Understanding the mind by simulating the body:

virtual humans as a tool for cognitive science research

Jonathan Gratch

Department of Computer Science

and

the Institute for Creative Technologies

University of Southern California

12015 Waterfront Dr.

Playa Vista, CA 90094

gratch@ict.usc.edu

310-448-0306

in *The Oxford Handbook of Cognitive Science*. Susan Chipman (ed.).

Oxford University Press (in press)

Understanding the mind by simulating the body: virtual humans as a tool for cognitive science research

In helping to define the field of cognitive science, Herb Simon emphasized the importance of "understanding by simulating" (Simon 1969, 17-22). By this, he was stressing key insights that arise from viewing psychological theory from the perspective of computation. For example, the exercise of translating a conceptual theory into a working computer program forces a researcher to concretize his or her assumptions and often reveals gaps or internal inconsistencies in the original theory. Further, working programs allow a researcher to empirically study how theoretical assumptions interact with each other and unfold over time as they act on complex environments. As Simon vividly illustrated through is metaphor of an ant on the beach, such empirical simulations can often reveal that the apparent complexity of cognitive processes can be illusory – that complexity can arise from dynamic interaction of simple processes – and that such consequences are difficult to envision without the benefit of simulation (e.g., see Simon 1969 pp. 63-64; Gratch 2012). Cognitive scientists have taken this advice to heart, spawning the subfield of cognitive modeling which, in turn has resulted in detailed and predictive computational simulations of many aspects of human behavior. Cognitive modeling research has undoubtedly yielded important insights into cognitive processes, yet it abstracts away many important details about how minds interact with the environment. In this chapter I argue that only by simulating the body, and the important constraints it imposes on cognition, can we truly understand the mind.

Cognitive simulations have provided important theoretical insights and advanced the field, but they have often been criticized as being too reductionist – breaking cognition into individual components without consideration of how these parts fit into a complete whole (e.g., Barsalou 2008; Rosenbloom 2011). For example, researchers may develop detailed stand-alone models human memory, learning or decision-making without concern or how these models might interoperate. Partly in response to such criticisms, several cognitive scientists have begun to investigate more comprehensive models of cognitive capabilities under such rubrics as general artificial intelligence or human-level artificial intelligence. For example, Bringsjord and colleagues have attempted to create cognitive models that can understand and solve problems from standard IQ tests (Bringsjord and Schimanski 2003). But as any social psychologist would immediately object, IQ tests and other conventional measures of intelligence are hardly a comprehensive measure of the rich range of cognitive skills people must exhibit in their daily lives. Indeed, one aspect of everyday cognition that is inescapable but largely absent from cognitive simulations is the skills associate with the human body.

From the perspective of cognitive science, virtual humans provide an exciting domain within which to examine a broader and more integrated array of problems that are traditionally explored by cognitive scientists. Virtual humans are computer generated anthropomorphic characters that exist within virtual worlds but are designed to perceive, understand and interact with real-world humans. Virtual humans are essentially digital robots: they aspire to address many of the research issues that arise from physical robots but dispense from many of the engineering challenges that robotacists must confront. Although typically conceived as practical tools to assist in a range of application (such as training, health care and entertainment), in this chapter I will outline their potential for advancing cognitive science research. In their most general form, virtual humans aspire to simulate the cognitive abilities of people, but also many of the "embodied" aspects of human behavior more traditionally studied in fields outside of cognitive psychology, such as nonverbal behavior recognition and production. More importantly, these abilities are not ends in themselves, but rather the means for virtual humans to successfully interact with people, and the success of these interactions is a gold standard by which to judge the success of underlying cognitive models. Through this emphasis on integration of multiple cognitive capabilities (e.g., language, gesture, emotion, and the control problems associated with navigating and interacting with a simulated virtual world) and the requirement that these capabilities work together, in real-time, to interact with human partners, creates a unique and challenging environment within which to develop and validate cognitive theories and models.

In this chapter, I will discuss the general technical problems and current techniques involved in creating virtual humans and briefly discuss some of the applications to which state-of-the-art virtual humans have been applied. I will then emphasize the potential of virtual humans as a challenge problem for advancing the state of cognitive models. I will review some existing research that holds this goal and outline some future open issues and directions.

Virtual Humans

Virtual humans are software artifacts that look like, act like and interact with humans but exist in virtual environments. The motivations for constructing such artifacts are many and varied. Within human-computer interaction research, recent years have seen an interest in making computation more "human-centric" meaning, in part, that computers should strive to interact with people in ways that are closer to how people interact with each other. Thus, providing computers with the ability to communicate through natural language and gesture can potentially make human-computer interaction more effective and efficient. Within media psychology, virtual humans have gained interest for they influence they have over human impressions and decisions. For example, extensive research has shown that people treat computers very differently when they embody human-like characteristics – they treat them more like other people (Reeves and Nass 1996; Pütten, Krämer et al. 2010). This phenomenon opens the possibility to use virtual humans to study human social processes, but also has important practical implications. For example, people are more persuaded by a message delivered by a virtual human, making the well-suited to communicating medical advice (Bickmore, Gruber et al. 2005) or selling products.[1]

Building virtual humans draws on expertise from a variety of disciplines. In that virtual humans should look, think and move like people, simulation researchers draw heavily on the human sciences. For example, virtual human "brains" are often inspired by psychological theories of human cognition (Laird 2001; Swartout, Gratch et al. 2006), language (Cassell and Stone 1999; Traum 2008) and emotion (Gratch and Marsella 2005). Virtual human bodies, as well, are informed by knowledge of physiological and biomechanical processes (Badler, Phillips et al. 1993; Maciel, Nedel et al. 2002; Honglun, Shouqian et al. 2007). Translating these theories and findings into working software requires the integration of advanced capabilities from a number of domains of computer science research including machine perception, artificial intelligence, cognitive modeling, graphics and animation. Although specific training applications may not require it, in their most general form, virtual humans must sense, interpret and rapidly respond to a variety of events and situations (see Figure 1). To mimic the capabilities of people, virtual humans must perceive the verbal and nonverbal behaviors of human users as well as

