human and human–computer

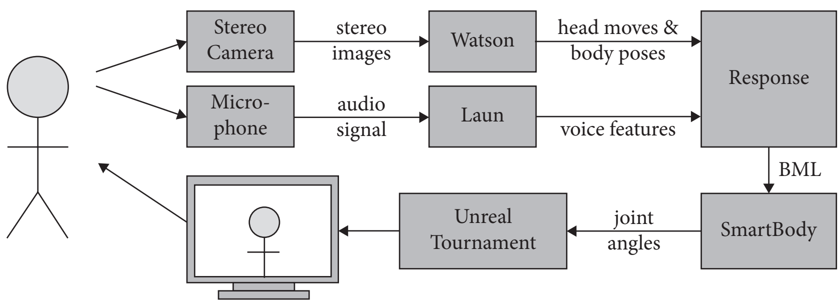

FIGURE 11.2 Rapport Agent architecture.

interaction. For example, we can demonstrate that such synthetic resonant behaviors alter participant feelings, impressions, and behavior, and that the strength of such effects are mediated by the same factors of positivity, contingency, and mutual attention that Tickle-Degnen and Rosenthal have posited for face-to-face interactions between people. Further, as posed by Nass and Reeves in their *media equation* (Reeves & Nass, 1996), such effects seem to occur even when participants know that such behaviors are "merely" produced by a computer.

In the remainder of this chapter, we outline the basic experimental paradigm by which we explore such questions and then summarize the findings of several studies. We highlight different aspects of these findings. First, we describe the consequences, subjective and behavioral, that such contingent nonverbal behaviors have on human participants. Next, we discuss some of the factors that seem to mediate these effects. These include properties of the agent's behavior, dispositional factors on the part of participants, and situational factors surrounding the interaction such as if participants believe they are interaction with another participant or a computer program.

Experimental Paradigm

As discussed above, our experiments largely follow the quasi-monolog paradigm of Welji and Duncan (Welji & Duncan, 2004). As illustrated in Figure 11.3, human participants sit in front of the Rapport Agent and are prompted (either by the experimenter or by the agent itself) to retell some previously experienced situation. This situation might be a controlled experimental stimulus: in one series of experiments, participants watch a short cartoon or a video on sexual harassment and then are instructed to retell it to the agent in as much detail as possible. Alternatively, the situation might draw on real-life experience: in another series of studies, the Rapport Agent, by playing a prerecorded animation sequence, "asks" the participant a sequence of increasingly intimate questions. In either situation, the agent displays some form of nonverbal feedback while the participant speaks, with the exact nature of the feedback being dictated by the specific experimental manipulation.

In this paradigm, participant rapport is assessed through a variety of subjective and behavioral measures. Subjective measures include scales assessing rapport, social presence (Biocca & Harms, 2002), helpfulness, distraction, and naturalness. Behavioral measures include the length people speak (as a measure of engagement), the fluency of their speech (e.g., how many repeated words, broken words, and filled pauses per minute), the intimacy of their disclosure, the facial expressions produced, and amount of mutual gaze. Behavioral measures are assessed through a mixture of automatic annotation techniques and hand annotations by multiple coders.

Part of the power of this paradigm is that we can systematically manipulate aspects of the Rapport Agent's appearance and nonverbal behavior. For example, Figure 11.3 (from Gratch, Wang, Gerten, & Fast, 2007) illustrates a study that examined the impact of appearance (human vs. computer-generated human), behavior (human-generated vs. computer-generated), and contingency (carefully timed vs. random listener feedback). Going clockwise in this figure from the upper left, this experiment compared face-to-face interaction (in which a visible human participant displayed natural listening behaviors), the Rapport Agent (which exhibited computer-generated behavior and appearance), a noncontingent version of the Rapport Agent (which exhibited behaviors identical to the Rapport Agent in terms of their frequency and dynamics, but not contingent on the behaviors of the speaker), and a "mediated agent" (in which a real participant's listening behaviors were displayed on a computer-generated avatar).

Social Effects of the Rapport Agent

The picture emerging from a series of studies is that the Rapport Agent elicits beneficial social effects from human participants similar to what can be found in rapportful face-to-face interactions. Further, these benefits seem to depend on the nonverbal resonances produced by the Rapport Agent, and speaker impressions and behavior degrade when these behaviors are absent.

The presence of contingent nonverbal feedback changes participants' feelings about themselves, the agent, and the quality of their interaction. Subjective effects include:

- Greater feelings of self-efficacy (Kang, Gratch, Wang, & Watt, 2008b)
- Less tension (Wang & Gratch, 2010) and less embarrassment (Kang, Gratch, Wang, & Watt, 2008b)
- Greater feelings of rapport (Wang & Gratch, 2010)
- A greater sense of mutual awareness (von der Pütten, Krämer, & Gratch, 2009)

