Using high-fidelity virtual reality to study perception in freely moving observers

Peter Scarfe

School of Psychology and Clinical Language Sciences, University of Reading, Reading, Berkshire, UK

Andrew Glennerster

School of Psychology and Clinical Language Sciences, University of Reading, Reading, Berkshire, UK

Technological innovations have had a profound influence on how we study the sensory perception in humans and other animals. One example was the introduction of affordable computers, which radically changed the nature of visual experiments. It is clear that vision research is now at cusp of a similar shift, this time driven by the use of commercially available, low-cost, high-fidelity virtual reality (VR). In this review we will focus on: (a) the research questions VR allows experimenters to address and why these research questions are important, (b) the things that need to be considered when using VR to study human perception, (c) the drawbacks of current VR systems, and (d) the future direction vision research may take, now that VR has become a viable research tool.

Introduction

Nearly all animals gain three-dimensional (3-D) information about the world by moving around, exploring, and dynamically and iteratively responding to their environment. As such, our view of the world is rarely static; the sensory information we receive about it is constantly changing as we sample our environment in a task-dependent way (Clark, 2013; Gibson, 1950, 1979; Land, Mennie, & Rusted, 1999; O'Regan & Noe, 2001). Despite this, perception is rarely studied in such situations. Participants in a typical perceptual experiment are treated as passive observers of their environment. They view sparse, briefly presented, reduced-cue stimuli, from a static viewpoint, often even preventing eye movements. Restricting stimuli in such a way acts to isolate the sensory information under study, but at the cost of investigating perception in a way that often bears little resemblance to anything we normally experience. Because of this, there is active debate as to whether conclusions based on such artificial stimuli

really reflect the way in which the visual system works (De Gelder & Bertelson, 2003; Felsen & Dan, 2005; Rust & Movshon, 2005).

While tightly constrained stimuli clearly have an important role in vision research, our goal here is to draw attention to the specific benefits of studying perception in a more naturalistic context now that this is a realizable goal. One of the primary reasons why perception is not studied in more naturalistic situations is that technological limitations have made it impossible to produce well-controlled computer-generated stimuli that are updated in real time as an observer moves. Even though virtual reality (VR) systems were first investigated in the late 1960s (Sutherland, 1968), it is only in the last 10 years that has it been possible to use high-fidelity stereoscopic VR, coupled with full-body motion tracking over large-scale volumes, to investigate perception. Recent developments in VR raise the prospect that high-quality VR will soon be within the reach of most researchers in the field. With the acquisition of VR startup Oculus by Facebook for $2 billion, and increased interest from companies such as Sony, Samsung, HTC, Raser, and Google, it appears highly likely that VR will become a mass-market commodity.

Why use VR to study human perception?

Using immersive VR allows experimenters to move closer to the goal of investigating human perception in a controlled and principled way, but with highly realistic stimuli that can be actively explored by experimental participants. This level of ecological validity is simply not possible with traditional experimental techniques, but is far more akin to the way we normally interact with the world. VR allows experi-

menters to manipulate the environment in complex and even physically impossible ways, so as to examine the contribution of different sources of sensory information to perception as observers move through a scene (Bruggeman, Zosh, & Warren, 2007; Jain & Backus, 2010). Research based on these principles has revealed surprising aspects of perception, such as how massive distortions of the environment can go completely unnoticed (Glennerster, Tcheang, Gilson, Fitzgibbon, & Parker, 2006).

Studies such as these have been the basis of a systematic investigation of the principles underlying spatial representation in moving observers and the necessary conditions for perceiving a stable world (Pickup, Fitzgibbon, & Glennerster, 2013; Svarverud, Gilson, & Glennerster, 2010, 2012). VR is ideal for the study of navigation, either in large-scale physical space (Foo, Warren, Duchon, & Tarr, 2005; Tarr & Warren, 2002), or through using a treadmill (Schwaiger, Thummel, & Ulbrich, 2007; Souman et al., 2011). Using VR, it is simple to manipulate cues to distance, such as the height of the horizon (Messing & Durgin, 2005), and to make worlds that are dynamically reconfigured as people navigate within them (Schnapp & Warren, 2007). This contrasts starkly with traditional experimental techniques where the observer is unable to interact in any meaningful way with what they see. Thus, the use of VR brings into focus core debates about human perception and offers novel and unique ways in which to address them.

For example, key differences exist in depth perception for moving and static observers that can only be studied when the observers are free to move and interact with their environment (Tcheang, Gilson, & Glennerster, 2005; van Boxtel, Wexler, & Droulez, 2003; Wexler, Panerai, Lamouret, & Droulez, 2001; Wexler & van Boxtel, 2005). There has been intense debate over the etiology of human perceptual biases in attributes such as distance, depth, and shape (Brenner & Landy, 1999; Johnston, 1991; Todd, Chen, & Norman, 1998; Todd & Norman, 2003; Todd, Tittle, & Norman, 1995), with recent research showing that with the simple adoption of a more naturalistic viewpoint, bias in perceived 3-D shape can be completely eliminated (Scarfe & Hibbard, 2013). The elimination of bias suggests that observers utilize different visual cues in more naturalistic settings, which allow them to accurately estimate object properties, and it is the removal of these cues in the constrained reduced-cue setting that causes the perceptual bias. These issues are closely linked to the inherent problems in isolating single cues (Todd, Christensen, & Guckes, 2010; Zabulis & Backus, 2004) and the active debate in how to generalize from single-cue to multi-cue situations (Mon-Williams & Bingham, 2008).

VR cannot provide a magic solve-all solution to these questions about cue utilization and integration but, in combination with traditional experimental techniques, it offers a bridge towards studying sensory perception in a controlled, principled fashion using stimuli that realistically reflect how we interact with the world. This means far more than just presenting a range of different sources of sensory information; instead VR recreates the natural way in which perception and action are intimately entwined with the environment (Clark, 1997, 2008, 2013; Ellis, 1991; Shapiro, 2011; Tarr & Warren, 2002; Thelen & Smith, 1994; Varela, Thompson, & Rosch, 1991). Without this shift in perspective, researchers risk solely studying how observers behave in an experiment, rather than how they behave in real life. It is the latter that most experimenters *hope* they are studying.

