

Iowa State University

From the Selected Works of Jonathan W. Kelly

August, 2004

Perception of Shared Visual Space: Establishing Common Ground in Real and Virtual Environments

Jonathan W. Kelly, *University of California, Santa Barbara*

Andrew C. Beall, *University of California, Santa Barbara*

Jack M. Loomis, *University of California, Santa Barbara*



Available at: https://works.bepress.com/jonathan_kelly/7/

Jonathan W. Kelly

j_kelly@psych.ucsb.edu

Andrew C. Beall

beall@psych.ucsb.edu

Jack M. Loomis

loomis@psych.ucsb.edu

Department of Psychology

University of California at Santa

Barbara

Santa Barbara, CA 93106-9660

Perception of Shared Visual Space: Establishing Common Ground in Real and Virtual Environments

Abstract

When people have visual access to the same space, judgments of this shared visual space (shared vista) can facilitate communication and collaboration. This study establishes baseline performance on a shared vista task in real environments and draws comparisons with performance in visually immersive virtual environments. Participants indicated which parts of the scene were visible to an assistant or avatar (simulated person used in virtual environments) and which parts were occluded by a nearby building. Errors increased with increasing distance between the participant and the assistant out to 15 m, and error patterns were similar between real and virtual environments. This similarity is especially interesting given recent reports that environmental geometry is perceived differently in virtual environments than in real environments.

I Introduction

The success of collaborative work in a multiperson environment depends heavily on the establishment of common ground, a concept that encompasses, among other things, shared knowledge and mutual awareness of environmental state (Clark & Marshall, 1981; Clark & Wilkes-Gibbs, 1986; Kraut, Fussell, Brennan, & Siegel, 2002; Olson & Olson, 2000). Shared visual space, one aspect of common ground, refers to portions of the environment that are visually accessible to two or more individuals simultaneously. The space visible to one person is herein referred to as a “vista,” and the areas common to the vistas of two or more people is referred to as the “shared vista.” Often this shared vista involves physical copresence of the individuals, as when they occupy the same room or nearby environment. Alternatively, a shared vista can be mediated through technologies such as video conferencing or virtual reality (VR). In both cases, a shared vista provides the potential for mutual awareness of the same environment, which helps establish common ground. Kraut, Fussell, and Siegel (2003) demonstrated this facilitative effect in a collaborative repair task (in this case, fixing a bicycle), where an expert remotely assisted a novice by way of audio contact or video-plus-audio contact. The ensuing conversations involved more pointing and deictic expressions (e.g., “this one” and “over there”) when the expert had access to the novice’s visual space by way of head-mounted cameras, resulting in more efficient communication. While task success was comparable in the two conditions, the addition of the shared vista

allowed information to be offloaded from verbal communication channels and onto other nonverbal channels. The same principles apply in side-by-side interaction, where the shared vista helps two people converse about the same object and also affords judgments of what the other person can or cannot see from his or her vantage point. These judgments must be made before one considers coordinating action with respect to some object or location.

A report on special weapons and tactics (SWAT) teams by Jones and Hinds (2002) underscores the importance of shared vistas in planning actions during a distributed task. The researchers monitored communication between SWAT team officers throughout four training missions, each involving approximately 25 team members surrounding a building. The tactical commander (TC), usually positioned at some distance from the building, was in charge of coordinating the efforts of all team members. The following conversation between the TC and two officers (Officers W and B) illustrates an attempt to assess the shared vista and plan subsequent actions based on this knowledge (Jones & Hinds, 2002, p. 377):

TC: W, do you have a visual on the suspect?

Officer W: No, (there is a) large stack of boxes between me and location (where I hear what) I believe is the suspect.

TC: B, do you see a location for W to egress to that remains in cover?

Officer B: Yes, there is a desk with a computer immediately to his left when he comes around the stack that he should be able to get to.

TC: Did you get that W?

Officer W: Affirmative, moving to the desk.

In this case, Officer B has made a judgment of the vista shared by him and the suspect. By assimilating Officer W's view of the layout with Officer B's view and the suspect's view, the TC can send out coordinated orders to the different members. To establish the necessary common ground for coordinated action, the TC was interested in not only the shared vista of Officer B and the suspect, but also the space not shared by the two, which afforded safe egress for Officer W.