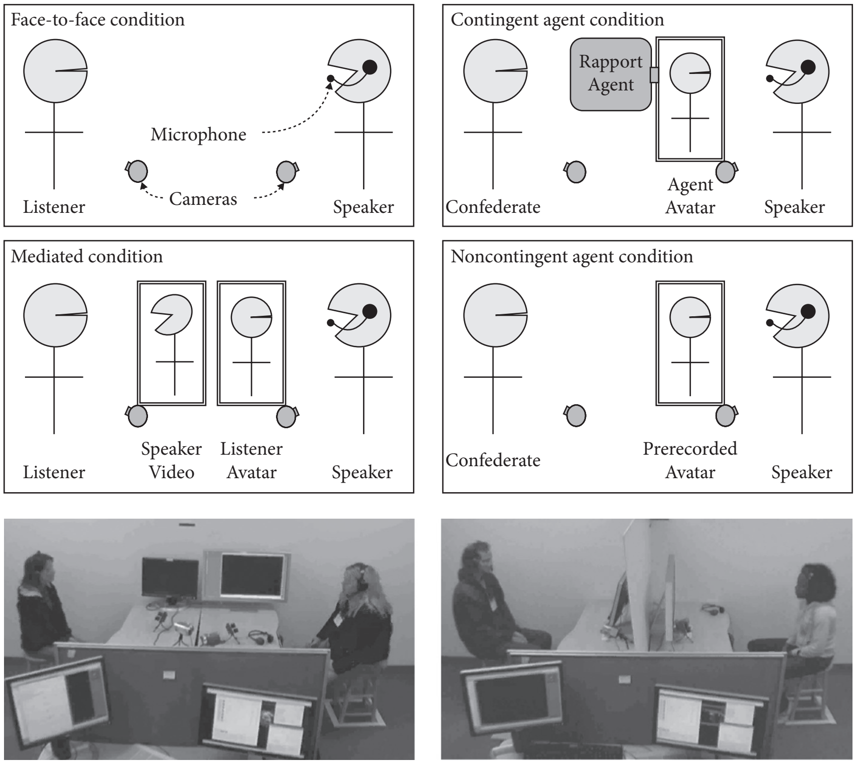

FIGURE 11.3 Graphical depiction of the four conditions. The actual face-to-face condition is illustrated on the lower left, and the setup for the other three conditions on the lower right.

- Greater feelings of trustworthiness on the part of the agent (Kang, Gratch, Wang, & Watt, 2008b)

The contingent nonverbal feedback of the rapport agent also changes participants' behavior. Behavioral effects include:

- More disclosure of information, including longer interaction times and more words elicited (Gratch et al., 2006, 2007; von der Pütten et al., 2009; Wang & Gratch, 2010)
- More fluent speech (Gratch et al., 2006, 2007; von der Pütten et al., 2009; Wang & Gratch, 2010)
- More mutual gaze (Wang & Gratch, 2010)
- Fewer negative facial expressions (Wang & Gratch, 2009b)

Factors That Influence These Effects

One focus of our empirical research has been directed at illuminating factors that mediate or moderate the observed social effects of the Rapport Agent, and more generally, explore the validity of alternative theoretical constructs for interpreting these results and guiding future agent design. We organize this discussion into three basic questions. First, what properties of the Rapport Agent are necessary or sufficient for promoting rapport? Second, what characteristics of people lead them to be more or less influenced by the agent's behavior? Finally, we consider the more general question of the usefulness of social psychological theory (which was developed to explain human-to-human interaction) as a framework for guiding the design of computer systems. Blascovich (Chapter 3, this volume) suggests that interactions might unfold very differently depending on whether people believe they are interacting *with computers* or *through computers* (i.e., the Rapport Agent might have different social effects depending on if participants believed its behavior was generated by a computer or if they believed the behavior corresponded to the movements of an actual human listener). This last question, depending on the answer, could have profound effects for the value of interdisciplinary research on social artifacts.

Characteristics of the Agent That Impact Rapport

If we adopt the former perspective and apply social psychological theory directly, Tickle-Degnen and Rosenthal's theory argues that three broad characteristic of agent behavior should promote the establishment of rapport between participants and the Rapport Agent. These factors include *positivity*, meaning that rapport will be enhanced by positive nonverbal cues, including positive facial expressions and encouraging feedback such as head nods; *coordination*, meaning that rapport will increase as the behavior of one participant is perceived as contingent upon (i.e., causally related to) the behavior of the other; and *mutual attentiveness*, meaning that rapport will increase as participants attend to each other nonverbally, for example, through mutual gaze. Thus, our empirical studies have sought to manipulate these factors independently and observe their impact on rapport.

Positivity

In face-to-face conversations, positivity is conveyed through a variety of nonverbal signals such as facial expressions and head nods. To explore the impact of positivity, we have operationalized this dimension in terms of the presence or absence of head nods generated by the Rapport Agent. Although we have yet to manipulate facial expressions (due to technical challenges in producing these at natural points in conversation), we have examined the relationship between participants' expressions and their perceptions of rapport.

Our findings illustrate that the presence of listener nods significantly enhances feelings and behavioral manifestations of rapport (Wang & Gratch, 2010). The strength of this effect seems moderated by the perceived contingency of the nods (Kang, Gratch, Wang, & Watt, 2008b) and dispositional factors (Pütten, Krämer, Gratch, & Kang, 2010) as will be discussed. We also explored whether participant facial expressions could be indicators of rapport, specifically via communicating positivity (Wang & Gratch, 2009b). We looked at participants' facial expressions, which were analyzed using the Facial Action Coding System and the Computer Expression Recognition Toolbox (see Whitehill, Bartlett, and Movellan, Chapter 6, this volume). In both of human-to-human and human-to-virtual human (Rapport Agent) interactions, more positive facial expressions were associated with greater rapport. Collectively, these findings are in line with Tickle-Degnen and Rosenthal's predictions.