[1] e.g., see Alaska Airlines "Ask Jenn": https://www.alaskaair.com/alaskawebui/Agent.aspx

events in both the physical and virtual environments that the user and virtual human share. They must be able to fuse these multiple modalities, together with prior observations and domain knowledge, into a coherent understanding of the physical and social environment. They must be able to act on the user and the world by constructing and executing plans and communicating through natural language including all the non-verbal communication that accompanies human speech (e.g., eye contact and gaze aversion, facial displays, and gestures). All of these activities must work in close coordination and be subject to strict real-time requirements, for example, responding to a user's question in about 250ms. The field of virtual human research is making rapid progress in both these individual component capabilities and their integration into a comprehensive working system.

Insert Figure 1 about here.

The complexity of creating a virtual human can be daunting, given the number of capabilities that must be integrated into a working system. Fortunately, some of the research in this area has focused on the development of modular, sharable software architectures to facilitate application development. For example, several core virtual human capabilities are now freely available for research purposes. This includes software for sensing human signals including facial expressions (Bartlett, Littlewort et al. 2006), gestures (Morency, Sidner et al. 2005) and speech (Schuller, Lehmann et al. 2009). Other software exists for selecting realistic social behaviors (Lee & Marsella, 2006) and synthesizing those behaviors (Thiebaux, Marshall et al. 2008). There are even integrated packages that provide fully functioning virtual humans and authoring tools for certain application genres (e.g., see the Virtual Human Toolkit at http://vhtoolkit.ict.usc.edu/ or the Relational Agent project http://relationalagents.com/demos/index.html).

Although still limited in their capabilities, virtual humans have been successfully applied to a variety of application domains. One promising application are has been their use as “virtual role players” for interpersonal-skills training (e.g., Hart, Gratch et al. in press). For example, several groups have developed virtual medical patients that medical students can interact with to practice their interviewing and examination skills (Johnsen, Raij et al. 2007; Parsons, Kenny et al. 2008). The Tactical Questioning system allows for military personnel to practice talking to individuals that they may meet while on patrol or individuals that may have information of military value (Traum, et al., 2007) and similar applications have been developed for law enforcement training that include subject interviewing skills training and line-up identification (Frank, et al., 2002; Daugherty, Babu, Cutler, & Hodges, 2007). The BiLAT game-based simulation environment allows students to practice their negotiation skills (Kim, et al., 2009). In each of these systems, the user or training participant interacts with a computer controlled virtual human in a conversational style similar to the way a human would talk to another human or interact in a live role-playing simulation.

In addition to training applications, virtual humans are increasingly employed to interact with general users over the web to accomplish a variety of tasks. One promising area has been health applications. These applications augment traditional web-based interaction with the more personalizing, rapport-building and persuasion techniques that people use in face-to-face interaction. For example, Bickmore and colleagues have illustrated how a human-like character and specific rapport-building behaviors can lead patients to better adhere to their medical treatments than other computer-based techniques (Bickmore and Pfeifer 2008).

Virtual Humans vs. Physical Robots

Robots have recently received considerable interest as a tool for expanding cognitive science theories (Pfeifer and Bongard 2007; Pezzulo, Barsalou et al. 2011) but virtual humans can be far better suited to exploring cognitive science problems than physical robots. Virtual human facilitate exploration of the same challenges involved in perceiving and acting in the world, but dispense with the engineering challenges involved in creating physical systems. These engineering challenges have tended to focus robotics research on the lower-level perceptual-motor problems of navigating and manipulating objects (e.g., Berkeley 2008) and yield robotic bodies with very different biomechanical constraints than human physiology imposes (Collins, Wisse et al. 2001). In contrast, virtual human research has been able to focus on tasks involving higher-level cognition such as language and social interaction and more realistic behaviors and bodily constraints.

Virtual humans' biggest advantage is methodological. As a tool for scientific studies, robots are severely limited. They are expensive, slow, difficult to customize and prone to mechanical failures. In contrast, virtual humans can be replicated without cost, run faster than real time, and often trivial to customize. This can be a huge advantage for empirical investigations. For example, suppose one wishes to explore theories of how emotional contagion influences crowds behaviors. With virtual humans, it is possible to run large scale monte carlo simulations of how crowds will flee from a disaster (Tsai, Bowring et al. 2011). Alternatively, suppose one wishes to investigate how a cognitive model parameter impacts social behavior. One could simultaneously run hundreds of participants over the web on different parameter settings (Dehghani, Gratch et al. 2012).

Virtual Humans in Cognitive Science

From a cognitive science perspective, virtual humans offer many opportunities to enrich cognitive science research. Here I outline ways in which virtual humans can be seen as challenge problem for cognitive science research (expanding on the range of inputs, outputs, and problem domains typically investigated in cognitive science research).

Virtual humans and embodied cognition

Cognitive science is often (and only somewhat unfairly) criticized for studying "brains in a vat" meaning that there has been a strong emphasis on cognitive processes but less on how these processes interact with the body or the physical environment. Cognition is often cleanly separated from the body and environment, with the body represented, if at all, as an input-output device that executes top-down cognitive commands. Of course, the situated and embodied nature of cognition has a long tradition in cognitive science. Simon emphasized that apparent cognitive complexity is best seen as the interaction of simple processes interacting with complex environments (Simon 1969) and many have argued persuasively how much of the information we use to solve problems is stored, not in memory but in the state of the world (e.g., see Suchman 1987). More recent research on "embodied cognition" emphasizes that even the state of our bodies contains important information that informs and influences cognitive processes (Wilson 2002; Barsalou 2008) and this work has begun to influence computational models of cognitive processes (e.g., see Ritter and Young 2001).

