Avatars

Jeremy N. Bailenson

Department of Communication

Stanford University

Jim Blascovich

Research Center for Virtual Environments and Behavior

University of California, Santa Barbara

Introduction

Webster's dictionary defines an *avatar* as "the incarnation of a Hindu deity". According to the Hindu religion, a god needs some type of a representational vehicle to embody his holy being when interacting with humans. The deity appears to humans via an avatar of either human or animal form.

In the late 20th century, this term avatar was adopted by popular culture and scientists studying human-computer interaction as a definition for a digital representation of humans in some type of online or virtual forum (i.e., virtual reality). In other words, the human needs some type of representational vehicle to embody his or her being in virtual reality. Neal Stephenson, in his seminal science fiction novel Snow Crash (1993), is largely credited with applying this religious term towards virtual reality, though the reference appeared as early as 1984 in online Multi-User Dungeons (Morningstar & Farmer, 1991) and the concept, though not the term, appeared in earlier works of fiction (e.g., Gibson, 1984; Vinge, 1981; Brunner, 1975).

This chapter will explore avatars: their underlying concepts, their current use in society, the scientific research exploring them, and the ethical issues behind digital human representation.

Avatars: The Underlying Concepts

Definitions

In order to properly discuss avatars, it is necessary to define a number of related terms. An avatar is a perceivable digital representation[1] whose behaviors are executed in real-time by a human being. An agent is some type of mathematical or computational

formula designed to accomplish a specific goal. An embodied agent is an agent whose behaviors are executed by some type of perceivable digital representation. In other words, humans control avatar behavior, and agents control embodied agent behavior. Figure 1 demonstrates the relations among these terms.

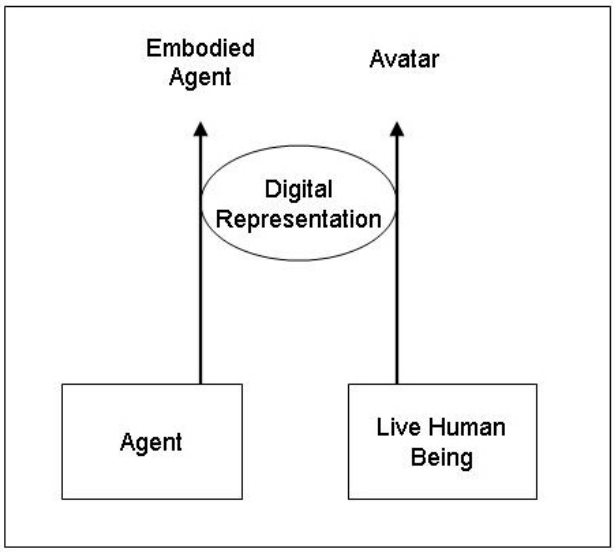

Figure 1: A representational schematic of avatars and embodied agents. When a given digital representation is controlled by a human it is an avatar, and when it is controlled by a computational agent it is an embodied agent. Central to the current definition is the ability for real-time behavior, in that the digital representation exhibits behaviors by the agent or human as they are performed.

Figure 1 illustrates the fact that the specific digital form used to represent the person or the agent is inconsequential to the defining status of the representation. An agent or person can drive the same digital representation, and the specific details of how that representation looks or sounds is inconsequential to determining its status. A human can represent herself with an avatar that looks like a computer, and an agent can embody itself with an extremely realistic humanoid digital representation.

However, the distinction between agents and avatars often blurs. Relatively few true avatars appear in contemporary virtual environment implementations. For real-time representations of human behavior (e.g., collaborative virtual environments used for

conferencing or online video games and chat-rooms) it is extremely difficult, given today's technology, to accurately provide a one-to-one mapping of a human's actions (e.g., movements, sounds, touches, scents) and her or his avatar's actions. As a consequence, some behaviors are sometimes rendered nonveridically; for example, a human presses a button to make his avatar in a chatroom smile instead of having equipment sense and render every small detail of his or her facial movements. As a result, the behaviors are slightly changed; the smile rendered by the computer algorithm may be drastically different from any given human's smile. Other behaviors (e.g., reflexive gestures such as blinking or breathing) are often either omitted or alternatively rendered onto the digital representation via a computer algorithm. In other words, an avatar can be and often is a hybrid of an embodied agent and an avatar: the human controls the symbolically meaningful verbal and nonverbal gestures and an agent controls the more mundane automatic behaviors. As Supreme Court Justice Potter Stewart wrote in 1964, "I can't define pornography, but I know it when I see it."—similarly, it may not be possible to mathematically delineate the defining features that separate avatars from embodied agents, but there is little doubt that a meaningful distinction exists.