Coordination

Coordination occurs when two participants' behaviors are mutually related in a timely and positive manner. Within the Rapport Agent, we have operationalized this factor by manipulating whether behaviors (such as nods or posture shifts) were generated in response to the participant's behavior (a coordinated condition) or in response to some unrelated factor (an uncoordinated condition). For example, in one study, we would create noncontingent behavior by showing a participant the same nonverbal behavior that was generated for the previous participant. This manipulation only breaks contingency while ensuring a similar frequency and distribution of behaviors are generated.

Overall, participants exhibit more subjective and behavior characteristics of rapport when coordination is present. For example, breaking the contingency of nonverbal feedback leads participants to talk less and produced more disfluent speech (Gratch et al., 2007). These effects were especially strong in participants who scored high in a scale of social anxiety as such participants, in addition to these behavioral effects, feel less subjective rapport and greater embarrassment (Kang, Gratch, Wang, & Watt, 2008b). We further explored the effects of coordination on learning (Wang & Gratch, 2009a). We hypothesized that contingent feedback would be perceived as helpful and can help students learn better. The learning task of users was storytelling interactions also used in other studies (Gratch et al., 2007; Kang et al., 2008a, 2008b]. We found that users reported higher rapport with the contingent feedback than the noncontingent one. We also discovered that rapport helped users' interaction and enhanced their self-performance, which improved learning gains. Collectively, these findings are in line with Tickle-Degnen and Rosenthal's predictions.

Mutual Attentiveness

Mutual attentiveness occurs in a conversation when both participants attend closely to each other's words or movements. In our empirical research, we have operationalized this concept in terms of gaze—for example, are participants looking at each other during a conversation? Prior research in gaze suggests there should be a curvilinear relationship between gaze and feelings of rapport. In other words, continuously staring at another person or completely avoiding his or her gaze would tend to be disruptive, but, short of these extremes, rapport should be enhanced by more visual attention.

Consistent with this, we found a similar curvilinear relationship between gaze and rapport. For example, an agent that continuously gazed at the participant without companying other timely positive gestures caused more distractions, less rapport, and more speech disfluency in storytelling interaction (Ning & Gratch, 2010). Similarly, an agent that failed to maintain gaze with the participant was equally disruptive. Collectively, these findings are in line with Tickle-Degnen and Rosenthal's predictions.

Characteristics of Participants That Impact Rapport

The previous studies and findings emphasized the impact of differences in the agent's behavior, but numerous social psychological studies emphasize that the trajectory of a social interaction is heavily shaped by the "baggage" people bring to a situation. In human-to-human interactions, people who are extroverted will more easily establish rapport than introverts, and we might expect that these dispositional tendencies will carry over into interactions with the Rapport Agent. Indeed, in a series of studies we have found that dispositional factors shape interactions with virtual humans in a similar manner to their influence in face-to-face interactions between people.

From the social science literature, we would expect that several dispositional traits might moderate the Rapport Agent's social impact. Broad personality dimensions, such as the "Big Five" (Costa & McCrae, 1992)—that is, extraversion, agreeableness, openness, conscientiousness, and neuroticism—influence many aspects of social behavior, and previous research illustrates that such effects can carry over to virtual interactions (Bickmore, Gruber, & Picard, 2005; Yee, Harris, Jabon, & Bailenson, 2010, in press). More specific personality traits also seem relevant to the study of rapport. For example, we might expect the impact of the Rapport Agent's nonverbal behavior to depend on participants' *public self-consciousness* (Scheier & Carver, 1985), which refers to their sensitivity to attend to the actions of others and adjust their own behavior to create favorable impressions. Rapport can also be impacted by one's tendency to be anxious in social situations: talking with a stranger (or an even stranger virtual human) can be threatening, and people vary in their tendency to react fearfully to novel social situations. Thus, measures such as shyness or social anxiety may illuminate mechanisms that underlie the impact of virtual humans (Cheek, 1983).

Indeed, several dispositional factors have been found to influence the effectiveness of the Rapport Agent. With respect to the Big Five, extroversion and agreeableness influence interactions in ways that are consistent with their impact in human-to-human interactions. Extroverts tend to talk more, and more fluently, and feel better about their interaction, and similar findings hold for participants who score high in agreeableness (Kang, Gratch, Wang, & Watts, 2008a; von der Pütten, Krämer, & Gratch, 2010). Social anxiety also plays a moderating role: interestingly, we found dispositionally anxious subjects felt more trust toward the Rapport Agent than in their interactions with human conversational partners (Kang, Gratch, Wang, & Watt, 2008b).