Work on embodying cognitive models, while proceeding, is often piecemeal, as vividly illustrated in Figure 2, which shows EPIC (Meyer and Kieras 1997), a cognitive architecture which integrates cognitive processes with several perceptual-motor capabilities while leaving out most of the actual body. In contrast, current embodied theories of cognition emphasize a more comprehensive and pervasive connection between bodily and cognitive process than this figure suggests.

Insert Figure 2 about here.

Embodied theories argue that the brain and body are tightly linked: the configuration and state of one's body can have a profound influence on cognitive processes and vice versa. Obviously, physiological processes such as hormones and cardiovascular arousal can shape decisions and actions (Schachter and Singer 1962). Perhaps more surprisingly, simply the physical configuration of the body also seems to influence thought. For example, posture can impact how easily we are persuaded (Petty, Wells et al. 1983) or how motivated we are to perform a task (Ahn, Teeters et al. 2007); gestures and language are closely coupled such that disruption in gesture production can impact speech (Iverson and Thelen 1999; McNeill 2005); facial expressions can influence our emotions and (Niedenthal, Mermillod et al. 2010) and peripheral physiological processes.

Embodied theories further argue that cognitive processes are heavily influenced by the body's relationship to aspects of the physical environment and the constraints environmental objects impose on physical action. For example, when wearing a heavy backpack, people overestimate the distance to a goal or the steepness of a hill (Iverson and Thelen 1999); the physical proximity of objects can influence their perceived value (Kahneman, Knetsch et al.

1991); and filling out questionnaires in a disgusting room can enhance judgments of moral outrage (Schnall, Haidt et al. 2008).

Finally, there is evidence that even the perception of one's own body influences thought. Some of the most interesting evidence for this comes from the use of virtual environments and virtual humans to manipulate these perceptions. For example, Mohler and colleagues used virtual environments to show that people use the size of their own bodies to estimate physical distances (Mohler, Bülthoff et al. 2008). Bailenson and colleagues showed that social characteristics like extroversion could be altered by manipulating the height or attractiveness of participants' virtual bodies in a virtual environment (Bailenson, Swinth et al. 2005).

Together, these findings suggest that cognitive science, and cognitive modeling research in particular, would benefit by a greater emphasis on the connection between cognitive and physical processes (see also Pezzulo, Barsalou et al. 2011), and that virtual humans can facilitate the exploration of this connection. Such a partnership would act as a forcing function, not only on the development of cognitive models but also virtual human research. With regard to cognitive models, it would require a greater focus on the relationship between cognitive and bodily processes. Interestingly, virtual human research suffers many of the same problems. Although virtual human researchers represent the physical body, they ignore many of the physiological and biomechanical processes implicated in embodied cognition. Further, although many virtual human researchers have rich models of physical bodies, these bodies are often disconnected from the world – for example, a character may be used to speak and gesture but, for all practical purposes, is floating in a vacuum. Taking the connection between the mind, body, and environment seriously can potentially produce a flowering of new research questions and technological advancement.

One example of this partnership is illustrated by the work of Sprague, Ballard and Robinson (2007) on visual perceptual-motor coordination. In this work, they combine a detailed cognitive model of visual perception with a virtual human model known as Walter that represents many of the kinetic constraints associated with human vision “in the wild.” The model is tested on several tasks that require visual scanning and navigation in the virtual environment. For example, the virtual human simulates a simple navigation task that involves walking down a street, avoiding obstacles and collecting trash. The body represents the kinematic constraints of moving in the environment as well as perceptual constraints associated with vision: e.g., a visual representation must be constructed from multiple rapid saccades. The model is compared with eye tracking data of human participants engaged in the same task and they are able to show the main features of the model are present in the human data.

Virtual Humans and social cognition

Humans are social animals that evolved to survive through social relationships, yet cognitive science and especially cognitive modeling research has tended to downplay the social antecedents and consequents of cognition. Everyday cognition is largely about relationships: influencing others, eliciting social support, managing reputations, seeking love, companionship and entertainment. And in these relationships, our physical bodies play a substantial role – we watch, we speak, and we touch. In tackling social cognition, however, classical cognitive scientists have avoided the need for reasoning about bodily processes by viewing the bodies of social actors as mere conduits that exchange information between brains (Reddy 1979). In this view, information in the brain of one social actor is *encoded* into bodily signals that are transmitted to other social actors and *decoded* in order to recover the original meaning (see

Figure 3). This abstraction allows the cognitive scientist to profitably study social cognition as an exchange of information, without becoming distracted by the messy details of how this information is generated and perceived.

Insert Figure 3 about here.

Several lines of research argue that this clean separation between mind and body leads to a distorted view of social cognition. Close observers of face-to-face communication emphasize that social interactions rarely involve situations where fully-formed messages are "transmitted" from one party to another. Rather, people are simultaneously transmitting and receiving a wide array of social signals and any "fully-formed" message in the mind of one conversation partner rarely survives first-contact with the other partner. For example, several researchers have emphasized how subtle bodily feedback (e.g., nods, smiles and furrowed brows) lead to rapid adjustments in how a message is delivered (Yngve 1970; Tickle-Degnen and Rosenthal 1990; Bavelas, Coates et al. 2000; Parkinson 2009) to such an extent that it is often argued that meaning exchanged between participants is best seen as negotiated or co-constructed (Bavelas, Coates et al. 2000).