It also may be useful to distinguish an avatar from an online identity (Turkle, 1995). An online identity captures the notion of the distributed representation of people in the digital age. In other words, people today are known to each other by email and home pages on the World Wide Web, and as a consequence, each of us has an 'online identity', constituted by the distributed representation of all information relevant to us. In this chapter, however, we refer the very specific definition depicted in Figure 1 as an avatar.

<u>How real are avatars?</u>

Avatars look and behave like the people they represent. But exactly how realistic do avatars need to be in order to be effective, or even to be considered an avatar? Avatars can resemble their human counterparts along a number of dimensions (See Blascovich, 2001, for a review); the two which have received the most attention in the literature is behavioral resemblance (number of a given human's behaviors the avatar exhibits) and photographic resemblance (how many of a given human's static visual features the avatar possesses).

The degree of behavioral resemblance is largely governed by the ability of the system to track behavior (know exactly what the human is doing at every given moment), and then to render behavior (transpose that exact behavior onto the digital representation). Currently, real-time behavioral tracking technology for avatars, while improving steadily, is extremely far from matching expectations instilled by popular culture, for example, online representations of Neo from the Matrix (Warner Brothers, 1999), Hiro from Snow Crash (Stephenson, 1992), or Case from Neuromancer (Gibson, 1984). In those fictional accounts, the movements and gestures of avatars are seamless; when Case moves his arm while jacked into the computer deck, his avatar performs the exact motion at the exact same time in cyberspace, and the actions of human and avatar are perceptually indistinguishable. Outside the fictional realm, however, real-time behavior tracking is extremely difficult. While there have been advances in tracking of gesture through various forms mechanical, optical, and other systems (see Turk & Kolsch, 2003, for a review), the gap between actual movements and real-time tracked movements remains large.

Furthermore, once the movements and behaviors have been captured, they need to be rendered onto the digital representation of the avatar. This process is not trivial, and many issues arise in terms of the quality of the movements and behaviors when applied to a digital model that contains only a fraction of the degrees of motion freedom that the human body does. In other words, currently, digital models simply don't have enough landmarks, joints, and meaningful clusters of polygons to support completely naturalistic and realistic movements. While the advances in motion capture rendering and inverse kinematics have been quite good with non-real-time representations (see any film by Pixar), the same is not true with real-time avatars in which these motions need to be expressed on the fly.

On the other hand, there are fewer barriers to achieving high photographic resemblance. The use of three-dimensional scanners, as well as photogrammetric software allows for the realistic recreation of static, digital human heads and faces that are nearly real enough to function as an analog to a real face (Bailenson, Beall, Blascovich, & Rex, 2003). The key challenge, though, is designing faces and bodies in high enough detail to allow for the realistic rendering of behavior described above. In sum, static avatars currently can look quite a bit like their human controllers; however, avatars can only perform a small subset of a dynamic human's actions in real-time.

Avatars: Current Use

Avatars are being used in a number of venues currently. Depending on how loosely one defines 'a digital representation', the argument can be made that avatars are quite pervasive in society. For example, sound is transformed into digital information as it travels over fiber-optic cables and cellular networks; consequently, the audio

representation we perceive over phone lines is actually an avatar of the speaker. This classification may seem trivial at first, but becomes less trivial when preset algorithms are applied to the audio stream to cause subtle changes in the avatar, such as cleaning and amplifying the signal. In other words, because the voice is translated into digital information, it is a representation of the human, as opposed to raw perceptual input from the speaker.

In practice, scholars and online communities tend to refer most often to visual representations when using the term avatar. Currently, millions of people employ avatars in Massively Multi-User Online Role-Playing Games (MMORPGs; Yee, 2003) as well as in some three-dimensional chat-rooms used for virtual conferencing. In these communities, users interact with one-another using either a keyboard or a joystick, and type messages to one-another as well as see each others' avatars move around digital space. However, these are avatars in the most minimal sense of the word; behavioral and photographic similarity of these online is quite low due to the inaccessibility and expense of tracking and digital modeling software. Instead, typically users navigate these worlds using stock avatars with limited behavioral capabilities.

Currently, there are few arenas in which extremely realistic avatars are common. In the following section we discuss the area in which avatars are most common: scientific research.

Avatars: Scientific Research

Computer Scientists and other researchers are dedicating great efforts towards developing systems capable of producing functional and effective avatars (e.g., Cassell & Vilhjálmsson, 1999; Thalmann & Thalman, 1999; Benford, Bowers, Fahlén, Greenhalgh,

& Snowdon; Badler, Phillips, & Webber, 1993). These researchers strive to develop the graphics, logic, and the tracking capabilities to a) veridically render actual movements by humans on digital avatars, and b) augment those veridical movements by employing control algorithms to 'take over' when there is missing tracking data or information about static visual features.