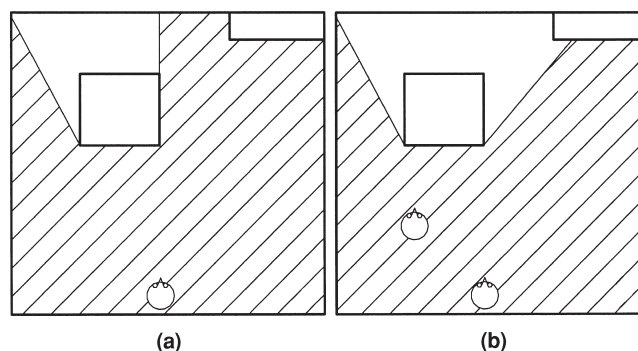


Figure 1. Panels A and B show a plan view of a rectangular room with a rectangular column in the upper left quadrant. The crosshatch area in Panel A depicts the vista of a single viewer. The crosshatch area in Panel B depicts the area of intersection between the vistas of two viewers. This area is referred to as a “shared vista.”

The process of establishing common ground through shared visual space also facilitates direct correspondence between team members (rather than mediated correspondence by the TC). Any subsequent coordinated efforts should take the shared vista into account during planning as well as online monitoring of any actions taken. In this work, we compare perceptual performance in a shared vista task in real and virtual environments. Given that much of the contemporary research and applications in VR technology involve multiple users sharing and interacting in a common space (e.g., collaborative environments, online games, and entertainment) (Leigh, DeFanti, Johnson, Brown, & Sandin, 1997; Mania & Chalmers, 1998; Schwartz et al., 1998; Normand et al., 1999; Lanier, 2001), an understanding of the perception of shared visual space is becoming increasingly important.

Accurate judgment of shared visual space is not a simple process, and a thorough analysis begs a broader understanding of physical and perceptual space. Benedikt (1979) provided an insightful analysis of the geometric and statistical properties of the environment visible from any given vantage point. Figure 1(a) shows the visible region or vista (Benedikt used the term “isovist”) of a person in a simple environment. We extend Benedikt's conceptualization of vistas to deal with the space formed by the intersection of two or more people's vistas, which

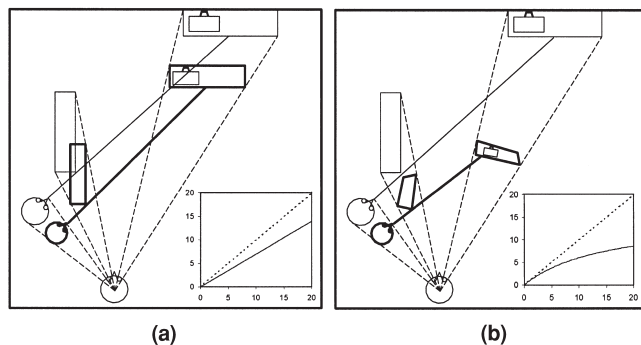


Figure 2. Panel A illustrates some of the factors involved in a shared vista judgment, including distances to objects and collinearity. Bold items represent perceived shape and locations after a linear distance compression (inset shows the linear function used). Note that the perceived shared vista is unchanged. Panel B shows how interobject relations are affected by a nonuniform compression of perceptual space based on a hyperbolic function (graphically represented in the inset). The perceived shared vista no longer corresponds to the actual shared vista.

we call a shared vista. Figure 1(b) shows a shared vista for two persons.

While Benedikt (1979) was interested in properties of physical space, we are interested in perceptual space. A reasonable starting point is to assume that accurate judgment of any shared vista should depend on the accurate perception of environmental geometry, including distances and directions to all relevant objects. Figure 2 shows how an observer might determine whether an object is visible to someone else in the room. In this case, he or she wants to know if the other person can see a briefcase lying on a table across the room. One way to perform the task requires the observer to determine the locations of the other person, the edge of the occluding column, and the table (this requires perception of both distance and direction). Based on these perceived locations, the observer can then extrapolate an imaginary line from the person to the occluding edge to the table (this is essentially a collinearity judgment). If that imaginary line falls on the left side of the briefcase, it is visible to the other person; if it falls on the right side, it is not visible from that vantage point. Team sports provide many such situations where judgments of

another person's visual space are critical to team success. An alert soccer player, for instance, will be aware of which teammate has an open view of the goal before deciding where to pass the ball. If the ball carrier is able to accurately compute the relevant geometric relationships from his or her own vantage point, then it should be a straightforward process to predict the visibility of objects from another vantage point. However, in the case of human observers, neither the access to accurate distance and direction information nor the ability to compute general geometric relationships can be assumed.