In this sense, the body and bodily processes serve as important constraints that govern how two or more individuals can socially interact, and attending to these constraints can reveal much about social cognition. For example, bodily signals communicate meaning but only some of these signals are under voluntary control. Blushing, for instance, is often interpreted as a signal of guilt or embarrassment but most people cannot strategically choose to blush. In contrast, smiling indicates pleasure but is far easier to feign. Perhaps not surprisingly, blushing is viewed as a more authentic indicator of underlying emotional state and acts as an important cue for shaping social interactions (Dijk, Koenig et al. 2011). Another constraint about bodies is that they must breath. When speaking to another individual, we must periodically pause to take a breath and these pauses seem to have important social meaning. In this line, McFarland (2001) found that individuals that liked each other tended to synchronize their breathing patterns, perhaps to telegraph good points to allow the other party to take turns in the conversation. These observations don't invalidate the conduit metaphor of social interaction, but rather they emphasize that encoding and decoding processes cannot be cleanly separated from models and theories of cognition. If one accepts this claim, then virtual humans become invaluable tools for incorporating these bodily processes into cognitive science research.

Bodily constraints on cognition (behavior encoding)

With regard to behavior encoding, virtual human bodies help concretely represent and simulate the bodily constraints that govern interpersonal behavior generation. Indeed, much of the research within the field of virtual human research has explored how to create cognitive models that encode internal representations (beliefs, desires, and intentions) into external behavior, subject to the constraints that physical bodies impose. For example, imagine a virtual human is engaged in some joint task with a human observer such as negotiating an agreement (e.g., see Traum, Gratch et al. 2008). The agent has certain aspirations and preferences that are unknown to the user and similarly, the user has goals, limits and priorities that are unknown to the agent. Further, the agent might wish to strategically convey false information that might extract greater concessions (e.g., "I am uncompromising"). Because it has a complete body, how are different communication channels best utilized to achieve a good outcome? More concretely, how is communicative intent translated into speech, prosodic information, gestures and facial expressions?

Current virtual human research on behavior encoding draws heavily on traditional artificial intelligence and has much to learn from cognitive science. For example, in many virtual humans, speech and gesture generation are very loosely coupled. Many systems use standard text generation techniques to produce text and then send this to a behavior scheduler to lay down corresponding gestures (Cassell, Vilhjálmsson et al. 2001). This can be adequate for very structured social interactions (e.g., where agents can form complete sentences without interruption). However, this process is inconsistent with evidence from psycholinguists showing a closer coupling between these processes (McNeill 2005) and there is a growing sense in the virtual human community that loose modular approaches will be unable to support richer and more fluid social interactions (e.g., see Wang, Lee et al. 2011).

Virtual confederates (behavior decoding)

When it comes to behavior decoding, virtual humans also have much to offer research on social cognition. One especially promising idea is the use of virtual humans as *virtual confederates* (Bente, Kraemer et al. 2001; Bailenson, Beall et al. 2004). A major issue in studying social interaction is the problem of experimental control. In order to test a specific theory of social interaction, scientists face the problem of how to systematically manipulate social behavior in order to demonstrate causality. One common solution is to employ confederates. For example, if a scientist wants to test the theory that postural mimicry produces greater feelings of rapport, they can hire a graduate student to mimic participants in an experimental study. The criticism of such work, however is that the graduate student is obviously not blind to the experimental condition (i.e., they know if they are mimicking or not) and the act of performing this behavior might fundamentally change the nature of the social interaction. For example, to mimic a participant, the graduate student has to closely attend to the participant's nonverbal behavior and perhaps attend less to the content of their speech.

Virtual confederates hold the potential to address many of the limitations of human confederates and allow for a more systematic and controlled manipulation of bodily communicative processes. For example, in an experimental study, one can change the race, gender or gestures of a virtual human while leaving all other aspects of appearance and behavior the same. However, this approach only makes sense if people will act towards these computers programs as if they are people. This is a tall order, and some evidence suggests people treat computers very differently. For example, negotiators are more likely to accept unfair offers if they believe the offer comes from a computer program (Sanfey, Rilling et al. 2003). Interestingly, this difference diminishes if the computer program incorporates social cues such as a human-like voice or face. Indeed, there is now a substantial body of research that people respond socially to virtual humans (Nass, Robles et al. 2003; Bailenson and Yee 2005; Kraemer 2008; Pütten, Krämer et al. 2010). These effects seem to increase as the visual appearance and behavior becomes more human like (Bailenson, Yee et al. 2006), but even rudimentary characters can provoke social responses if participants are led to believe the character's behavior corresponds to the behavior of another human participant (Blascovich 2002), and indeed, a number of studies using virtual humans use this sort of deception (e.g., Bailenson and Yee 2005; Gratch, Wang et al. 2007).

Virtual confederates have several important advantages over human confederates for studying human social cognition: (a) confederates can introduce subtle but important differences in non-verbal expression when implementing the same condition across participants; (b) it is complex to keep confederates blind to their condition which might hinder the performance of genuine non-verbal expressions. Unlike virtual agents, it is known that humans have a hard time controlling or faking certain aspects of their expressions (DePaulo, Blank et al. 1992); (c) confederates are expensive; (d) it is easy to run a study online with a virtual confederate.

To concretely illustrate these suggestions, let's consider one research effort that explores the role of bodily processes in a variety of social decision making tasks such as negotiations or social dilemmas. de Melo, Gratch and Carnevale examined how the bodily expression of emotion can influence tendencies for people to be cooperative or competitive in social dilemmas. In social dilemmas, people are faced with a decision between pursuing their own self-interest or trusting another person to reach mutual cooperation and maximize joint reward. In such dilemmas, decision theorists argue that the rational thing for a person to do is act so as to maximize expected utility, which corresponds to pursuing the choice that maximizes self-interest. The dilemma is that, if all parties act "rationally", then the collective outcome is worse for all. Interestingly, people often cooperate far more than predicted by classical models and the bodily expression of emotion seems to play an important role in these departures from "rational" behavior (Frank 1988).