Beyond pure vision research, VR is increasingly being used in the social sciences (Fox, Arena, & Bailenson, 2009). Studies have investigated how a participant's conscious perception of body ownership can be dramatically manipulated by altered sensory-motor contingencies in simulated multi-actor environments (Slater, Spanlang, Sanchez-Vives, & Blanke, 2010). Others have shown how experience in VR can modulate prosocial behavior (Rosenberg, Baughman, & Bailenson, 2013) and social influence (Bailenson & Yee, 2005). In applied settings, VR is being used in both rehabilitation (Jack et al., 2001; Rizzo & Kim, 2005; Sharkey, 2014) and medical training (Tse et al., 2010). Therefore, just as with the advent of computer-generated stimuli, VR offers the real possibility of a large-scale shift in the study of all aspects of perception. In the remainder of the paper we discuss the past, present, and (near) future of VR technology and the influence that the availability of this technology may have on the way that perception and motor control are studied.

VR: Technology, past, present, and (near) future

Past

Before the advent of affordable consumer computers, stimuli used for vision research were highly impoverished compared to those generated with modern day systems. Running an experiment often required building a unique mechanical apparatus capable of a very limited range of functions. The first VR system is widely considered to be a stereoscopic head-mounted display (HMD) developed by Ivan Sutherland and Bob Sproull at the University of Utah

(Sutherland, 1968). The system consisted of two head-mounted cathode ray tubes (CRTs), each offering 40° field of view (FOV). The headset was too heavy to be worn, so it had to be mounted on the ceiling. Mechanical and ultrasonic head position sensors allowed the virtual environment to be yoked to movement within a 6 ×6 × 3-ft. volume, with approximately 40° up-down head tilt. The simulated environment consisted of simple wireframe models and could be viewed either in isolation, or superimposed on the real world via prisms (this would now be called "augmented reality"). At the time, no general-purpose computers were fast enough to drive the displays, so special-purpose computing hardware and software had to be developed. This allowed the low-resolution displays to be driven at 30 Hz.

Although technology has advanced greatly, this work was so prescient that the problems Sutherland (1968) described are still key to achieving good VR today. In technological terms, the ultimate goal of VR is to be able to simulate an environment that is indistinguishable from the real world. For vision, this means that the images the eyes receive must be identical to those that they would receive if they were looking at real-world objects, and when the person moves these images must change in exactly the same way as if the observer moved relative to these objects in the real world (Sutherland, 1968; see Ellis, 1991, for a review of early technology and its context).

In the early to mid 1990s, VR systems appeared to be gaining traction, but many of the same problems still persisted. HMDs remained bulky and cumbersome and required expensive specialist workstations to simulate even a simple environment yoked to people's movements. Displays suffered from poor spatial and temporal resolution, and this, coupled with poor geometric calibration and high latency tracking, meant that people viewed a pixelated and distorted simulation of the world that lagged noticeably relative to their movements. This had problematic effects both behaviorally, such as the effects of a constrained FOV on distance perception (Creem-Regehr, Willemsen, Gooch, & Thompson, 2005; Knapp & Loomis, 2004), as well as physiologically, such as the risk of adverse symptoms including headache and nausea (Mon-Williams, Plooy, Burgess-Limerick, & Wann, 1998; Mon-Williams, Wann, & Rushton, 1993; Wann, Rushton, & Mon-Williams, 1995). Overall, these issues made VR unsuitable for scientific research or consumer use.

Present

Today's systems are still not perfect, but technology has progressed to such an extent that many more labs are now starting to use VR to investigate all facets of human perception. Interestingly, VR companies are becoming increasingly interested in basic research on human perception to design and optimize VR technology, offering clear scope for scientific research to shape future technology. An example of a well-characterized high-end HMD is the NVIS SX111 (NVIS, Inc., Reston, VA) used in the authors' lab (Figure 1). This weighs 1.3 kg and has a vertical FOV of 72°, a horizontal FOV of 102°, and 50° of horizontal binocular overlap. The LCD displays for each eye have a resolution of 1280 × 1024 pixels (3.6 arcmin per pixel) and a refresh rate of 60 Hz. The SX111 is a vast improvement on early systems. However, advances in VR technology mean that it will soon be outperformed in terms of resolution, refresh rate, FOV, binocular overlap, and other factors by cheaper, commercially available consumer headsets from companies such as Oculus, Samsung, Sony, Microsoft, HTC, and Raser (amongst others).

Currently, these consumer products are largely limited to preproduction developer kits, rather than finished commercially available products. Those HMDs that have been released to a wider consumer market are typically based on an enclosure to hold a mobile phone that acts as the display device. However, it is clear that in a short period of time, consumer VR products will mean that researchers no longer have to pay the premium prices that have been demanded for research-grade VR equipment, which has arguably held back its wider adoption. Currently, consumer brand companies are aware of the hype and subsequent failure of VR in the 1990s and want to ensure that rich VR consumer content is in place before officially releasing a commercial product. A side benefit of this is that companies such as Oculus and Raser have adopted an open-source software (and hardware) model, which will allow researchers the freedom they need for scientific research. However, this is not yet the case for all companies.