Some studies indicate correct judgment of distance and direction in full-cue outdoor environments (e.g., Fukushima, Loomis, & Da Silva, 1997; Loomis & Knapp, 2003), while others indicate errors in distance perception (Da Silva, 1985; Gilinsky, 1951). In particular, the question of whether perceived distance is related to physical distance by a linear transform or by a compressive nonlinearity is debatable. For the time being, consider how both cases might affect perception of a shared vista. Let us return to the situation in Figure 2(a). If distances to the other person, the column, and the briefcase are perceived to be 70% of their physical distance, the percept will be of a uniformly scaled room. Although this error might impact interactions with objects in the environment, it will not affect perception of the shared vista. In both cases, the observer concludes that the briefcase is out of view. However, if distance perception is nonlinear, the scaling of the perceived room will be nonuniform (Figure 2(b)), resulting in erroneous perception of the shared vista (in this hypothetical case, the observer incorrectly deems the briefcase visible from the other vantage point).

Virtual environment technology based on head-mounted displays (HMDs) produces visual stimulation that differs from viewing real environments in important ways. Some of the more important differences include reduced field of view, fixed accommodation, optical distortion (typically greatest in the periphery), reduced dynamic range of illumination, compressed color gamut, potential destabilization of the visual world due to tracking latencies, and decreased spatial resolution due to display quantization. How these artifacts factor

into altering the perception of visual space is a complex and poorly understood issue. The research that has been conducted testing visual space perception in virtual environments has consistently found that geometric properties of virtual environments are perceived differently than they are in the real world (Bingham, Bradley, Bailey, & Vinner, 2001; Ellis & Menges, 2001). Results show significant distortions of properties such as distance and size of objects (Loomis & Knapp, 2003; Thompson et al., in press). Given distortions such as these, the question arises of whether there is an impact on the perception of shared visual space in VR. In order to naturalistically coordinate and execute actions in a shared virtual environment, it is important that the virtual environment be perceived similarly to the environment being simulated. If two people cannot trust that the new technology will allow them to perceive shared visual space correctly, they will want to supplement nonverbal communication with verbal communication, resulting in a loss of efficiency. More seriously, if they falsely believe that the technology is providing accurate information when it is not, there may be outright miscommunication and errors in performance of collaborative tasks. If VR is to be considered a useful training tool for multiple interactants, it is important that skills acquired virtually be applicable in the real world. Specifically, strategies for establishing common ground in multiperson virtual environments should be effective in the real world as well.

To assess human performance at judging shared vistas in both real and virtual environments, we devised a simple task where a participant judges which parts of an environment are visible from a confederate's point of view.

2 Methods

2.1 Design

In a 2×3 fully factorial design, there were two levels of environment type (real and virtual) and three levels of distance to the assistant/avatar (5, 10, and 15 m). For each environment type, there were three geometrically equivalent locations in order to obtain

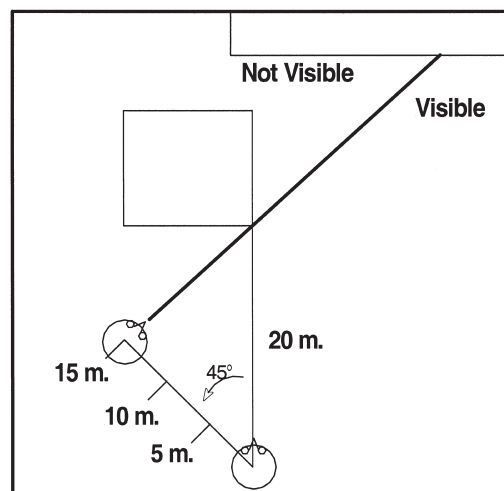


Figure 3. Plan view of a large outdoor scene used in the experiment. Geometry was the same or mirror-reversed for all environments, both real and virtual.

multiple judgments for each condition. Thus, each participant made 18 judgments in all.