To examine the role of emotional signals in participant cognitions, de Melo and colleagues created an interactive virtual human that could engage in a variety of social decision tasks including social dilemmas. In a series of studies, they were able to show that virtual human emotional expressions would influence decisions in reliable ways – for example, expressions of anger would tend to lower people's willingness to cooperate (de Melo, Zheng et al. 2009). Further, they were able to show that it was not the expression *per se* that produced these effects but rather the meaning conveyed by the expression. Specifically, the identical smile behavior would produce different decisions depending on if it was elicited in the context of a cooperative or exploitive act. In the former case, participants interpreted the smile as a signal that the agent had cooperative intentions whereas the latter case (where an agent smiled while exploiting the participant), was interpreted as a signal that the agent had malicious intent (see de Melo, Gratch et al. 2011). This data emphasizes that the configuration of the body (e.g., behaviors indicating emotion) must be seen in combination with environmental events (e.g., outcomes in a joint task) to make social inferences about another's mind.

As another example, consider the Rapport Agent of Gratch and colleagues (Gratch, Wang et al. 2007; Huang, Morency et al. 2011), which was created to assess social-psychological theories of how nonverbal behavior influences the effectiveness of face-to-face interactions. According to Ticle-Degnen and Rosenthal (1990), conversations that contain certain patterns of nonverbal behavior – positivity, mutual attention and coordination – produce a sense of rapport and are associated with many desirable interpersonal outcomes such as social engagement (Tatar 1997), success in teacher-student interactions (Bernieri & Rosenthal 1988), success in negotiations (Drolet and Morris 2000), improving worker compliance (Cogger 1982), psychotherapeutic effectiveness (Tsui and Schultz 1985), improved test performance in classrooms (Fuchs 1987) and improved quality of child care (Burns 1984).

To assess this theory, the Rapport agent used virtual human technology to simulate rapport-producing nonverbal behaviors and assess the impact of these on participant's subjective and behavioral responses. Specifically, the system used a vision system to detect participant gaze, posture, gestures and expressions and an audio analysis package to analyze features of their speech. The virtual human used this information to provide real-time feedback such as nods, smiles, and postural mimicry in order to create perceptions of positivity, mutual attention and coordination. In a series of studies, these behaviors have been shown to create subjective feelings of rapport, but also behavioral outcomes predicted by rapport theory, such as more fluent speech, more intimate speech and improved test performance (see Gratch, Kang et al. in press for a review).

Summary

Virtual humans can serve as an important tool for expanding cognitive science research. Simon emphasized that computers can serve as important tools for understanding cognition through simulation. Here, I've argued that an important way to understand the mind is by simulating the body. Virtual humans provide a powerful framework for facilitating this research in a variety of ways. The exercise of connecting a cognitive model to a detailed body simulation can reveal important constraints that cognition must obey in order to act on the world or other social actors. Having a body also allows exploration of the possibility that the mind relies on bodily processes, as argued by embodied theories of cognitions. Finally, virtual humans serve as important methodological tools for studying human social processes.

Many challenges remain before exploiting the potential of this technology in cognitive science research. From the perspective of embodied cognition, the biggest challenge is that most virtual humans are mere shells of a body. They don't provide realistic biomechanical or physical simulations that can act as plausible constraints on thought. From the perspective of virtual confederates, are increasingly realistic when it comes to physical appearance but far from realistic when it comes to behavior. Indeed, this disconnect between appearance and behavior can be disturbing (Mori 2005; Groom, Nass et al. 2009). Thus, advances are needed both in enhancing behavioral realism and in finding clever experimental manipulations that work within the limits of the medium. As these limitations are addressed, virtual humans will become indispensable for any serious cognitive scientist.

References

Ahn, H.-i., A. Teeters, et al. (2007). Stoop to Conquer: Posture and Affect Interact to Influence Computer Users' Persistence

Affective Computing and Intelligent Interaction. A. Paiva, R. Prada and R. Picard, Springer Berlin / Heidelberg. **4738:** 582-593.

Badler, N. I., C. B. Phillips, et al. (1993). Simulating Humans. New York, Oxford University Press.

Bailenson, J., A. Beall, et al. (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. and N. Yee (2005). "Digital Chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments." Psychological Science **16**: 814-819.

Bailenson, J., N. Yee, et al. (2006). "The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition and copresence in dyadic interaction." Presence: Teleoperators and Virtual Environments **15**(4): 359-372.

Bailenson, J. N., K. R. Swinth, et al. (2005). "The independent and interactive effects of embodied agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in Immersive Virtual Environments." PRESENCE: Teleoperators and Virtual Environments **14**: 379-393.

Barsalou, L. W. (2008). "Grounded Cognition." Annual Review of Psychology **59**(1): 617-645.

Bartlett, M. S., G. C. Littlewort, et al. (2006). Fully automatic facial action recognition in spontaneous behavior. 7th International Conference on Automatic Face and Gesture Recognition, Southampton, UK.

Bavelas, J. B., L. Coates, et al. (2000). "Listeners as Co-narrators." Journal of Personality and Social Psychology **79**(6): 941-952.

Bente, G., N. C. Kraemer, et al. (2001). "Computer Animated Movement and Person Perception: Methodological Advances in Nonverbal Behavior Research." Journal of Nonverbal Behavior **25**(3): 151-166.