Furthermore, behavioral scientists are examining humans as they interact with one-another via avatars (Beall, Bailenson, Loomis, Blascovich, & Turk, 2003; Slater, Howell, Steed, Pertaub, & Garau, 2000; Slater, Sadagic, Usoh, & Schroeder, 2000). These researchers strive to understand social presence (also known as copresence). Social presence reflects the degree to which people respond socially towards an avatar during interaction, compared to the degree to which they respond to normal physical humans.

Blascovich and colleagues (Blascovich et al. 2002; Blascovich, 2001) provide a threshold model for social presence that provides specific predictions for the interplay of photographic and behavioral realism of avatars. According to that model, photographic anthropomorphism, that is, how realistic an avatar looks or how much the avatar was designed to look like an actual human, is only an important driving factor when it relates specifically to behavioral realism. In other words, inclusion of certain visual features are necessary to perform important, socially relevant behavioral actions (e.g., an avatar needs to have recognizable eyebrows in order to frown). In sum, according to the Blascovich model, photographic realism is only important in that it allows for social behaviors.

There is ample data relating to this prediction, some of which supports the Blascovich model, other of which does not. Bailenson, Blascovich, Beall, and Loomis (2001) demonstrated that making a digital representation more realistic looking (i.e.,

adding a texture map from an actual photograph of a face) does not increase copresence over an agent that is more cartoon-like, so long as both types of agents demonstrate realistic gaze behaviors. Along similar lines, Garau et al. (2003) did not demonstrate an overall advantage for more photographically realistic avatars; moreover, those researchers demonstrated that increasing photographic realism of an avatar can actually cause a decrease in copresence, if the behavioral realism is not also increased.

In sum, research on avatars is largely in its infancy; this early work is paving the way for our understanding of computer-mediated human interaction. As avatars become more commonplace in communication systems, collaborative learning environments, and entertainment systems, research geared towards understanding these applications will become more prevalent.

Avatars: Ethical Implications

Interacting via avatars allows for a phenomenon that is simultaneously a luxury and a danger: the removal of objective truth from interaction. Bailenson, Beall, Loomis, Beall, & Turk (2003) describe a paradigm called Transformed Social Interaction (TSI). Using an avatar to interact with another person is qualitatively different from other forms of communication, including face-to-face interaction, telephone conversations, and videoconferencing. Via an avatar (which is constantly redrawn in real-time), interactants possess the ability to systematically filter their physical appearance and behavioral actions in the eyes of their conversational partners, amplifying or suppressing features and nonverbal signals in real-time for strategic purposes. In other words, interactants can use avatars to bend, twist, or entirely remove truth from a conversation.

These TSI algorithms can have a drastic impact on interactants' abilities to influence their conversational partners. For example, online teachers lecturing to more than one student at once can individually tailor their nonverbal behaviors towards each of the students simultaneously. Student A might respond well to a teacher who smiles, and Student B might respond well to a teacher with a neural expression; using an avatar that is rendered separately for each student, a teacher can project, simultaneously, an avatar to both Student A and B designed to work best for each individual student. Research by Beall et al. (2003) has used avatars to employ such a strategy using eye-contact, and has demonstrated an increased amount of student attention to the teacher using TSI.

While there are many applications in which avatars augmented with TSI are advantageous, there are also some ethical problems that arise when interacting within a medium in which there is no objective reality. In the hands of advertisers, unscrupulous politicians, lawyers, or any person with the goal of influencing interactants, one can paint a potentially dismal picture of the future of interaction, one in which nobody is who they seem to be, and online human representations are stretched so far away from their actual physical essences that the act of conversation becomes merely a shell of its former self. Nonetheless, the potential for subjective identities is not a phenomenon intrinsic to avatars in virtual reality. Currently, people wear makeup and costumes, get plastic surgery, wear braces, and engage in a host of practices geared towards changing their appearance as well as their behaviors. With avatars, however, this practice only becomes easier. Early research (Bailenson et al., 2003; Beall et al., 2003) has demonstrated that TSI of avatars is often difficult to detect. However, it is the challenge of researchers to

determine the best way to manage this potential for TSI as the use of avatars becomes more prevalent.

Conclusions

In the current chapter, we have reviewed the concept of the avatar in human-computer interaction. Currently, there are ample examples of humans interacting with one-another via avatars. For the most part, these avatars are simplistic and unrealistic. The exception to this trend occurs in research laboratories, in which scientists are beginning to develop and test avatars that are extremely similar in appearance and behavior to their human counterpart. As these avatars become more prevalent in society, there is a realistic potential for qualitative changes in the manner in which we communicate due to the decoupling of behavior from human to avatar. While there are some potential pitfalls to transforming behaviors as they pass from physical actions to digital representations, doing so presents many opportunities for both the user of online systems as well as for researchers of human-computer interaction.