Tracking systems have also improved vastly compared to early systems and now offer the high spatial and temporal resolution needed for accurate real-time rendering. As an example, the 14-camera Vicon tracking system (MX3 and T20S cameras; Vicon, Oxford, Oxfordshire) used in the authors' lab tracks small passive markers resolved to submillimeter precision over a large viewing volume and generates coordinates at 240 Hz (Figure 2; Movie 1). Using such a system, we are able to achieve an end-to-end latency of 32 ms (Gilson & Glennerster, 2012). The drawback with these systems at the moment is that, to achieve this accuracy, passive or active markers must be placed on the tracked objects (Gilson, Fitzgibbon, & Glennerster, 2006, 2011; Glennerster et al., 2006). Upcoming consumer headsets incorporate tracking markers into the headset itself or small handheld devices used to interact with the virtual world. This is an elegant and cost-effective solution, but
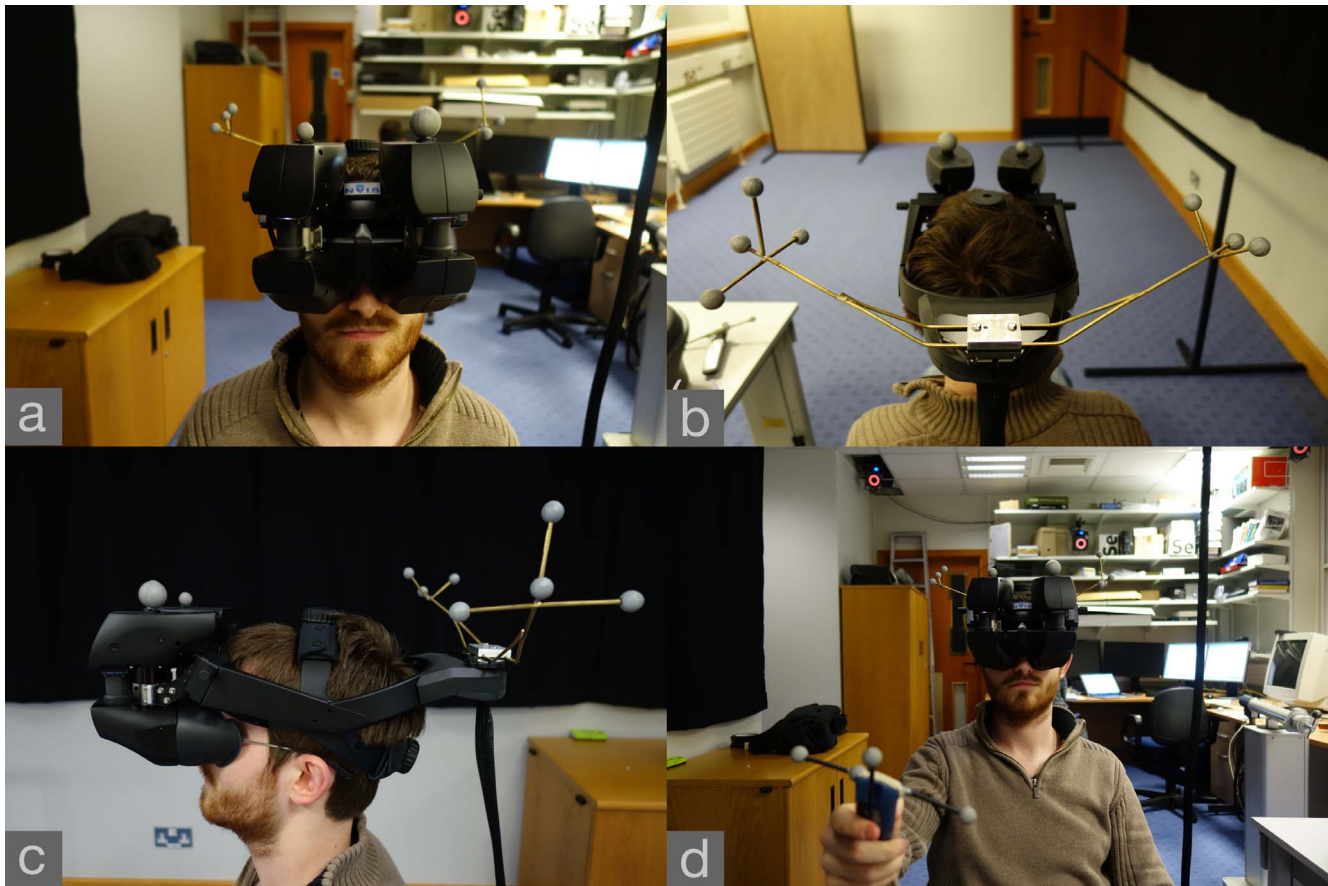
Figure 1. Shows four photos of a person wearing the headset currently used in the author's lab (NVIS SX111), each taken from a different position. The main components of the headset can be seen. (a) Shows a front view highlighting the positions of the two screens, one for each eye, each with a reflective tracking marker on their top. (b) Shows a back view of the "antlers," which hold the majority of the tracked markers. These ensure that markers are not occluded during movement. The markers define a headset model, which is tracked as the person moves within the lab. (c) Shows a side view with the screens, head straps, and antlers visible. (d) Shows a wider shot, including two of the Vicon tracking cameras (the lab has 14 cameras), and a tracked button box used for recording responses (e.g., judgments of visual direction).

currently movements can be tracked over a relatively small spatial volume compared to those achieved by specialist tracking equipment. Some systems choose to track larger volumes with much lower spatial and temporal precision, while others allow tracking only of the head, not other objects, and in some cases this is limited to rotations, not translations.

One of the key innovations made in the authors' lab has been the ability to calibrate the HMD so as to provide geometrically correct perspective projection of the 3-D scene as the observer moves through the virtual environment (Gilson et al., 2011). This greatly improves user comfort in using VR and is essential in order to make accurate inferences about behavior. The visual system often exhibits biases, even with real-world stimuli (Watt, Akeley, Ernst, & Banks, 2005), so in order to study these VR it is essential to be certain that they are not due to the use of VR itself. The calibration process is automated and allows the recovery of the extrinsic and intrinsic parameters of the left and right eye frustums of the HMD. These relate the 3-D coordinates of objects in the simulated world to the image space of the HMD. This allows the experimenter to generate geometrically correct perspective projection of the simulated environment as a person moves within it (Figure 3; Movie 2). Similar techniques can be used with future commercially available headsets.