2.2 Participants

Twelve students at the University of California, Santa Barbara were paid \$10 for their participation. Participation took approximately 1 h. The age range of the participants was 18–23, with six males and six females.

2.3 Stimuli and Apparatus

The geometric structure of all environments, both real and virtual, is depicted in Figure 3. In real environments, participants stood at the origin at all times while the assistant stood either 5, 10, or 15 m away. The assistant faced and looked at the occluding edge of a building, which was always 20 m from the participant. Angular separation between the assistant and the occluding edge was held constant at 45° . For real world environments, participants were given a photograph that depicted a 70° horizontal by 35° vertical view of the scene in front of them, taken from their perspective. Participants were then asked to judge which parts of the

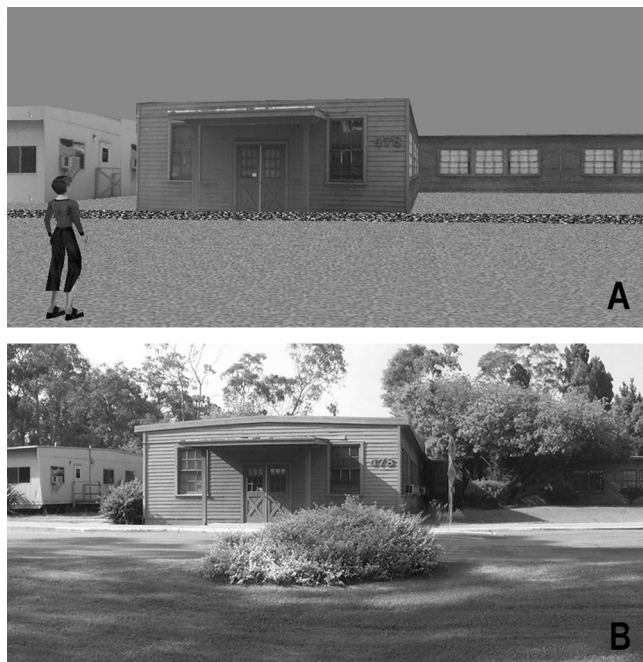


Figure 4. Panel A is a screenshot from one virtual environment used. Panel B is a photograph of the real world scene (assistant not shown).

background scene would be visible to the assistant, and to indicate the perceived point of occlusion on the photograph.

The virtual models of the aforementioned environments were somewhat photorealistic, using texture maps captured from the real world environments. All virtual worlds included models of the relevant objects (i.e., assistant, occluding buildings, and background scene). The avatar that replaced the assistant was a polygon-based model of a Caucasian female. She was positioned at the same distances from the participant (5, 10, and 15 m), and always faced the occluding edge of the building. Figure 4 shows one of the locations in both the real and virtual conditions. In VR, subjects indicated the perceived point of occlusion using a pointer in the virtual world (rather than having to refer to a photograph, as in the real world condition). It should be noted that the judgment was the same in both real and virtual environments. In both cases, subjects were judging the point of occlusion on the background scene

from the confederate/avatar's vantage point. The only difference was in the method of response.

The head-mounted display used to present the virtual environments was a Virtual Research V8 HMD (a stereoscopic display with dual 680×480 resolution LCD panels that refresh at 60 Hz). The visual scene spanned 50° horizontally by 38° vertically. Projectively correct stereoscopic images were rendered by a 1 GHz Pentium 4 processor computer with a GeForce 2 Twinview graphics card. The simulated viewpoint was continually updated by the participant's head movements. The orientation of the participant's head was tracked by a three-axis orientation-sensing system (Intersense IS300), while the location of the participant's head was tracked three dimensionally by a passive optical position-sensing system (developed in-house and capable of measuring position with a resolution of 1 part in 30,000, or approximately 0.2 mm in a 5 m^2 workspace). The system latency, or the amount of delay between a participant's head or body motion and the concomitant visual update in the HMD was 42 ms maximum.

2.4 Procedure

Participants completed all real world conditions first, followed by all virtual conditions. Six participants proceeded through the locations in one order, and six went in a reverse order. The order of distances (from participant to assistant/avatar) was randomized within each location.