References

Badler, N., Phillips, C., Webber, B. (1993). *Simulating Humans: Computer Graphics, Animation, and Control*. Oxford University Press, 1993.

Bailenson, J.N., Beall, A.C., Blascovich, J., & Rex, C. (2003, in press). Examining Virtual Busts: Are Photogrammetrically-Generated Head Models Effective for Person Identification? *PRESENCE: Teleoperators and Virtual Environments*.

Bailenson, J.N., Beall, A.C., Loomis, J., Blascovich, J., & Turk, M. (2003, in press). Transformed Social Interaction: Decoupling Representation from Behavior and Form in Collaborative Virtual Environments. *PRESENCE: Teleoperators and Virtual Environments*.

Bailenson, J.N., Blascovich, J., Beall, A.C., & Loomis, J.M.. (2001). Equilibrium revisited: Mutual gaze and personal space in virtual environments. *PRESENCE: Teleoperators and Virtual Environments*, 10, (2001) 583-598.

Beall, A. C., Bailenson, J. N., Loomis, J., Blascovich, J., & Rex, C. (2003). Non-zero-sum mutual gaze in immersive virtual environments. *Proceedings of HCI International 2003*, Crete.

Benford, S. D., Bowers, J. M., Fahlén, L. E., Greenhalgh, C. M. and Snowdon, D. N. User Embodiment in Collaborative Virtual Environments, in Proc. *1995 ACM Conference on Human Factors in Computing Systems (CHI'95)*, Denver, Colorado, USA, May 1995, ACM Press.

Blascovich, J. (2001). Social influences within immersive virtual environments. In R. Schroeder (Ed.) *The social life of avatars*. Springer-Verlag.

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2001, in press). Immersive Virtual Environment Technology as a Methodological Tool for Social Psychology . *Psychological Inquiry, 13*, 146-149.

Brunner, J. (1975). *Shockwaver Rider*. Ballantine Books.

Cassell, J. and Vilhjálmsson, H. (1999). Fully Embodied Conversational Avatars: Making Communicative Behaviors Autonomous. *Autonomous Agents and Multi-Agent Systems 2(1)*: 45-64.

Garau, M., Slater, M., Vinayagamoorhty, V., Brogni, A., Steed, A., & Sasse, M. A. (2003). The Impact of Avatar Realism and Eye Gaze Control on Perceived Quality of Communication in a Shared Immersive Virtual Environment. *Proceedings of the SIG-CHI conference on Human factors in computing systems*, April 5-10, 2003, Fort Lauderdale, FL, USA.

Gibson, W. (1984). *Neuromancer*. New York : Ace Books.

Morningstar, C. & Farmer, (1991). The Lessons of Lucasfilm's Habitat, in *Cyberspace: First Steps*, Michael Benedikt (ed.), MIT Press, Cambridge, Mass.

Slater, M., Howell, J., Steed, A., Pertaub, D., Garau, M. & Springel, S. (2000) Acting in Virtual Reality, *ACM Collaborative Virtual Environments, CVE'2000*, 103-110.

Slater, M., Sadagic, A., Usoh, M., & Schroeder, R. (2000) Small Group Behaviour in a Virtual and Real Environment: A Comparative Study *Presence: Teleoperators and Virtual Environments 9(1)*, 37-51.

Stephenson, N. (1993). Snow crash. New York: Bantam Books.

Thalmann, M. N, & Thalmann D. (eds), *Computer Animation and Simulation 99*, Springer Verlag, Wien, 1999.

Turk, M. & Kolsch, M. (2003). Perceptual Interfaces. In Medioni, G. & Kang, S. B. (eds.) *Emerging Topics in Computer Vision*, Prentice Hall (in press).

Turkle, S. (1995). *Life on the Screen: Identity in the Age of the Internet.* New York: Simon and Schuster.

Vinge, V. (1981). *True Names*, in Binary Star #5, edited by James R. Frenkel. Dell, New York.

Yee, N. (2002). Befriending Ogres and Wood Elves - Understanding Relationship Formation in MMORPGs. *Http://www.nickyee.com/hub/relationships/home.html.*

Footnotes

1. We restrict the discussion in the current paper to digital avatars, excluding physical avatars such as puppets and robots. There are a number of reasons for focusing on digital representations, such as the ability to quickly send and replicate digital information and the disparity of advances in virtual human digital and physical animation technology. Currently, the majority of digital avatars are visual or auditory information though there is no reason to restrict the definition as such.