In addition to good spatial calibration, advances in consumer grade computers and graphics cards, fueled by Moore's Law,[1] have been greatly improved by the ability to render highly detailed virtual environments yoked to full-body movements with minimal temporal lag. Indeed, it is now possible to render images that are indistinguishable from photos of real objects on standard computing equipment (Figure 4), although these can require processing times measured in days. To render virtual scenes, labs typically use general-purpose software environments such C/C++, Python, and
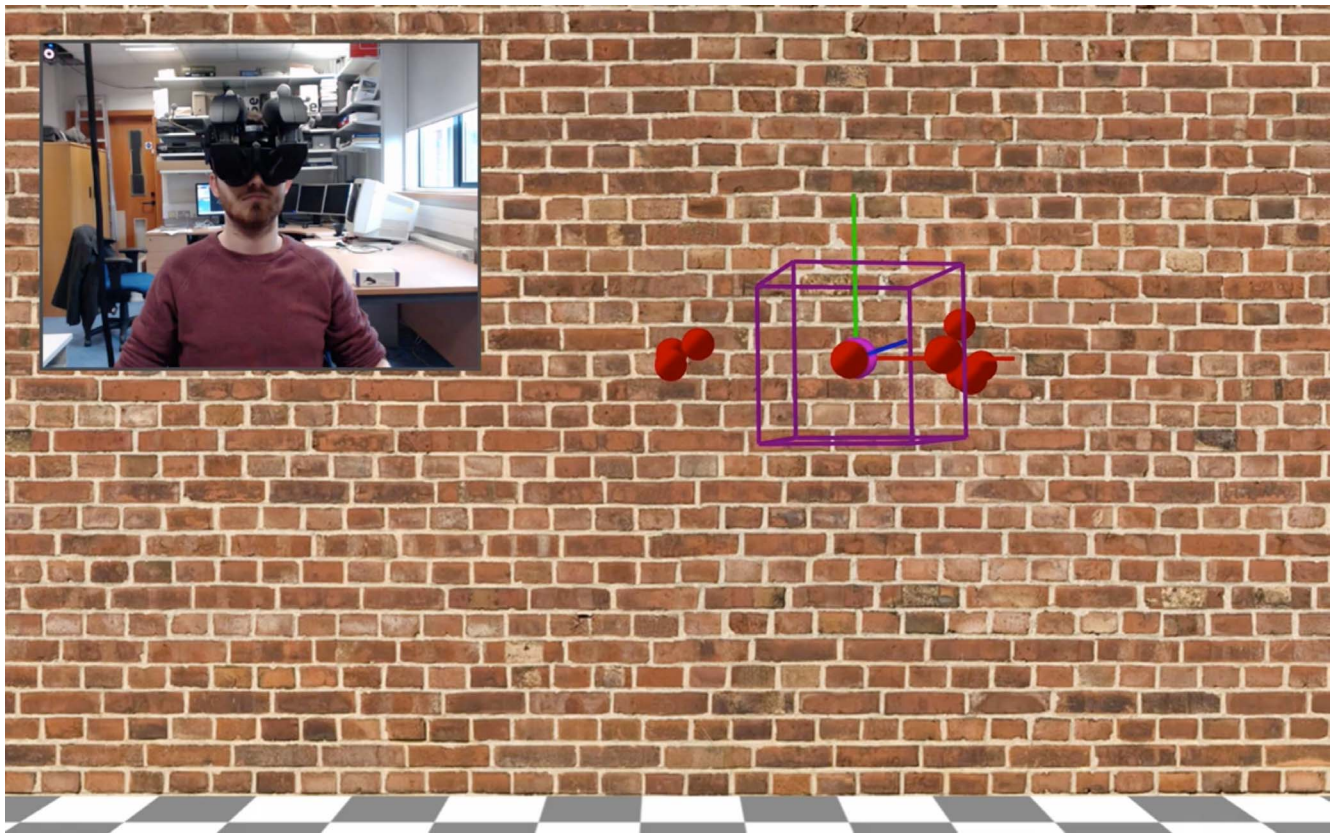
Figure 2. Rendering of a virtual environment with brick-textured walls and a checkerboard-textured floor. In this rendering the headset markers are shown as red spheres, the tracked center of the headset model is shown as a purple sphere, with an attached head-centered coordinate frame shown by the red (*x*-dimension), green (*y*-dimension), and blue (*z*-dimension) axes. The purple wireframe cube is rendered with the position and orientation of the person's head. Inset is a photo showing a person wearing the headset in the lab, at the exact moment when the coordinates for the rendering were captured. In the top left of the inset image you can see one of the 14 Vicon cameras used for tracking. A dynamic version showing head movement and rotation is shown in Movie 1.

Matlab. The use of scripting languages such as Matlab and Psychtoolbox (Kleiner, Brainard, & Pelli, 2007) is especially important as it opens up the use of VR to a much wider potential pool of experimenters, including those without specialized programming experience. In addition to research-led computing languages, game development engines now allow highly realistic stimuli to be simulated for experimental research.

Current game engines include Unreal Engine, Unity, and CryEngine, all of which offer graphical user interfaces, in addition to traditional scripting. Because this is a key area for commercialization, VR companies such as Oculus and motion-tracking vendors such as Vicon already provide integration with these systems. However, one area in which game engines are currently suboptimal for vision research is in the precise control of stimulus timing. This is because, traditionally, games can run at much lower (and variable) refresh rates than needed for motion-tracked VR. This is set to change with the adoption of VR by game development companies, hand in hand with the development of VR itself. An additional benefit of game development

environments is that code can be deployed across multiple platforms, including the mobile phones that are used in a number of commercial headsets.