Berkeley, I. S. N. (2008). CajunBot: A Case Study of an Embodied System. andbook of Embodied Cognitive Science. P. C. Garzon and A. Gomila. London, Elsevier**:** 81-98.

Bickmore, T., A. Gruber, et al. (2005). "Establishing the computer-patient working alliance in automated health behavior change interventions." Patient Education and Counseling **59**(1): 21-30.

Bickmore, T. and L. Pfeifer (2008). Relational Agents for Antipsychotic Medication Adherence. CHI'08 workshop on Technology in Mental Health. Florence, Italy.

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

Bringsjord, S. and B. Schimanski (2003). What is artificial intelligence? psychometric AI as an answer. Proceedings of the 18th international joint conference on Artificial intelligence. Acapulco, Mexico, Morgan Kaufmann Publishers Inc.**:** 887-893.

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

Cassell, J. and M. Stone (1999). Living Hand to Mouth: Psychological Theories about Speech and Gesture in Interactive Dialogue Systems. AAAI Fall Symposium on Narrative Intelligence, AAAI Press.

Cassell, J., H. Vilhjálmsson, et al. (2001). BEAT: The Behavior Expressive Animation Toolkit. SIGGRAPH, Los Angeles, CA, ACM Press.

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

Collins, S. H., M. Wisse, et al. (2001). "A Three-Dimensional Passive-Dynamic Walking Robot with Two Legs and Knees." The International Journal of Robotics Research **20**(7): 607-615.

de Melo, C., J. Gratch, et al. (2011). Reverse Appraisal: Inferring from Emotion Displays who is the Cooperator and the Competitor in a Social Dilemma. Cognitive Science Conference. Boston.

de Melo, C., L. Zheng, et al. (2009). Expression of Moral Emotions in Cooperating Agents. 9th International Conference on Intelligent Virtual Agents. Amsterdam, Springer.

Dehghani, M., J. Gratch, et al. (2012). Interpersonal effects of emotions in morally-charged negotiations. 4th Annual Meeting of the Cognitive Science Society, Sapporo, Japan.

DePaulo, B. M., A. L. Blank, et al. (1992). "Expressiveness and Expressive Control." Personality and Social Psychology Bulletin **18**(3): 276-285.

Dijk, C., B. Koenig, et al. (2011). "Saved by the blush: Being trusted despite defecting." Emotion **11**(2): 313-319.

Drolet, A. L. and M. W. Morris (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.

Frank, R. (1988). Passions with reason: the strategic role of the emotions. New York, NY, W. W. Norton.

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. (2012). "The Sciences of the Artificial Emotions: Comment on Aylett and Paiva." Emotion Review **4**(3): 266-268.

Gratch, J., S.-H. Kang, et al. (in press). Using social agents to explore theories of rapport and emotional resonance. Social emotions in nature and artifact: emotions in human and human-computer interaction. J. Gratch and S. Marsella. Cambridge, MA, Oxford University Press.

Gratch, J. and S. Marsella (2005). "Lessons from Emotion Psychology for the Design of Lifelike Characters." Applied Artificial Intelligence **19**(3-4): 215-233.

Gratch, J., N. Wang, et al. (2007). Creating Rapport with Virtual Agents. 7th International Conference on Intelligent Virtual Agents. Paris, France.

Groom, V., C. Nass, et al. (2009). "Evaluating the effects of behavioral realism in embodied agents." International Journal of Human-Computer Studies **67**: 842-849.

Hart, J., J. Gratch, et al. (in press). How virtual reality training can win friends and influence people. Fundamental Issues in Defense. R. Sottilare. Surrey, UK, Ashgate Press.

Honglun, H., S. Shouqian, et al. (2007). "Research on virtual human in ergonomic simulation." Computers & Industrial Engineering **53**(2): 350-356.

Huang, L., L.-P. Morency, et al. (2011). Virtual rapport 2.0. Proceedings of the 10th international conference on Intelligent virtual agents. Reykjavik, Iceland, Springer-Verlag**:** 68-79.

Iverson, J. M. and E. Thelen (1999). "Hand, mouth and brain. The dynamic emergence of speech and gesture." Journal of Consciousness Studies **6**(11-12): 19-40.

Johnsen, K., A. Raij, et al. (2007). The validity of a virtual human experience for interpersonal skills education. Proceedings of the SIGCHI conference on Human factors in computing systems. San Jose, California, USA, ACM**:** 1049-1058.

Kahneman, D., J. L. Knetsch, et al. (1991). "Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias." The Journal of Economic Perspectives **5**(1): 193-206.

Kraemer, N. (2008). Social Effects of Virtual Assistants. A Review of Empirical Results with Regard to Communication. Proceedings of the 8th international conference on Intelligent Virtual Agents. Tokyo, Japan, Springer-Verlag.

Laird, J. E. (2001). It Knows What You're Going To Do: Adding Anticipation to a Quakebot. Proceedings of the Fifth International Conference on Autonomous Agents, Montreal, Canada, ACM Press.

Maciel, A., L. P. Nedel, et al. (2002). Anatomy Based Joint Models for
Virtual Humans Skeletons. IEEE COMPUTER ANIMATION**:** 220-224.

McFarland, D. H. (2001). "Respiratory Markers of Conversational Interaction." Journal of Speech, Language, and Hearing Research **44**: 128-143.

McNeill, D. (2005). Gesture and Thought, University of Chicago press.

Meyer, D. E. and D. E. Kieras (1997). "A computational theory of executive cognitive processes and multiple-task performance: Part 2. Accounts of psychological refractory-period phenomena." Psychological Review **104**(4): 749-791.

