Presence and The Sixth Sense

Abstract

This paper discusses the notion that presence may be considered as a selection mechanism that organizes the stream of sensory data into an environmental gestalt or perceptual hypothesis about current environment. A particular environmental gestalt results in scan-sensing of the world in a particular pattern reminiscent of saccades and fixations in eye scan paths. The environment hypothesis is continually reverified or else a break in presence occurs. Presence is therefore compared to visual hypothesis selection in the work of Richard Gregory and Lawrence Stark. The implications for measurement are discussed, and it is concluded that physiological measures indicating breaks in presence are worthy of study, and that the study of presence is also the study of what maintains an environmental gestalt.

1 Introduction

Presence is often thought of as the sense of 'being there' in the virtual environment. Draper, Kabur, and Usher (1998) and IJsselsteijn, Freeman, and de Ridder (2001) provide reviews of this area. In this paper, we outline another way of thinking about presence, one that does not rely on a characterization of subjective experience. Presence is considered as a perceptual mechanism for selection between alternative hypotheses: "I am in this place" and "I am in that place" ("I am confused"). This idea is based on the approaches to perception of Richard Gregory (1998) and Lawrence Stark et al. (2001). This view of presence was first discussed by Slater and Steed (2000). In this paper, we further reflect on this idea and its consequences for measurement.

Presence, *Vol. 11, No. 4, August 2002,* 435–439

2 Presence and Hypothesis Selection

An important current in perceptual theory, exemplified by the work of Richard Gregory, is that perception involves the act of selection between hypotheses. This is illustrated by the famous Kanisza triangle (figure 1).

It is far more likely that a white triangle is overlaying the three black circles rather than the (physically correct) hypothesis that three circles have had parts cut out of them and then arranged perfectly to create a triangular shape. The hypothesis is typically accepted so strongly that triangle edges between the circles are usually seen, although there are no such edges, of course.

The scan path theory of Lawrence Stark postulates a top-down cognitive model of visual perception. A visual hypothesis results in a repeated pattern of eye movements consisting of alternate saccades and fixations. There is a way to look at the world when the hypothesis about what we are seeing is one thing rather than another. The recurring fixations are as if the observer is repeatedly verifying the hypothesis: this is a kitchen rather than a bedroom, a café rather than a bathroom, an indoor experience rather than an outdoor one, a duck rather than a rabbit. (See figure 2.) Notice how your pattern of eye movements changes according to how you decide to see this figure. There is a way to look at a duck that is different from how to look at a rabbit. Of course, the physical shading that defines the figure is invariant.

A top-down cognitive mechanism in conjunction with the "stuff" of the image determines your eye movements, rather than eye movements and perception being determined by what is out there. In other words, we see

Mel Slater
m.slater@cs.ucl.ac.uk
Department of Computer Science
University College London