(Near) future

VR technology is currently progressing rapidly. In the near future, HMDs are likely to be far lighter and less bulky, making them much easier and less intrusive to use. At the moment, most HMDs need tethering to the computer equipment used for rendering, which restricts the scope of movement. Wireless options exist, but usually at the cost of increased latency, so experimenters trade freedom of movement for temporal precision. Tracking systems still require the use of markers placed on the objects to be tracked. These can either be passive, such as the Vicon tracking system used in our lab, or active, such as Phase Space. The future of tracking is likely to be markerless, requiring little or no set-up. Affordable markerless tracking solutions, such as *Leap Motion* (Leap Motion, Inc., San
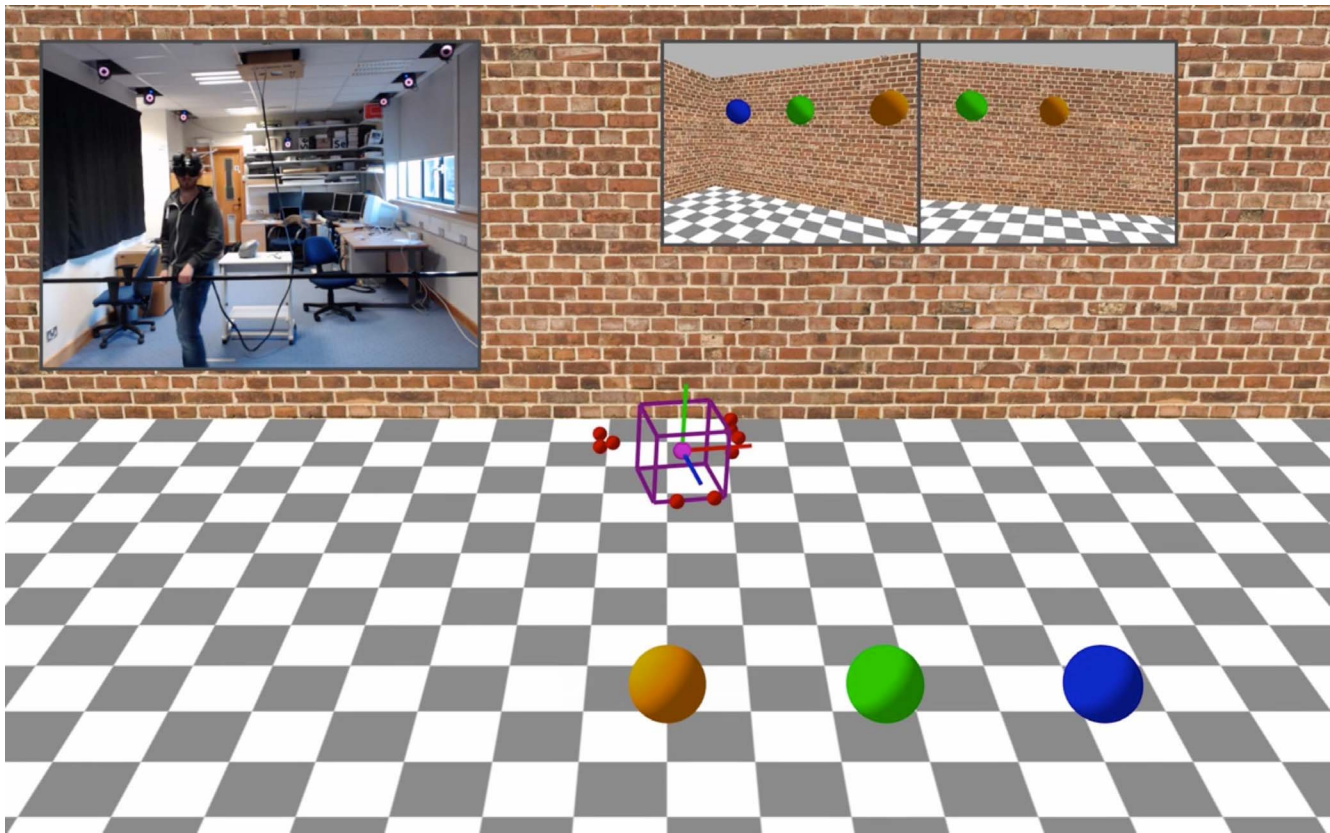
Figure 3. Same format as Figure 2; however, the viewpoint of the virtual camera is elevated to show a greater portion of the simulated environment. Three colored spheres have been placed in the simulated room, and the dual panel inset to the top right of the image shows the perspective-correct left- and right-eye images of the simulated environment, as seen by the observer. As in Figure 2, the inset video shows a person moving within the lab, the tracked coordinates of the headset driving the simulated environment in real-time. With the wider FOV, many more of the ceiling-mounted Vicon cameras can be seen. During an experiment the lab would be in darkness. A dynamic version of this Figure showing the observer walking through the virtual scene is shown in Movie 2.

Francisco, CA) and *Microsoft Kinect* (Microsoft, Redmond, WA) already exist. However, these can only track body movements over small spatial volumes, and with lower spatial and temporal precision. Interestingly, Leap motion is already being incorporated into VR headsets so that users can manipulate virtual objects with a live 3-D rendering of their hands.

Current trends in technology mean that the pixel density of screens in HMDs is increasing rapidly. Some of the most pixel-dense screens available are used in slim form factor mobile phones, which is why companies such as Oculus use these screens in their headsets. Others, such as Samsung, use the phone itself, slotted into a headset mount, as the HMD, thus exploiting the phone's high pixel-density screen, powerful processors, and built-in accelerometers and gyroscopes that can be used for tracking. The pixel density of current consumer headsets can be better than traditional HMDs, such as the NVIS SX111, but these headsets have a much-reduced FOV. Those systems with larger FOV have a reduced pixel density (e.g., Oculus Development Kit 2,

$\sim$90° FOV, $\sim$5.6 arcmin per pixel). A key requirement of our own research is a wide binocular FOV (area of the world seen by both eyes). This varies greatly over different headsets, as does the level of image distortion (e.g., barrel, pincushion, and radial distortions), and chromatic aberration (different wavelengths of light focused in different focal planes).