Mohler, B. J., H. H. Bülthoff, et al. (2008). A full-body avatar improves egocentric distance judgments in an immersive virtual environment. Proceedings of the 5th symposium on Applied perception in graphics and visualization. Los Angeles, California, ACM**:** 194.

Morency, L.-P., C. Sidner, et al. (2005). Contextual Recognition of Head Gestures. 7th International Conference on Multimodal Interactions, Torento, Italy.

Mori, M. (2005). On the Uncanny Valley. Proceedings of the Humanoids-2005 workshop: Views of the Uncanny Valley. Tsukuba, Japan.

Nass, C., E. Robles, et al. (2003). "Speech-Based Disclosure Systems: Effects of Modality, Gender of Prompt, and Gender of User." International Journal of Speech Technology **6**(2): 113-121.

Niedenthal, P. M., M. Mermillod, et al. (2010). "The simulation of smiles (SIMS) mode: embodied simulation and the meaning of facial expression." Behavioral and brain sciences **33**: 417-480.

Parkinson, B. (2009). "What holds emotions together? Meaning and response coordination." Cognitive Systems Research **10**: 31-47.

Parsons, T. D., P. Kenny, et al. (2008). "Objective Structured Clinical Interview Training using a Virtual Human Patient." Studies in Health Technology and Informatics **132**: 357-362.

Petty, R. E., G. L. Wells, et al. (1983). "The Effects of Recipient Posture on Persuasion." Personality and Social Psychology Bulletin **9**(2): 209-222.

Pezzulo, G., L. W. Barsalou, et al. (2011). "The Mechanics of Embodiment: A Dialog on Embodiment and Computational Modeling." Frontiers in Psychology **2**(5): 1-21.

Pfeifer, R. and J. C. Bongard (2007). How the Body Shapes the Way We Think-A New View of Intelligence. Cambridge, MA, MIT Press.

Pütten, A. M. v. d., N. C. Krämer, et al. (2010). ""It doesn't matter what you are!" Explaining social effects of agents and avatars." Computers in Human Behavior **26**(6): 1641-1650.

Reddy, M. (1979). The conduit metaphor - a case of frame conflict in
our language about language. Metaphor and
thought. A. Ortony. Cambridge, UK, Cambridge University Press**:** 284-297.

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

Ritter, F. E. and R. M. Young (2001). "Embodied models as simulated users: introduction to this special issue on using cognitive models to improve interface design." International Journal of Human-Computer Studies **55**(1): 1-14.

Rosenbloom, P. (2011). Bridging Dichotomies in Cognitive Architectures for Virtual Humans. AAAI Fall Symposium on Advances in Cognitive Systems (FS-11-01). Washington D.C., AAAI Press**:** 265-272.

Sanfey, A. G., J. K. Rilling, et al. (2003). "The Neural Basis of Economic Decision-Making in the Ultimatum Game." Science **300**(5626): 1755-1758.

Schachter, S. and J. E. Singer (1962). "Cognitive, social, and physiological determinants of emotional state." Psychological Review **69**: 3790399.

Schnall, S., J. Haidt, et al. (2008). "Disgust as Embodied Moral Judgment." Personality and Social Psychology Bulletin **34**(8): 1096-1109.

Schuller, B., A. Lehmann, et al. (2009). Blind Enhancement of the Rhythmic and Harmonic Sections by NMF: Does it help? NAG/DAGA. Rotterdam.

Shannon, C. E. and W. Weaver (1948). A mathematical theory of communication. Urbana, IL, University of Illinois Press.

Simon, H. (1969). The sciences of the Artificial. Cam bridge, MA, MIT Press.

Sprague, N., D. Ballard, et al. (2007). "Modeling embodied visual behaviors." ACM Trans. Appl. Percept. **4**(2): 11.

Suchman, L. A. (1987). Plans and situated actions: The problem of human-machine communication. New York, Cambridge University Press.

Swartout, W., J. Gratch, et al. (2006). "Toward Virtual Humans." AI Magazine **27**(1).

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

Thiebaux, M., A. Marshall, et al. (2008). SmartBody: Behavior Realization for Embodied Conversational Agents. International Conference on Autonomous Agents and Multi-Agent Systems. Portugal.

Tickle-Degnen, L. and R. Rosenthal (1990). "The Nature of Rapport and its Nonverbal Correlates." Psychological Inquiry **1**(4): 285-293.

Traum, D. (2008). Talking to Virtual Humans: Dialogue Models and Methodologies for Embodied Conversational Agents. Modeling Communication with Robots and Virtual Humans. I. Wachsmuth and G. Knoblich. Berlin, Springer**:** 296-309.

Traum, D., J. Gratch, et al. (2008). Multi-party, Multi-issue, Multi-strategy Negotiation for Multi-modal Virtual Agents. 8th International Conference on Intelligent Virtual Agents. Tokyo, Japan, Springer.

Tsai, J., E. Bowring, et al. (2011). Empirical Evaluation of Computational Emotional Contagion Models

Intelligent Virtual Agents. H. Vilhjálmsson, S. Kopp, S. Marsella and K. Thórisson, Springer Berlin / Heidelberg. **6895:** 384-397.

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

Wang, Z., J. Lee, et al. (2011). Towards More Comprehensive Listening Behavior: Beyond the Bobble Head

Intelligent Virtual Agents. H. Vilhjálmsson, S. Kopp, S. Marsella and K. Thórisson, Springer Berlin / Heidelberg. **6895:** 216-227.

Wilson, M. (2002). "Six views of embodied cognition." Psychonomic Bulletin & Review **9**(4): 625-636.

Yngve, V. H. (1970). On getting a word in edgewise. Sixth regional Meeting of the Chicago Linguistic Society.