Dispositional factors don't uniquely determine the outcome of a social interaction but, rather, interact with aspects the situation. For example, someone who is confident in social situations might perform well regardless of the behavior of his or her conversational partner. However, someone who is less secure might seek constant reassurance by carefully monitoring his or her partner's nonverbal feedback: if this feedback is positive, he or she may perform well and report positive feelings; if this feedback is negative, the opposite may occur. We see similar interaction effects in our data (Kang, Gratch, Wang, & Watt, 2008a). For example, extroverts seem insensitive to manipulations that impact the quality of the Rapport Agent's nonverbal feedback, whereas participants who score high in social anxiety are quite disrupted by negative feedback (such as that produced by the noncontingent agent). We found that people having greater social anxiety felt less rapport and more embarrassment, and exhibited worse performance (e.g., more disfluent speech) when presented with noncontingent feedback.

Overall, our studies suggest that both agent behavior and dispositional factors interact to determine the overall quality of an experience with a virtual human. Further, these effects are largely consistent with predictions from the social psychological literature on human-to-human interactions. However, virtual humans are clearly different from people in many respects, and we now turn to turn to theoretical predictions that suggest these differences can profoundly shape the nature of human interactions with such technological artifacts.

Media Effects

The previous findings suggest that a social psychological theory such as rapport has some explanatory value in predicting how behaviors of an artificial agent will influence human subjective impressions and behavior. However, there are many important differences between humans and virtual humans, and some theoretical perspectives argue that these differences are crucial in determining the trajectory of a human/virtual human interaction. In Chapter 3 of this volume, Blascovich describes the threshold model of social influence within digital immersive virtual environments, which posits that a number of "media factors" can promote or disrupt the applicability of standard social psychological findings to virtual settings. For example, *agency* refers to the "perceived human sentience" (meaning the attribution of consciousness and free will) of a human-appearing agent. In the context of the Rapport Agent, *agency* would refer to whether participants believed the Rapport Agent's listening behavior corresponded to the actual movements of a sentient human participant in another room (i.e., that the Rapport Agent was an *avatar* of another participant) or whether they believed they were generated by a computer agent. The threshold model predicts that social effects of the Rapport Agent will be moderated by agency attributions. Numerous studies by different research groups show that people react socially toward both forms of representations—agents and avatars (c.f., e.g., Bailenson, Blascovich, Beall, & Loomis, 2004; Bickmore, Gruber, & Picard, 2005; Cassell et al., 2001; Gratch, Wang, Gerten, Fast, & Duffy, 2007; Krämer, Iurgel, & Bente, 2005; Nass, Moon, Morkes, Kim, & Fogg, 1997; Reeves & Nass, 1996; see Krämer, 2008, for an overview). However, it is still unclear whether people react in the same way toward agents and avatars.

We investigated whether users' belief in interacting with either an avatar (a virtual representation of a human) or an agent (autonomous virtual person) leads to different social effects in self-disclosure interview interaction (von der Püten et al., 2010). We specifically tested two dominant models of social effects in human–agent interaction: the threshold model of social influence and Nass' Ethopoeia concept. According to the threshold model of social influence (Blascovich et al., 2002), the social influence of real persons who are represented by avatars will always be high, whereas the influence of an artificial entity depends on the realism of its behavior. Conversely, the Ethopoeia concept (Nass & Moon, 2000) predicts that automatic social reactions are triggered by situations as soon as they include social cues. We found that the outcome did not demonstrate a significant difference in users' perception of agency between an agent and an avatar. On the other hand, manipulating the presence of social cues (i.e., the presence or absence of positive, contingent feedback on the part of the agent) significantly influenced users' subjective experience as well as their actual behavior: Users experienced more feelings of mutual awareness with their partner and revealed more information in the interaction when social cues of rapport were present. This outcome supports the assumption that "the more computers present characteristics that are associated with humans, the more likely they are to elicit social behavior" (Nass & Moon, 2000, p. 97). Like in face-to-face interactions, the evaluation of people is first and foremost dependent on what agents do—even if it is merely subtle differences in nonverbal behavior. In conclusion, the Ethopoeia concept by Nass and colleagues is more suitable as an approach to explain the social effects we found than the threshold model of social influences by Blascovich and colleagues.

Discussion and Conclusion

Across a wide range of studies, we have consistently shown that simple nonverbal cues on the part of a computer program can provoke a wide range of beneficial subjective and behavioral outcomes. Overall, our studies and related findings add further evidence that the nonverbal behavior of embodied virtual agents influences the behavior of the humans who interact with them in ways that are consistent with psychological findings on human-to-human interaction. Further, these effects increase as a virtual agent exhibits more human-like behavioral characteristics. More specifically, studies support Tickle-Degnen and Rosenthal's claims that rapport is promoted by social behavioral cues that are positive and contingent and convey mutual attention, and that these effects are moderated by personality traits of the human user. Interestingly, participant's belief as to whether he or she was interacting with another human or a computer program did not have much effect, reinforcing claims by Nass and colleagues that computer programs will produce social effects to the extent that they incorporate human-like social cues, and reinforcing the validity of using social psychological theory to inform agent design.