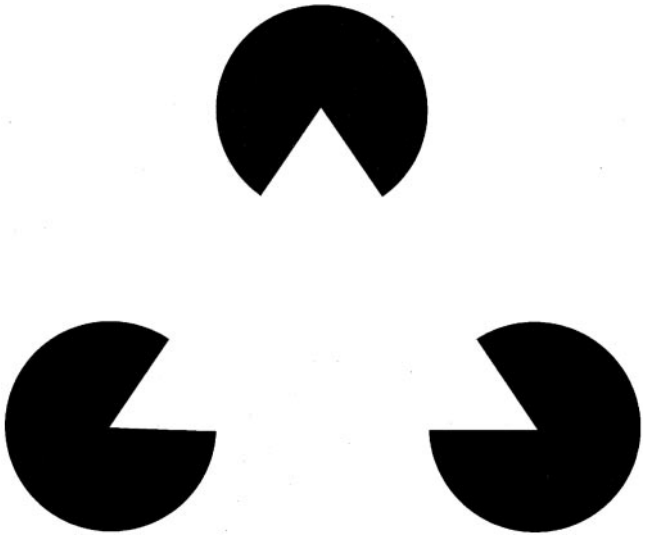

Figure 1. *The Kanisza Triangle.*

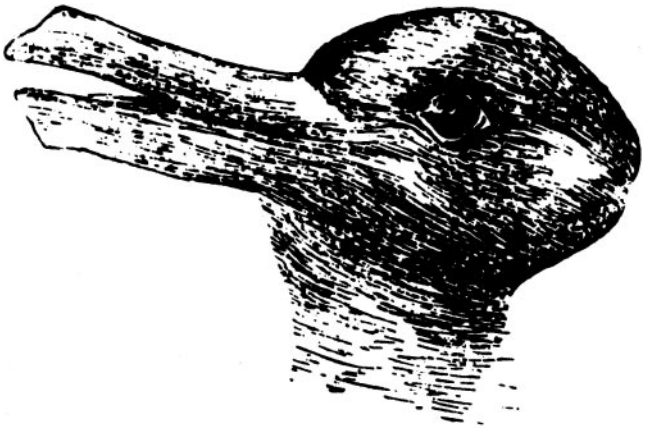

Figure 2. *A duck or a rabbit?*

what is already in our mind's eye. In the movie *The Sixth Sense* (directed by M. Night Shyamalan), dead people were said to "see what they want to see." It seems that the same may be true for live people.

At any time, but especially when we are immersed in a VE, we may be receiving competing signals from several different environments. The issue of presence is interesting only when there are competing signals from at least two environments. This is fundamental to the concept of presence in this paper. If (somehow) an actor were receiving signals from only one environment, then by definition that actor is present in that environment. Of course, there are all sorts of other things that may be studied: the degree to which the actor is interested in, paying attention to, involved in, that environment. But these are not presence.

The critical issue is how will the actor act? To which set of signals will the actor respond?

In the pit room (figure 3) (Usoh et al., 1999), the participant walks into a virtual room with an 18 m precipice at its center. In fact, the person is standing in a CAVE-like system or wearing a head-mounted display, feeling the cables, temperature, and the physical signals of that real place. But visual signals indicate that they are standing in a room with a dangerous precipice. Presence in the virtual room at any moment results in choice of the hypothesis that indeed this is a room with a precipice rather than the physical place of a CAVE. Of course, the participant has abstract knowledge that "really" they are in a CAVE, but visual perception overrides this knowledge and the bodily system reacts as if they were in the pit room: heart rate rises, locomotion is carefully judged, and the subject reports symptoms of anxiety (Meehan, Insko, Whitton, & Brooks, 2001).

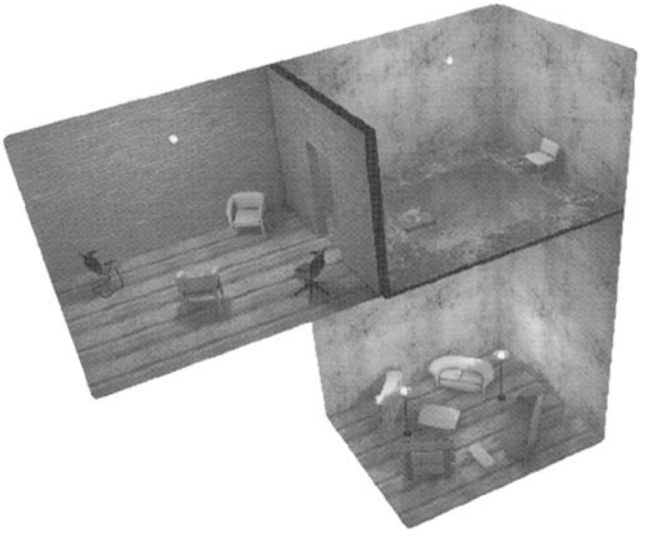

Figure 3. *The pit room (UNC, UCL experiment; © ACM 1999).*

Given competing signals from different environments, at any moment action is chosen on the basis of selection between alternative hypotheses, alternative interpretations. Some signals (such as the weight of the cables) will be in the background in one interpretation ("I am

in the pit room"), and come to the foreground ("I am in a CAVE") in another. How the participant scan-senses the environment will depend upon which interpretation is dominant. The visual signals of the pit room may be discounted and defocused when the interpretation "I am really in a CAVE" becomes the dominant reality. Meehan added static haptics representing the edge of the precipice, and Basdogan, Ho, Srinivasan, and Slater (2000) added force feedback to a shared virtual manipulation task, both resulting in increased reported presence. Perhaps the greater the number of consistent sensory signals belonging to an environment, the greater the probability that the interpretation corresponding to that environment will be selected.