While the pixel density of screens used in current headsets is good compared with the CRTs they replaced, the temporal precision of these screens is far worse; this is a trend that has affected HMDs in the same way as computer monitors in general. Low refresh rates result in greater end-to-end latency, jittery motion, and motion blur caused by image persistence. Research-grade solutions exist, but these are expensive and currently limited to full-size monitors. OLEDs and other technologies look promising and are starting to be incorporated into consumer HMDs. For example, the OLED screens in the Oculus DK2 run at 75 Hz. Another benefit of OLEDs is that they have a high dynamic range. However, we are clearly quite some way from an

Figure 4. Simulated rendering of a wool scarf laying on a planar textured surface. This shows current state-of-the-art photo-realistic rendering. The image is produced using the Mistuba software renderer, developed by Wenzel Jakob (2010), and the simulation data were produced by Jonathan Kaldor (Kaldor, James, & Marschner, 2008, 2010; for further details on the rendering process see Jakob, Arbree, Moon, Bala, & Marschner, 2010) and voxelization by Manual Vargas Escalante and Manolis Savva. Both the renderer and example scene are freely available to download (http://www.mitsuba-renderer.org/download.html).

optimal screen technology for vision research, which has a high (deterministic) refresh rate, accurate color reproduction, high dynamic range, and wide color gamut. Happily, VR companies are aware of the deficiencies in current screen technology, so that in most instances the needs of researchers also align with those of consumers.

In addition to hardware, it is becoming easier to generate ray-traced photo-realistic images with free-ware packages, such as PBRT (Pharr & Humphreys, 2010), Radiance (Ruppertsberg & Bloj, 2008; Ward Larson & Shakespeare, 1998), Mitsuba (Jakob, 2010), Blender (Stichting Blender Foundation, 2014), and Rendertoolbox (Heasly, Cottaris, Lichtman, Xiao, & Brainard, 2014). Film studios, such as Pixar, produce films that are entirely computer generated and make the software they use to render these movies freely available (e.g., Renderman). Future technology will allow real-time ray tracing of photorealistic virtual environments yoked to an observer's movements. It is also becoming clear that in addition to faithfully simulating how the world looks, we also need to accurately simulate the physics of how it behaves (Battaglia, Hamrick, & Tenenbaum, 2013; Scarfe & Glennerster, 2014). This is becoming a realizable possibility for real-time VR using technologies such as PhysX and Bullet. These can be used with languages such as C/C++ and Matlab, and are built into commercial game engines.

As with any computer-generated stimuli, cue conflicts generally exist between the environment that is being simulated and the sensory information that would be generated from real-world objects. For example, focus cues from the HMD display typically conflict with the 3-D objects being simulated and can contribute to distortions in perceived depth and visual discomfort (Hoffman, Girshick, Akeley, & Banks, 2008; Watt, Akeley, Ernst, & Banks, 2005). Multiplane volumetric displays exist to provide near-correct focus cues (Akeley, Watt, Girshick, & Banks, 2004; Watt, Akeley, Girshick, & Banks, 2005), but these are too bulky to be portable at present. However, recent advances in miniature high-speed switchable lenses (Love et al., 2009) raise the prospect of eliminating cues conflicts such as these in future HMDs. Eye trackers are available for some HMDs, but currently these provide rather coarse information. Future systems will provide the high-resolution information needed for fine-scale, gaze-contingent rendering.

VR systems are also increasingly being integrated with linear and omnidirectional treadmills (Souman et al., 2011; Steinicke, Visell, Campos, & Lécuyer, 2013) and robotic motion platforms (Barnett-Cowan, Mei-linger, Vidal, Teufel, & Bulthoff, 2012). These have the benefit of simulating a totally unconstrained environment in a smaller physical space. The commercialization of systems such as these by game development companies will make these available to many more researchers. VR is already being integrated with haptic robotics to enable people to touch and interact with simulated 3-D objects, both to investigate human sensory perception (Atkins, Fiser, & Jacobs, 2001) and for purposes of rehabilitation (Harwin, Murgia, & Stokes, 2011; Jack et al., 2001; Loureiro, Harwin, Nagai, & Johnson, 2011). This trend is only set to continue, with haptics game controllers already being used in applied research (Tse et al., 2010). With the convergence of these technologies, the focus will be on achieving real-time multimodal VR.

Using VR for the study of perception: New horizons

VR has the potential to radically alter the way that vision research is carried out and even the way neuroscientists think about visual processing in the brain. The goal of visual processing is often seen as reconstructing the world, rather then simply interacting with it. Participants in experiments move around continually, so models developed to explain behavior in moving observers will inevitably be quite different from those that are applicable to an experiment in which the participant is fixed to a bite bar or immobilized in an fMRI machine. Immersive VR open up new possibilities for studying sensory information processing in much more natural conditions. This is true for the study of

vision, which has been the traditional domain for VR, as well as other sensory modalities, such as haptics. Realistic multimodal simulations of the environment are now within the scope of today's technology.

Using this technology it is possible to tightly control aspects of the stimulus in ways that would not be possible in the real world, but the observer can explore his or her environment and carry out tasks naturally, which is the key difference between classical experiments and the real world. Of course, this freedom comes with certain problems. For example, it is no longer possible to guarantee that the retinal stimulus will be identical on every trial. Nevertheless, the experimenter has a record of where the observer was during the experiment, and it is possible to recreate the visual stimulus at each moment. This leads to a different approach to analyzing data, where a method of constant stimuli is unlikely to be useful; instead, some experimental parameters will vary considerably during the course of the experiment depending on how the participant interacts with the stimulus. The key is that candidate models must include, and take account of, these parameters when seeking to explain the data.

Techniques such as this are already common in applied computer vision, where data mining is used to assess the relative likelihood of candidate models given a rich set of data (Prince, 2012). Once modeling of this type is applied to the large amount of data that is available in immersive VR experiments, the trend is likely to be towards experiments that are less and less constrained and more and more like real-world natural tasks. In some ways, in an ideal situation, observers should be unaware that they are in an experiment; they could play interactive games, or be absorbed in a chase, without realizing that their movements and responses are being monitored, recorded, and analyzed. The experiment would no longer "get in the way" of what the person is doing, and we would begin to study how sensory systems respond naturally in everyday life, rather than in the contexts of the experiment itself. The day participants queue up to do our experiments, we will know that psychophysics research has really moved on.