Participants were led to each real world location in a manner that prevented them from gaining any information about the scene from the vantage point that the assistant would assume. When the assistant was standing at the proper distance, looking at the occluding edge of the building, participants were asked to judge which parts of the background scene were visible to the assistant and which parts were occluded by the building. Subjects were read the following instructions:

This is an experiment to study what we call vistas. In particular, we're interested in how well someone can imagine someone else's vista. By vista, we mean the view of an environment at a particular location and

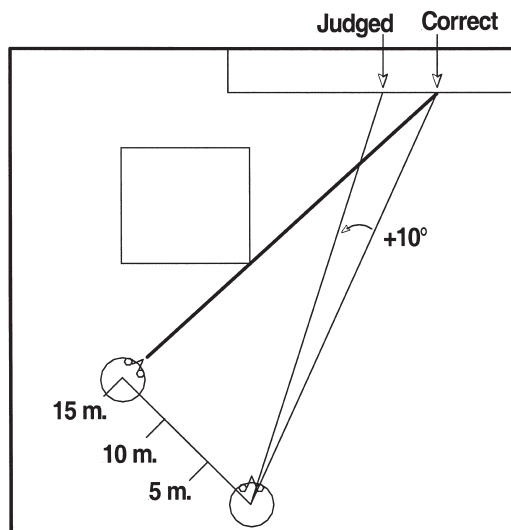


Figure 5. Depiction of $+10^\circ$ error. Here, the observer has overestimated the shared vista by 10° of visual angle.

what is visible and not visible at that location. Your task is to imagine what the scene would look like from another location. Try to visualize the exact location in the far scene that would just be the breaking point between what would be visible and what would not be visible. That's where I want you to draw a line.

Observers then responded by drawing a vertical line indicating the perceived point of occlusion on the photograph. Once all three judgments were completed (with the assistant 5, 10, and 15 m away), the task was repeated at two more locations that provided the same underlying geometric configuration of distances. Once all real world conditions were completed, participants were led back to the lab and completed the same task in VR. As noted above, the judgment was the same but the response method was slightly different. Rather than drawing on a photograph of the vista before them, they aimed a pointer at the perceived point of occlusion.

3 Results

All data were computed in terms of angular error, where an error of $+1^\circ$ represents an overestima-

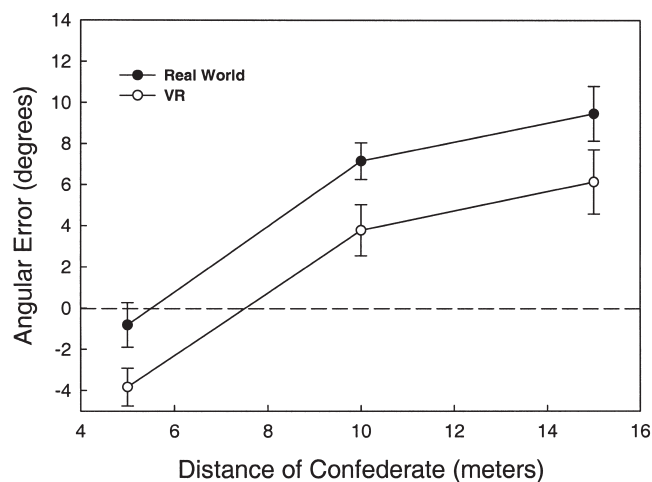


Figure 6. Mean angular error in judgments of the shared vista as a function of distance to the assistant/avatar. Error bars represent \pm one standard error of the mean.

tion of the area visible to the assistant by one degree of visual angle from the participant's perspective (see Figure 5). For the VR trials, this value was directly defined by the angular difference between the pointer and the location of the correct response in polar coordinates, with the observer at the origin. In the real world condition, this value was extracted from the photograph on which subjects recorded their responses.

Using this definition of error, Figure 6 shows that the shared vista is increasingly overestimated as the assistant moves from 5 to 10 to 15 m away, for both real and virtual environments. When the assistant is 10 or 15 m away, mean observer estimates indicate that the assistant can see more of the background than is geometrically possible.

A two-way repeated measures ANOVA was conducted to evaluate the effect of environment (real world or virtual) and the distance of the assistant (5, 10, or 15 m) on perception of shared visual space. The environment main effect was significant, $F(1,35) = 9.54$, $p < .01$, as was the distance main effect, $F(2,34) = 46.59$, $p < .01$, with no significant interaction, $F(2,34) = 0.02$, ns.