Despite the apparent success of the rapport agent, one should be cautious before concluding that people will always be so easily manipulated by simple nonverbal behaviors. According to the social inference taxonomy presented in the introductory chapter of this book (see Figure 3of the Introduction), the Rapport Agent is leveraging social reactivity (i.e., certain behaviors such as nods or smiles might trigger automatic responses in participants) and simple social inference (i.e., by giving certain expressive behaviors at the right time, the participant is—via some process such as reverse appraisal—attributing mental state to the agent even if this state is an illusion) but is otherwise limited. Our experimental settings (storytelling and interviews) are, by design, simple for agents to navigate. The Rapport Agent conveys understanding without actual understanding, a behavior that most of us have engaged in from time to time (for example, when carrying on a conversation in a unfamiliar language or in a noisy room), but such a charade only goes so far before it ends in embarrassment. For example, in a similar storytelling paradigm, Janet Bavelas illustrated that "generic" feedback (similar to what the Rapport Agent provides) is easy to produce without actually attending to the meaning of a conversation—she had participants listen while solving complex mathematical problems—but at certain points speakers need more meaningful "specific" feedback. Bavelas et al. (2000) found that when speakers were telling personally emotional stories, they expected emotional feedback at key dramatic moments. When they failed to receive it, they felt embarrassment and had difficulty proceeding with their story. The inability for the Rapport Agent to produce meaningful feedback is a clear limitation, and its mindless feedback can easily backfire if participants realize its true nature.

Part of our research can be seen as pushing the boundary of just how far one can go with simple contingent feedback. Our early studies explored "safe" and impersonal content such as Tweety and Sylvester cartoons. We subsequently examine the more daring content such as sexual harassment. Most recently, we've been asking people to elicit very personal information such as their greatest disappointments. At each stage we continue to show robust subjective and behavioral effects of contingent positive feedback; however, we must be careful before concluding that the agent is performing well. What we've shown is that the agent performs about as well as one would when discussing personal matters with a stranger (something many people feel is a stressful experience) and better than an agent that provides no or negative feedback. While important, this is a low bar, and much can be done to improve the performance and effects of such agents.

Recently, we've been speculating that rapport can be conceptualized as a phenomenon occurring on three levels: the emotional, the behavioral, and the cognitive. Emotionally, rapport is an inherently rewarding experience; we feel liking, emotional support, and acceptance. Cognitively, we share an understanding with our conversation partner; there is a convergence of beliefs or views, a bridging of ideas or perspectives. Behaviorally (or interactionally), there is a convergence of movements with our conversational partner; observers report increased synchrony, fluidity, and coordination in partners' movements. We argue that the Rapport Agent produces effects at each of these levels (people feel liking and support; they believe they are understood; they recognize that their partner is in sync with their behavior), but only the behavioral level is truly implemented.

At the cognitive level, agent head nods merely create the illusion of mutual understanding. According to theories of conversational grounding, speakers in a conversation expect frequent and incremental feedback from listeners that their communication is understood (Nakano, Reinstein, Stocky, & Cassell, 2003; Traum, 1994). When listeners provide grounding feedback, speech can proceed fluently and presumably with a greater sense of rapport. Such feedback often takes the form of nods, such as produced by the Rapport Agent, and presumably speakers are (incorrectly) interpreting these nods as grounding cues. This illusion can be maintained to an extent, but as Bavelas' example above illustrates, it will eventually backfire and lead participants to view such feedback with suspicion. If we want to create true rapport, a major challenge in creating true rapport is how to recognize and respond meaningfully to a speaker with the rapidity seen in human dyads. Indeed, recent research in dialogue systems is making progress in this direction (Sagae, DeVault, & Traum, 2010).

At the emotional level, agent head nods are likely creating the illusion of empathy and emotional support. When a participant tells a positive story, as in a Tweety and Sylvester cartoon, this illusion is easily maintained. However, when participants are relating their fears and disappointments, they expect corresponding expressions of sympathy or empathy on the part of their conversational partner: when someone has just revealed that he or she was sexually abused, simply nodding doesn't cut it and will likely backfire. If we want to create "real" rapport, a system must not only extract meaning from a participant's speech (a prerequisite of the cognitive level or rapport) but also exhibit coherent and supportive emotional reactions to this content. Our earlier chapter on EMA (Chapter 4) illustrates how such emotional reactions could be calculated, and current efforts are aimed at integrating appraisal with the low-level behavioral feedback produced by that Rapport Agent.

Conclusion

In this book, we have argued that the study of human emotion both informs and is informed by computational methods, and the Rapport Agent is a vivid example of this relationship. By starting with a psychological theory of social processes (in this case the theory of rapport by Tickle-Degnen and Rosenthal), we are able to construct a social artifact that instantiates the theory. The resulting artifact then becomes a means for testing the theory. According to rapport theory, simple contingent behaviors could produce feelings of mutual understanding and connectedness, and more effective social interactions. But the causality of this theory is hard to confirm. Are these contingent behaviors causes or consequences of successful social interactions? Artifacts like the Rapport Agent provide the means to experimentally tease apart these factors in ways that avoid the disadvantages and potential confounds introduced by human confederates. In this chapter, we outlined a series of studies that demonstrate that, indeed, contingent behaviors cause rapport. Thus, social theory has allowed us to build a better computer and return the favor through experimental support for theory. This is a true partnership between the social and computational sciences of social emotions.

Note

1. Although we adopt the terminology and definitions of Tickle-Degnen and Rosenthal, the nonverbal patterning seen in rapport is closely related to a number of social science concepts including *interpersonal sensitivity* (Hall & Bernieri, 2001), *social glue* (Lakin et al., 2003), *interactional synchrony* (Bernieri & Rosenthal, 1991), *behavioral mimicry* (Chartrand & Bargh, 1999), backchanneling (Yngve, 1970), and *interpersonal adaptation* (Burgoon, Stern, & Dillman, 1995).