The presence selection mechanism is in answer to a fundamental question: Where am I? A way to think of this is that the totality of signals may be organized into alternative gestalts, with each one corresponding to a different interpretation and different set of actions (Slater & Steed, 2000).

Actions may be of several types. They may be involuntary or voluntary. For example, there may be involuntary muscular tremors or heart racing when the pit is first seen. A voluntary action is a decision to walk around the ledges at the sides of the pit room rather than just walk across the "empty space." Actions may be conscious or unconscious. A participant may not be aware of postural sway in response to an event, for example. Actions may be supported or denied by the participant. A London firefighter when once seeing an early version of the pit room clearly jumped back in order to avoid "falling in." Later, he denied that there was any such occurrence and said that he felt nothing when seeing the pit. This discussion of actions has consequences for measurement of presence.

3 Measuring Presence

Questionnaires are by far the most common means of measuring the degree of presence (IJsselsteijn et al., 2001). However, questionnaire responses can provide an integration over time of only conscious, voluntary, and supported responses.

Behavioral measures such as the looming response require events in the environment that may have nothing to do with the application. Physiological measures (Meehan et al., 2001) appear to offer a promising way forward. However, physiological measures for anxiety (as in the pit room) have been used, but they are appropriate for only a very small minority of applications. What is the physiological response appropriate to seeing a virtual chair, or a room, or the deck of a ship? Basing physiological responses on specific events or objects within the VE does not offer a general solution.

A break in presence (BIP) is the moment of switch between responding to signals with source in environment X to those with source in environment Y. It is equivalent to the *aha*! moment in gestalt psychology: the switch between seeing the duck and the rabbit, for example. If we take X as the VE and Y as the physical world in which the experience is located, then one measure of presence is the number of transitions per unit time from X to Y. It has been shown how this can be modeled as a stochastic process (Slater & Steed, 2000). The big problem with this approach is that the participant has to somehow signal the moment of transition; in the experiment reported in the earlier paper, this was done verbally. The problem is not that giving the signal breaks presence in the VE (which has already been broken) but that the count itself may be unreliable. If the count is zero, is this because there really were no transitions or because the subject forgot to report any?

A way forward with respect to measurement may be a combination of physiological measures and BIPs. Do not attempt to use physiological measures to monitor participants' responses to content, but to physiological responses heralding BIPs. We know many of the types of event that cause BIPs, and we can deliberately generate them. Is there a common pattern of physiological response to such BIPs independently of the actual event that caused them? Whether the application is concerned with architecture or ships or conversations with virtual people, is there a common physiological-BIP response that is invariant and which can be isolated under many different conditions? If so, we're in business.

4 The Usual Results

The norm in our more than ten years of experimentation with presence in VEs is that questionnaire-elicited presence is usually reported as high, especially relative to the paucity of the VE compared to the real world. See, for example, Usoh, Catena, Arman, and Slater (2000), in which reported presence in a virtual office was the same as that for the real office. Although in that paper we used the result to criticize the utility of questionnaires, there is another explanation. It seems that some minimal set of sensory cues are needed to establish presence in a place and that the mind fills in the gaps, just as Stark suggests that it does in the perception of physical reality (Stark, 1995). Hence, once a VE depicting an office has been accepted as an office, there is the same presence as in a real office.