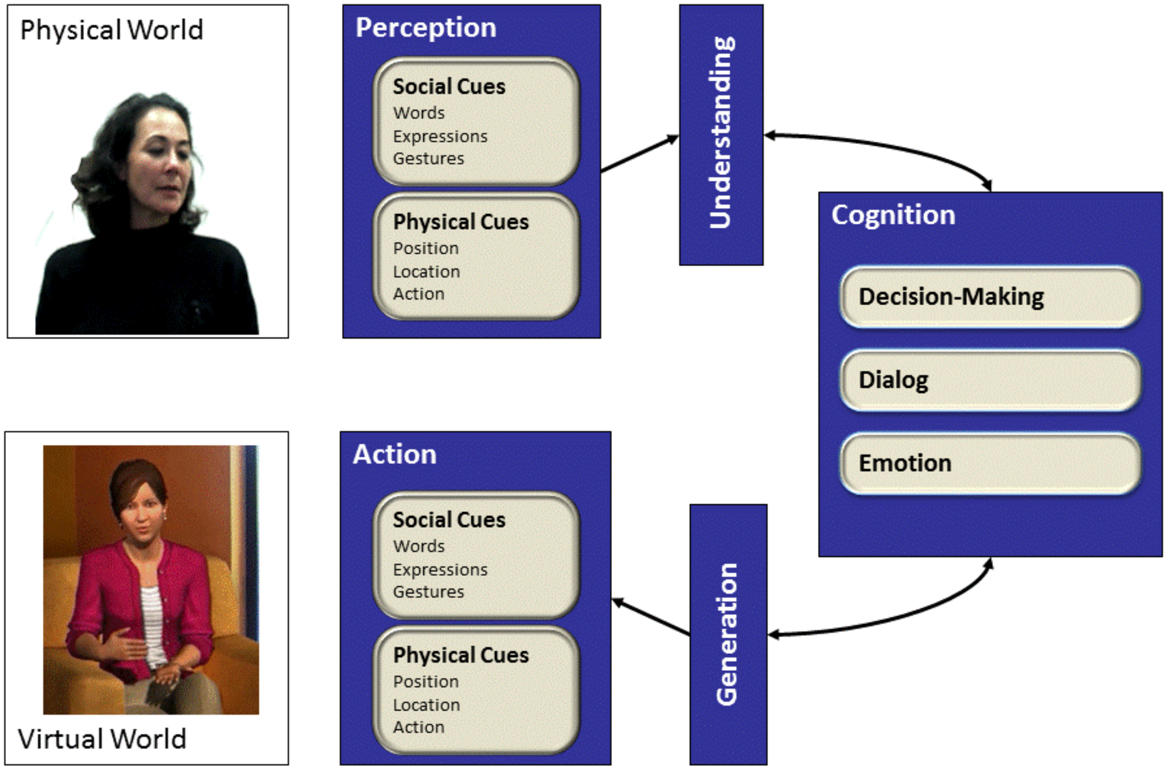

Figure 1: The component tasks that must be performed by the virtual humans discussed in this proposal. Such agents must recognize and understand physical and social signals in the real world, simulate human perceptual, cognitive and emotional capabilities, and act realistically in the virtual environment in ways that support training.

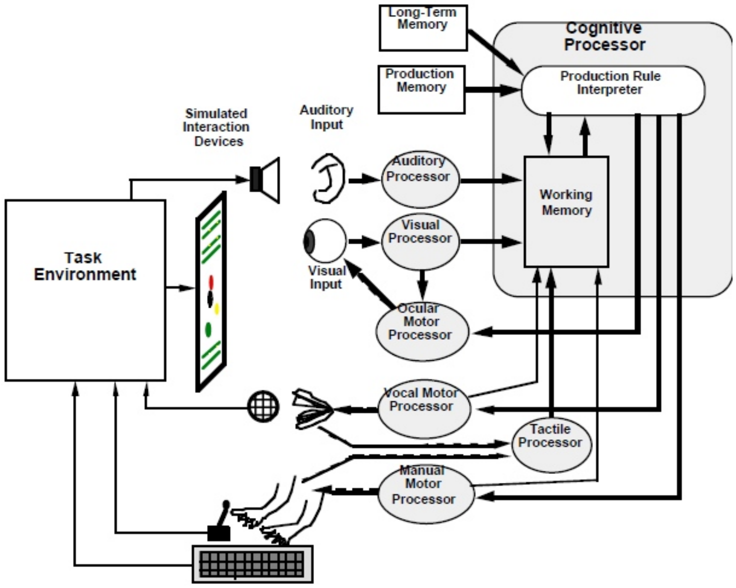

Figure 2: The EPIC cognitive architecture explores perceptual-motor linkages to cognition.

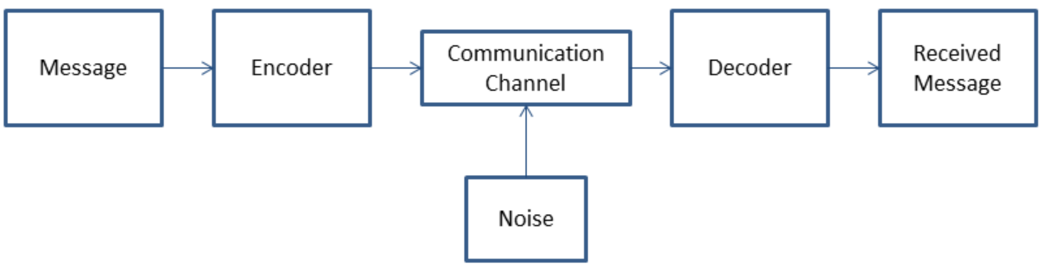

Figure 3: The conduit model of communication inspired was inspired by Shannon and Weaver's (1948) mathematical model of communication.