References

Adam, H., Shirako, A., & Maddux, W. W. (2010). Cultural variance in the interpersonal effects of anger in negotiations. *Psychological Science*, *21*(6), 882–889.

Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. *Advances in Experimental Social Psychology*, 32, 201–271.

Bailenson, J., Beall, A., Loomis, J., Blascovich, J., & Turk, M. (2004). Transformed social interaction: Decoupling Representation from Behavior and Form in Collaborative Virtual Environments. *PRESENCE: Teleoperators and Virtual Environments*, *13*(4), 428–441.

Bailenson, J., & Yee, N. (2005). Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological Science*, *16*, 814–819.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, *71*(2), 230–244.

Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology*, *79*(6), 941–952.

Bente, G., Kraemer, N. C., Petersen, A., & de Ruiter, J. P. (2001). Computer animated movement and person perception: Methodological advances in nonverbal behavior research. *Journal of Nonverbal Behavior*, *25*(3), 151–166.

Bernieri, F. J. (1988). Coordinated movement and rapport in teacher-student interactions. *Journal of Nonverbal behavior*, *12*(2), 120–138.

Bernieri, F. J., Gillis, J. S., Davis, J. M., & Grahe, J. E. (1996). Dyad rapport and the accuracy of its judgment across situations: A lens model analysis. *Journal of Personality and Social Psychology*, *71*(1), 110–129.

Bernieri, F. J., & Rosenthal, R. (1991). Interpersonal coordination: Behavior matching and interactional synchrony. In R. S. Feldman & B. Rimé (Eds.), *Fundamentals of Nonverbal Behavior*. Cambridge, UK: Cambridge Univ. Press.

Bickmore, T., Gruber, A., & Picard, R. W. (2005). Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient Education and Counseling*, 59(1), 21–30.

Biocca, F., & Harms, C. (2002, October). *Defining and measuring social presence: Contribution to the networked minds theory and measure*. Paper presented at the Presence, Porto, Portugal,

Blascovich, J. (2002). Social influence within immersive virtual environments. In R. Schroeder (Ed.), *The social life of avatars* (pp. 127–145): Springer-Verlag.

Brand, M. (1999, October). *Voice puppetry*. Paper presented at the 26th International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH), Los Angeles.

Breazeal, C., & Aryananda, L. (2002). Recognition of affective communicative intent in robot-directed speech. *Autonomous Robots*, *12*, 83–104.

Burgoon, J. (1983). Nonverbal violations of expectations. In J. M. Wiemann & R. P. Harrison (Eds.), *Nonverbal interaction* (pp. 11–77). Beverly Hils, CA: Sage.

Burgoon, J., Stern, L. A., & Dillman, L. (1995). *Interpersonal adaptation: Dyadic interaction patterns*. Cambridge, UK: Cambridge University Press.

Burns, M. (1984). Rapport and relationships: The basis of child care. *Journal of Child Care*, *2*, 47–57.

Cassell, J. (2000). Nudge nudge wink wink: Elements of Face-to-Face Conversation for Embodied Conversational Agents. In J. Cassell, J. Sullivan, S. Prevost, & E. Churchill (Eds.), *Embodied conversational agents* (pp. 1–27). Cambridge, MA: MIT Press.

Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjálmsson, H., et al. (1999, May). *Embodiment in conversational interfaces: Rea*. Paper presented at the Conference on Human Factors in Computing Systems, Pittsburgh, PA.

Cassell, J., Nakano, Y. I., Bickmore, T. W., Sidner, C. L., & Rich, C. (2001, July). *Non-verbal cues for discourse structure*. Paper presented at the Association for Computational Linguistics Joint EACL - ACL Conference, Toulouse, France.

Cassell, J., & Thórisson, K. R. (1999). The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *International Journal of Applied Artificial Intelligence*, *13*(4–5), 519–538.

Cathcart, N., Carletta, J., & Klein, E. (2003, April). *A shallow model of backchannel continuers in spoken dialogue*. Paper presented at the 10th Conference of the European Chapter of the Association for Computational Linguistics, Budapest, Hungary.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*(6), 893–910.

Cheek, J. M. (1983). *The Revised Cheek and Buss Shyness Scale (RCBS)*. Wellesley, MA: Wellesley College.

Clark, H. H., & Wasow, T. (1998). Repeating words in Spontaneous Speech. *Cognitive Psychology*, *37*, 204–242.

Cogger, J. W. (1982). Are you a skilled interviewer? *Personnel Journal*, *61*, 840–843.

Costa, P. T., Jr., & McCrae, R. R. (1992). *NEO PI-R professional manual*. Odessa, FL: Psychological Assessment Resources, Inc.

Drolet, A. L., & Morris, M. W. (2000). Rapport in conflict resolution: Accounting for how face-to-face contact fosters mutual cooperation in mixed-motive conflicts. *Experimental Social Psychology*, *36*, 26–50.

Forgas, J. P. (2007). The use of computer-mediated interaction in exploring affective influences on strategic interpersonal behaviours. *Computers in Human Behavior*, *23*(2), 901–919.