5 Conclusions

Let's review the suppositions of this paper and draw some conclusions:

- Moment by moment, a selection mechanism organizes streams of sensory signals into an environmental gestalt. Sensory data relevant to other environmental gestalts are relegated to the background. The participant scan-senses the world according to the present gestalt.
- We "see" in our mind's eye. Therefore, it is relatively easy to fool the "eye" into selecting the hypothesis that we are in the place depicted by a VE, notwithstanding the typical paucity of the VE compared to the real world. Hence, reported presence is high on the average.
- The hypothesis selection is not a once-and-for-all event. We continue to scan-sense the world in which we are present, repeatedly returning to and fixating on perceptually significant items, repeatedly testing the presence hypothesis. An anomaly associated with a perceptually significant item may lead to a break in presence: the reformation of sensory signals into another gestalt, presence in another environment.
- Anomalies in an environment are not equal in their significance: some will induce a break in presence, and others won't. For example, in the depiction of a virtual human, an anomaly in overall body shape is likely to be far less significant than the shape and movements around the eyes and mouth.

Therefore, we can argue that there are conditions that may be necessary but not sufficient for presence in a VE: those factors that we have called *immersion* (Slater, 2000), such as a surrounding stereo display with match between sensory signals and proprioception (including, for example, head-movement parallax). A setup can support all of these things, but still perceptually significant anomalies in the VE may cause BIPs. Hence, the study of presence must be concerned with what is essential for maintaining the perception of an environment. What is the minimal set of significant features that the scan-sensing fixations return to again and again in repeatedly verifying the hypothesis that this is "this" type of place or situation in which the participant is experiencing, and not "that" type?

At some deep level, our minds do not understand "virtual reality." Hence, only minimal cues are necessary for our presence selection mechanism to respond to a virtual pit, even though we know "really" that there is not one there. The sixth sense is this process of seeing what we expect to see, and it doesn't take much for a virtual reality to convince us: we respond to events in the virtual world much as we would to similar events in the physical world. It doesn't take much? The study of presence aims at finding out exactly what it does take.

References

Basdogan, C., Ho, C.-H., Srinivasan, M. A., & Slater, M. (2000). An experimental study on the role of touch in shared virtual environments. *ACM Transactions on Computer Human Interaction, 7*(4), 443–460.

Draper, J. V., Kaber, D. B., & Usher, J. M. (1998). Telepresence. *Human Factors, 40*(3), 354–375.

Gregory, R. L. (1998). *Eye and brain: The psychology of seeing* (Fifth Ed.). Oxford: Oxford University Press.

IJsselsteijn, W. A., Freeman, J., & de Ridder, H. (2001). Editorial: Presence: Where are we? *Cyberpsychology and Behavior, 4*(2) 179–182.

Meehan, M., Insko, B., Whitton, M., & Brooks, F. (2001). Objective measures of presence in virtual environments. *Presence 2001, 4th International Workshop,* Available at: http://www.cs.unc.edu/~meehan/presence2000/MeehanPresence2000.htm.

Slater, M. (1999). Measuring presence: A response to the Witmer and Singer questionnaire. *Presence: Teleoperators and Virtual Environments, 8*(5), 560–566.

Slater, M., & Steed, A. (2000). A virtual presence counter. *Presence: Teleoperators and Virtual Environments, 9*(5), 413–434.

Stark, L. W. (1995). How virtual reality works! The illusions of vision in "real" and virtual environments. *SPIE Proceedings: Symposium on Electronic Imaging: Science and Technology,* Feb. 5–10, San Jose, California.

Stark, L. W., Privitera, C. M., Yang, H., Azzariti, M., Ho, Y. F., Blackmon, T., & Chernyak, D. (2001). Representation of human vision in the brain: How does human perception recognise images? *Journal of Electronic Imaging (Special Issue on Human Vision), 10*(1), 123–151.

Usoh, M., Arthur, K., Whitton, M., Bastos, R., Steed, A. J., Slater, M., & Brooks, F. (1999). Walking > walking-in-place > flying, in virtual environments. *Computer Graphics (SIGGRAPH),* Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, July 1999, 359–364.

Usoh, M., Catena, E., Arman, S., & Slater, M. (2000). Presence questionnaires in reality. *Presence: Teleoperators and Virtual Environments, 9*(5), 497–503.