4 Discussion

4.1 Similarities between Environment Types

Of primary importance is the similarity between error patterns in real and virtual environments: both show monotonically increasing error of the same sign as the assistant moves from 5 to 10 to 15 m from the observer (see Figure 6). While the angular errors are approximately 3° larger in the real world environments, this effect is small relative to the 10° increase in error seen in both environment types as the assistant/ avatar moves away from the observer. How surprising this similarity is depends very much on the particular form of distortion introduced by VR. Loomis and Knapp (2003) showed that perceived distance in the virtual environments they studied was, to a first approximation, about one half of the simulated distance. In this case, a uniform underperception of scale would have no impact on judgments of certain properties, such as angles and collinearity. In light of the similar error patterns found here, we suspect that shared vista judgments are based on one of these “invariant” properties rather than on absolute egocentric distance.

4.2 Sources of Judgment Error and Strategies

The pattern of increasing angular error with increasing distance to the assistant (observed in both environments; see Figure 6) is similar to results obtained by Cuijpers, Kappers, and Koenderink (2000), where subjects oriented a remotely controlled pointer to point at a target. Essentially they made a judgment of exocentric direction (i.e., the direction of an imaginary line connected by the pointer and target) for targets ranging out to 4 m. In their study, angular error was dependent upon the ratio of egocentric distances to the pointer and the target. More recent work by Kelly, Loomis, and Beall (2004) on judgments of exocentric direction suggests that this error pattern is independent of egocentric distance out to at least 20 m. In the current study, the assistant can be construed as a pointer, and the target is represented by the occluding edge of the building. The close correspondence of error patterns sug-

gests that shared vista judgments can be reduced to judgments of exocentric direction, where overestimation of the shared vista represents overestimation of the angular orientation of an imaginary line connecting the assistant and the occluding edge of the building. Now the similarity in error patterns in the two environments makes more sense, since uniform scaling of absolute distance perception will not change the ratios of these distances. It should be noted that the stimulus in the current experiment was overdefined, in that observers could judge the direction of the imaginary line connecting the assistant and the occluding edge, or they could just assess the facing direction of the assistant (as the assistant was always looking directly at the occluding edge).

An alternative solution for judging shared visual space in this study can be performed on a 2D projection of 3D space. For objects on a planar surface, collinearity of the objects in 3D space implies collinearity of the corresponding images in a planar projection (e.g., the retinal image). Thus, in the shared vista task, if an observer knows that all three objects (the assistant, the occluding edge, and the background scene) lie on a ground plane, a shortcut strategy becomes available: they can find the point on the background scene collinear (in the projective image) with the assistant and the occluding edge.

In the current study, all relevant objects lay on the ground plane, and all judgments could potentially be based on a 2D perspective view. There is reason to believe subjects did not use this strategy. The work done by Cuijpers et al. (2000) on exocentric direction provides an excellent control, since they presented all objects at eye level, rendering the 2D strategy ineffective. Thus, the errors reported by Cuijpers et al. must be due to errors in 3D space perception. Given the striking similarity between those error patterns and the errors obtained here, it is safe to assume that observers based their judgments of the shared vista on perceived 3D layout, not 2D collinearity.

4.3 Implications for Common Ground in Virtual Reality

The similar error pattern found in natural and virtual environments indicates that certain aspects of virtual environments are treated similarly to the real world

environments they represent. Given that the shared vista contributes to the establishment of common ground in collaborative tasks, and that common ground is fundamental in planning complex coordinated actions, it is important that the collaborators extract accurate environmental information regarding the nature of the shared vista. However, participants in the current experiments made systematic judgment errors up to 10° in certain conditions. Errors of this type could cause false beliefs regarding the common ground between two interactants. In cases where precise judgment of shared visual space is critical to performance of a group task (such as SWAT team exercises), 10° errors could have serious consequences.

To the extent that training can ameliorate the perceptual judgment errors shown in the current experiments, training in a virtual environment should be transferable to real world environments. The applicability of VR for training on collaborative tasks is promising, especially for extreme work groups such as firefighters and SWAT teams where real world simulation is costly. In these situations, the ability to accurately perceive a shared vista is vital to the planning and control of group action.