Fuchs, D. (1987). Examiner familiarity effects on test performance: Implications for training and practice. *Topics in Early Childhood Special Education*, *7*, 90–104.

Gratch, J. (2008). True emotion vs. social intentions in nonverbal communication: Towards a synthesis for embodied conversational agents. In I. Wachmuth & G. Knoblich (Eds.), *Modeling communication with robots and virtual humans* (Vol. LNAI 4930, pp. 181–197). Berlin: Springer.

Gratch, J. (2010). Emotionally-resonant media: Advances in sensing, understanding and influencing human emotion through interactive media. In K. Döveling, C. von Scheve, & E. Konijn (Eds.), *Handbook of emotions and the mass media* (pp. 370–387). London: Routledge, Taylor and Francis Group.

Gratch, J., Okhmatovskaia, A., Lamothe, F., Marsella, S., Morales, M., van der Werf, R., et al. (2006, August). *Virtual rapport*. Paper presented at the 6th International Conference on Intelligent Virtual Agents, Marina del Rey, CA.

Gratch, J., Wang, N., Gerten, J., & Fast, E. (2007). *Creating rapport with virtual agents*. Paper presented at the 7th International Conference on Intelligent Virtual Agents, Paris, France.

Hall, J., & Bernieri, F. J. (Eds.). (2001). *Interpersonal sensitivity: Theory and measurement*. Mahwah, NJ: Psychology Press.

Huang, L., Morency, L.-P., & Gratch, J. (2010, May). *Parasocial consensus sampling: Combining multiple perspectives to learn virtual human behavior*. Paper presented at the 9th International Conference on Autonomous Agents and Multiagent Systems, Toronto, Canada.

Huang, L., Morency, L.-P., & Gratch, J. (2011, September). *Virtual rapport 2.0*. Paper presented at the Proceedings of the 10th International Conference on Intelligent Virtual Agents, Reykjavík, Iceland.

Kallmann, M., & Marsella, S. (2005, September). *Hierarchical motion controllers for real-time autonomous virtual humans*. Paper presented at the 5th International Working Conference on Intelligent Virtual Agents, Kos, Greece.

Kang, S.-H., Gratch, J., Wang, N., & Watt, J. (2008a, September). *Agreeable people like agreeable virtual humans*. Paper presented at the 8th International Conference on Intelligent Virtual Agents, Tokyo.

Kang, S.-H., Gratch, J., Wang, N., & Watt, J. (2008b, May). *Does contingency of agents' nonverbal feedback affect users' social anxiety?* Paper presented at the 7th International Conference on Autonomous Agents and Multiagent Systems, Estoril, Portugal.

Kang, S.-H., Gratch, J., Wang, N., & Watt, J. (2009, April). *The effect of affective iconic realism on anonymous interactants' self-disclosure*. Paper presented at Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, MA.

Kopp, S., Krenn, B., Marsella, S., Marshall, A., Pelachaud, C., Pirker, H., et al. (2006, August). *Towards a common framework for multimodal generation in ECAs: The behavior markup language*. Paper presented at the Intelligent Virtual Agents, Marina del Rey, CA.

Krämer, N. C., Iurgel, I., & Bente, G. (2005). Emotion and motivation in embodied conversational agents. Paper presented at the Proceedings of the Symposium "Agents that Want and Like", Artificial Intelligence and the Simulation of Behavior (AISB), Hatfield.

Krämer, N. C. (2008, January). Social effects of virtual assistants. a review of empirical results with regard to communication. In *Intelligent virtual agents* (pp. 507–508). Springer Berlin Heidelberg.

Lakin, J. L., Jefferis, V. A., Cheng, C. M., & Chartrand, T. L. (2003). Chameleon effect as social glue: Evidence for the evolutionary significance of nonconsious mimicry. *Journal of Nonverbal Behavior*, *27*(3), 145–162.

Lamothe, F., & Morales, M. (2006). *Response behavior* (No. ICT TR 01.2006). Marina del Rey, CA: University of Southern California.

Lanzetta, J. T., & Englis, B. G. (1989). Expectations of cooperation and competition and their effects on observers' vicarious emotional responses. *Journal of Personality and Social Psychology*, *45*, 543–554.

McFarland, D. H. (2001). Respiratory markers of conversational interaction. *Journal of Speech, Language, and Hearing Research*, *44*, 128–143.

Mead, G. H. (1934). *Mind, self and society*. Chicago: University of Chicago Press.

Morency, L.-P., de Kok, I., & Gratch, J. (2008, September). *Predicting listener backchannels: a probabilistic multimodal approach*. Paper presented at the 8th International Conference on Intelligent Virtual Agents, Tokyo.

Morency, L.-P., Sidner, C., Lee, C., & Darrell, T. (2005, October). *Contextual recognition of head gestures*. Paper presented at the 7th International Conference on Multimodal Interactions, Torento, Italy.

Nakano, Y., Reinstein, G., Stocky, T., & Cassell, J. (2003, July). *Towards a model of face-to-face grounding*. Paper presented at the Annual Meeting of the Association for Computational Linguistics, Sapporo, Japan.