Keywords: virtual reality, perception, stereopsis, 3-D vision

Acknowledgments

Commercial relationships: none.
Corresponding author: Peter Scarfe.
Email: p.scarfe@reading.ac.uk.
Address: Department of Psychology, University of Reading, Reading, Berkshire, UK.

Footnote

[1] Moore's law refers to the fact that the number of transistors on an integrated circuit, and thus a measure of computing power, doubles, roughly, every 2 years.

References

Akeley, K., Watt, S. J., Girshick, A. R., & Banks, M. S. (2004). A stereo display prototype with multiple focal distances. *ACM Transactions on Graphics*, *23*(3), 804–813.

Atkins, J. E., Fiser, J., & Jacobs, R. A. (2001). Experience-dependent visual cue integration based on consistencies between visual and haptic percepts. *Vision Research*, *41*(4), 449–461.

Bailenson, J. N., & Yee, N. (2005). Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological Science*, *16*(10), 814–819, doi:10.1111/j.1467-9280.2005.01619.x.

Barnett-Cowan, M., Meilinger, T., Vidal, M., Teufel, H., & Bulthoff, H. H. (2012). MPI CyberMotion Simulator: Implementation of a novel motion simulator to investigate multisensory path integration in three dimensions. *Journal of Visualized Experiments*, *63*, e3436, doi:10.3791/3436.

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences, USA*, *110*(45), 18327–18332, doi:10.1073/pnas.1306572110.

Brenner, E., & Landy, M. S. (1999). Interaction between the perceived shape of two objects. *Vision Research*, *39*(23), 3834–3848.

Bruggeman, H., Zosh, W., & Warren, W. H. (2007). Optic flow drives human visuo-locomotor adaptation. *Current Biology*, *17*(23), 2035–2040, doi:10.1016/j.cub.2007.10.059.

Clark, A. (1997). *Being there: Putting brain, body and world together again*. Cambridge, MA: MIT Press.

Clark, A. (2008). *Supersizing the mind: Embodiment, action and cognitive extension*. Oxford: Oxford University Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral & Brain Sciences*, *36*(3), 181–204, doi:10.1017/S0140525x12000477.

Creem-Regehr, S. H., Willemsen, P., Gooch, A. A., & Thompson, W. B. (2005). The influence of

restricted viewing conditions on egocentric distance perception: Implications for real and virtual indoor environments. *Perception*, *34*(2), 191–204, doi:10.1068/P5144.

De Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Science*, *7*(10), 460–467.

Ellis, S. R. (1991). Nature and origins of virtual environments: A bibliographical essay. *Computing Systems in Engineering*, *2*(4), 321–347.

Felsen, G., & Dan, Y. (2005). A natural approach to studying vision. *Nature Neuroscience*, *8*(12), 1643–1646, doi:10.1038/nn1608.

Foo, P., Warren, W. H., Duchon, A., & Tarr, M. J. (2005). Do humans integrate routes into a cognitive map? Map- versus landmark-based navigation of novel shortcuts. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *31*(2), 195–215, doi:10.1037/0278-7393.31.2.195.

Fox, J., Arena, D., & Bailenson, J. N. (2009). Virtual reality: A survival guide for the social scientist. *Journal of Media Psychology*, *21*(3), 95–113.

Gibson, J. J. (1950). *The perception of the visual world*. Boston: Houghton Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Gilson, S. J., Fitzgibbon, A. W., & Glennerster, A. (2006). Quantitative analysis of accuracy of an inertial/acoustic 6DOF tracking system in motion. *Journal of Neuroscience Methods*, *154*(1–2), 175–182, doi:10.1016/j.jneumeth.2005.12.013.

Gilson, S. J., Fitzgibbon, A. W., & Glennerster, A. (2011). An automated calibration method for non-see-through head mounted displays. *Journal of Neuroscience Methods*, *199*(2), 328–335, doi:10.1016/j.jneumeth.2011.05.011.

Gilson, S. J., & Glennerster, A. (2012). High fidelity virtual reality. In X. Tang (Ed.), *Virtual reality: Human computer interaction*. Rijeka, Croatia: InTech.

Glennerster, A., Tcheang, L., Gilson, S. J., Fitzgibbon, A. W., & Parker, A. J. (2006). Humans ignore motion and stereo cues in favor of a fictional stable world. *Current Biology*, *16*(4), 428–432.

Harwin, W. S., Murgia, A., & Stokes, E. K. (2011). Assessing the effectiveness of robot facilitated neurorehabilitation for relearning motor skills following a stroke. *Medical & Biological Engineering & Computing*, *49*(10), 1093–1102, doi:10.1007/s11517-011-0799-y.

Heasly, B. S., Cottaris, N. P., Lichtman, D. P., Xiao, B., & Brainard, D. H. (2014). RenderToolbox3: MATLAB tools that facilitate physically based stimulus rendering for vision research. *Journal of Vision*, *14*(2):6, 1–22, doi:10.1167/14.2.6. [PubMed] [Article]

Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. (2008). Vergence accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, *8*(3):33, 1–30, doi:10.1167/8.3.33. [PubMed] [Article]

Jack, D., Boian, R., Merians, A. S., Tremaine, M., Burdea, G. C., Adamovich, S. V., & Poizner, H. (2001). Virtual reality-enhanced stroke rehabilitation. *IEEE Trans on Neural Systems & Rehabilitation Engineering*, *9*(3), 308–318, doi:10.1109/7333.948460.

Jain, A., & Backus, B. T. (2010). Experience affects the use of ego-motion signals during 3D shape perception. *Journal of Vision*, *10*(14):30, 1–14, doi:10.1167/10.14.30. [PubMed] [Article]

Jakob, W. (2010). Mitsuba Renderer. Available from http://www.mitsuba-renderer.org/

Jakob, W., Arbree, A., Moon, J. T., Bala, K., & Marschner, S. (2010). *A radiative transfer framework for rendering materials with anisotropic structure. Proceedings of SIGGRAPH 2010*, *29*(4), Paper presented at the ACM Transactions on Graphics, Los Angeles, 2010.