Acknowledgments

This research was supported by ONR grant N00014-01-1-0098. We thank Sarah Meyer and Joe Hayek for their assistance.

References

- Benedikt, M. L. (1979). To take hold of space: Isovists and isovist fields. *Environment and Planning B*, 6, 47–65.
- Bingham, G. P., Bradley, A., Bailey, M., & Vinner, R. (2001). Accommodation, occlusion, and disparity matching are used to guide reaching: A comparison of actual versus virtual environments. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 145–168.
- Clark, H. H., & Marshall, C. E. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Weber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge, UK: Cambridge University Press.
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22(1), 1–39.
- Cuijpers, R. H., Kappers A. M. L., & Koenderink, J. J. (2000). Investigation of visual space using an exocentric pointing task. *Perception and Psychophysics*, 62(8), 1556–1571.
- Da Silva, J. A. (1985). Scales for perceived egocentric distance in a large open field: Comparison of three psychophysical methods. *American Journal of Psychology*, 98(1), 119–144.
- Ellis, S. R., & Menges, B. M. (2001). Studies of the localization of virtual objects in the near visual field. In W. Barfield & T. Caudell (Eds.), *Fundamentals of wearable computers and augmented reality* (pp. 263–293). Mahwah, NJ: Erlbaum.
- Fukushima, S. S., Loomis, J. M., & Da Silva, J. A. (1997). Visual perception of egocentric distance as assessed by triangulation. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 86–100.
- Gilinsky, A. S. (1951). Perceived size and distance in visual space. *Psychological Review*, 58, 460–482.
- Jones, H. L. & Hinds, P. J. (2002). Extreme work groups: Using SWAT teams as a model for coordinating distributed robots. *Proceedings of the ACM 2002 Conference on Computer Supported Cooperative Work (CSCW 2002)*, 372–381.
- Kelly, J. W., Loomis, J. M., & Beall, A. C. (2004). Judgments of exocentric direction in large-scale space. *Perception*, 33(4), 443–454.
- Kraut, R. E., Fussell, S. R., Brennan, S. E., & Siegel, J. (2002). Understanding the effects of proximity on collaboration: Implications for technologies to support remote collaborative work. In P. J. Hinds & S. Kiesler (Eds.), *Distributed work* (pp. 137–162). Cambridge, MA: MIT Press.
- Kraut, R. E., Fussell, S. R., & Siegel, J. (2003). Visual information as a conversational resource in collaborative physical tasks. *Human-Computer Interaction*, 18, 13–49.
- Lanier, J. (2001, April 1). Virtually there. *Scientific American*, 66–76.
- Leigh, J., DeFanti, T., Johnson, A., Brown, M., & Sandin, D. (1997). Global tele-immersion: Better than being there. *Proceedings of the Seventh International Conference on Artificial Reality and Tele-existence*, 10–17.
- Loomis, J. M. & Knapp, J. M. (2003). Visual perception of egocentric distance in real and virtual environments. In L. J. Hettinger and M. W. Haas (Eds.), *Virtual and adaptive environments* (pp. 21–46). Mahwah, NJ: Erlbaum.
- Mania, K., & Chalmers, A. (1998). A classification for user embodiment in collaborative virtual environments. *Proceed-*

- ings of the Fourth International Conference on Virtual Systems and Multimedia*, 177–182.
- Normand, V., Babski, C., Benford, S., Bullock, A., Carion, S., Chrysanthou, Y., et al. (1999). The COVEN project: Exploring applicative, technical, and usage dimensions of collaborative virtual environments. *Presence: Teleoperators and Virtual Environments*, 8(2), 218–236.
- Olson, G., & Olson, J. (2000). Distance matters. *Human Computer Interaction*, 15(2/3), 139–179.
- Schwartz, P., Bricker, L., Campbell, B., Furness, T., Inkpen, K., Matheson, L., et al. (1998). Virtual playground: Architectures for a shared virtual world. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, 43–50.
- Thompson, W. B., Willemsen, P., Gooch, A. A., Creem-Regehr, S. H., Loomis, J. M., & Beall, A. C. (in press). Does the quality of the computer graphics matter when judging distances in visually immersive environments? *Presence: Teleoperators and Virtual Environments*.