Nass, C. I., Moon, Y., Morkes, J., Kim, E. Y., & Fogg, B. J. (1997). *Computers are social actors: a review of current research, Human values and the design of computer technology*. Stanford, CA: Center for the Study of Language and Information.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103.

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

Roloff, M. E., & Campion, D. E. (1987). Communication and reciprocity within intimate relationships. In M. E. Roloff & G. R. Miller (Eds.), *Interpersonal processes: New directions in communication research* (pp. 11–38). Beverly Hills, CA: Sage.

Sagae, K., DeVault, D., & Traum, D. (2010, June). *Interpretation of partial utterances in virtual human dialogue systems*. Paper presented at the 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics.

Scheier, M. F., & Carver, C. S. (1985). The Self-Consciousness Scale: A revised version for use with general populations. *Journal of Applied Social Psychology*, *15*, 687–699.

Schröder, M., Cowie, R., Heylen, D., Pantic, M., Pelachaud, C., & Schuller, B. (2008, October). *Towards responsive Sensitive Artificial Listeners*. Paper presented at the Fourth International Workshop on Human-Computer Conversation.

Sonnby-Borgstrom, M., Jonsson, P., & Svensson, O. (2003). Emotional empathy as related to mimicry reactions at different levels of information processing. *Journal of Nonverbal Behavior*, 27(1), 3–23.

Tatar, D. (1997). *Social and personal consequences of a preoccupied listener*. Stanford, CA: Stanford University

Thiebaux, M., Marshall, A., Marsella, S., & Kallmann, M. (2008, May). *SmartBody: Behavior realization for embodied conversational agents*. Paper presented at the International Conference on Autonomous Agents and Multi-Agent Systems, Estoril, Portugal.

Tickle-Degnen, L., & Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychological Inquiry*, *1*(4), 285–293.

Tiedens, L. Z., & Fragale, A. R. (2003). Power moves: Complementarity in dominant and submissive nonverbal behavior. *Journal of Personality and Social Psychology*, *84*(3), 558–568.

Tosa, N. (1993, July). Neurobaby. 20th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH). San Diego, CA.

Traum, D. (1994). *A computational theory of grounding in natural language conversation*. (Unpublished doctoral dissertation). University of Rochester, Rochester, NY.

Tsui, P., & Schultz, G. L. (1985). Failure of rapport: Why psychotheraputic engagement fails in the treatment of Asian clients. *American Journal of Orthopsychiatry*, *55*, 561–569.

Van baaren, R. B., Holland, R. W., Kawakami, K., & van Knippenberg, A. (2004). Mimicry and prosocial behavior. *Psychological Science*, *15*(1), 71–74.

von der Pütten, A., Krämer, N., & Gratch, J. (2009, November). *Who's there? Can a virtual agent really elicit social presence?* Paper presented at the 12th Annual International Workshop on Presence, Los Angeles, CA.

von der Pütten, A., Krämer, N., & Gratch, J. (2010). How our personality shapes our interactions with virtual characters - Implications for research and development. In J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, & A. Safonova (Eds.), *Intelligent virtual agents* (Vol. 6356, pp. 208–221). Springer: Berlin/Heidelberg.

Pütten, A. M. von der, Krämer, N. C., Gratch, J., & Kang, S.-H. (2010). It doesn't matter what you are! Explaining social effects of agents and avatars. *Computers in Human Behavior*, *6*(6), 1641–1650.

Wang, N., & Gratch, J. (2009a, July). *Can a virtual human build rapport and promote learning?* Paper presented at the 14th International Conference on Artificial Intelligence in Education, Brighton, UK.

Wang, N., & Gratch, J. (2009b, September). *Rapport and facial expression*. Paper presented at the International Conference on Affective Computing and Intelligent Interaction, Amsterdam, The Netherlands.

Wang, N., & Gratch, J. (2010, April). *Don't just stare at me*. Paper presented at the 28th Annual CHI Conference on Human Factors in Computing Systems, Atlanta, GA.

Ward, N., & Tsukahara, W. (2000). Prosodic features which cue back-channel responses in English and Japanese. *Journal of Pragmatics*, *23*, 1177–1207.

Warner, R. (1996). Coordinated cycles in behavior and physiology during face-to-face social interactions. In J. H. Watt & C. A. VanLear (Eds.), *Dynamic patterns in communication processes* (pp. 327–352). Thousand Oaks, CA: SAGE publications.

Warner, R. M., Malloy, D., Schneider, K., Kooth, R., & Wilder, B. (1987). Rhythmic organization of social interaction and observer ratings of positive affect and involvement. *Journal of Nonverbal Behavior*, *11*(2), 57–74.

Welji, H., & Duncan, S. (2004, August). *Characteristics of face-to-face interactions, with and without rapport: Friends vs. strangers*. Paper presented at the Symposium on Cognitive Processing Effects of 'Social Resonance' in Interaction, 26th Annual Meeting of the Cognitive Science Society, Chicago, IL.

Yee, N., Harris, H., Jabon, M., & Bailenson, J. N. (2011). The expression of personality in virtual worlds. *Social Psychology and Personality Science*, *2*(1), 5–12.

Yngve, V. H. (1970, April). *On getting a word in edgewise*. Paper presented at the Sixth Regional Meeting of the Chicago Linguistic Society, Chicago, IL.