Johnston, E. B. (1991). Systematic distortions of shape from stereopsis. *Vision Research*, *31*(7–8), 1351–1360.

Kaldor, J. M., James, D. L., & Marschner, S. (2008). Simulating knitted cloth at the yarn level. *Proceedings of SIGGRAPH 2010*, *27*(3), Paper presented at the ACM Transactions on Graphics, Los Angeles, 2008.

Kaldor, J. M., James, D. L., & Marschner, S. (2010). Efficient yarn-based cloth with adaptive contact linearization. *Proceedings of SIGGRAPH*, *29*(4), Paper presented at the ACM Transactions on Graphics, Los Angeles, 2010.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*, 14–14.

Knapp, J. M., & Loomis, J. M. (2004). Limited field of view of head-mounted displays is not the cause of distance underestimation in virtual environments. *Presence: Teleoperators and Virtual Environments*, *13*(5), 572–577, doi:10.1162/1054746042545238.

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*(11), 1311–1328.

Loureiro, R. C., Harwin, W. S., Nagai, K., & Johnson, M. (2011). Advances in upper limb stroke rehabilitation: A technology push. *Medical & Biological Engineering & Computing, 49*(10), 1103–1118, doi: 10.1007/s11517-011-0797-0.

Love, G. D., Hoffman, D. M., Hands, P. J., Gao, J., Kirby, A. K., & Banks, M. S. (2009). High-speed switchable lens enables the development of a volumetric stereoscopic display. *Optics Express, 17*(18), 15716–15725, doi:10.1364/OE.17.015716.

Messing, R., & Durgin, F. H. (2005). Distance perception and the visual horizon in head-mounted displays. *ACM Transactions on Applied Perception, 2*(3), 234–250.

Mon-Williams, M., & Bingham, G. P. (2008). Ontological issues in distance perception: Cue use under full cue conditions cannot be inferred from use under controlled conditions. *Perception & Psychophysics, 70*(3), 551–561.

Mon-Williams, M., Plooy, A., Burgess-Limerick, R., & Wann, J. (1998). Gaze angle: A possible mechanism of visual stress in virtual reality headsets. *Ergonomics, 41*(3), 280–285, doi:10.1080/001401398187035.

Mon-Williams, M., Wann, J. P., & Rushton, S. (1993). Binocular vision in a virtual world: Visual deficits following the wearing of a head-mounted display. *Ophthalmic & Physiological Optics, 13*(4), 387–391.

O'Regan, J. K., & Noe, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral & Brain Sciences, 24*(5), 939–1031.

Pharr, M., & Humphreys, C. (2010). *Physically based rendering: From theory to implementation* (2nd ed.). Burlington, MA: Morgan Kaufmann.

Pickup, L. C., Fitzgibbon, A. W., & Glennerster, A. (2013). Modelling human visual navigation using multi-view scene reconstruction. *Biological Cybernetics, 107*(4), 449–464, doi:10.1007/s00422-013-0558-2.

Prince, S. J. D. (2012). *Computer vision: Models, learning, and inference.* New York: Cambridge University Press.

Rizzo, A., & Kim, G. J. (2005). A SWOT analysis of the field of virtual reality rehabilitation and therapy. *Presence: Teleoperators & Virtual Environments, 14*(2), 119–146, doi:10.1162/1054746053967094.

Rosenberg, R. S., Baughman, S. L., & Bailenson, J. N. (2013). Virtual superheroes: Using superpowers in virtual reality to encourage prosocial behavior. *PloS One, 8*(1), e55003, doi:10.1371/journal.pone.0055003.

Ruppertsberg, A. I., & Bloj, M. (2008). Creating physically accurate visual stimuli for free: Spectral rendering with RADIANCE. *Behavior Research Methods, 40*(1), 304–308.

Rust, N. C., & Movshon, J. A. (2005). In praise of artifice. *Nature Neuroscience, 8*(12), 1647–1650, doi:10.1038/nn1606.

Scarfe, P., & Glennerster, A. (2014). Humans use predictive kinematic models to calibrate visual cues to three-dimensional surface slant. *Journal of Neuroscience, 34*(31), 10394–10401, doi:10.1523/JNEUROSCI.1000-14.2014.

Scarfe, P., & Hibbard, P. B. (2013). Reverse correlation reveals how observers sample visual information when estimating three-dimensional shape. *Vision Research, 86,* 115–127, doi:10.1016/j.visres.2013.04.016

Schnapp, B., & Warren, W. (2007). Wormholes in virtual reality: What spatial knowledge is learned from navigation. *Journal of Vision, 7*(9):758, doi:10.1167/7.9.758. [Abstract]

Schwaiger, M., Thummel, T., & Ulbrich, H. (2007). Cyberwalk: Implementation of a ball bearing platform for humans. *Human–Computer Interaction, 4551,* 926–935.

Shapiro, L. (2011). *Embodied cognition.* Abingdon, Oxen: Routledge.

Sharkey, P. (2014). In E. A. Keshner, M. Levin, & T. Weiss (Eds.), *Virtual reality technologies for health and clinical applications: Physical and motor rehabilitation* (Vol. 1). Berlin, Heidelberg: Springer.

Slater, M., Spanlang, B., Sanchez-Vives, M. V., & Blanke, O. (2010). First person experience of body transfer in virtual reality. *PloS One, 5*(5), e10564, doi:10.1371/journal.pone.0010564.

Souman, J. L., Giordano, P. R., Schwaiger, M., Frissen, I., Thummel, T., Ulbrich, H., & Ernst, M. O. (2011). CyberWalk: Enabling unconstrained omnidirectional walking through virtual environments. *ACM Transactions on Applied Perception, 8*(4), 1–22, doi:10.1145/2043603.2043607.

Steinicke, F., Visell, Y., Campos, J., & Lécuyer, A. (2013). *Human walking in virtual environments: Perception, technology and applications.* Heidelberg, Germany: Springer.

Stichting Blender Foundation. (2014). Blender. Available from http://www.blender.org

Sutherland, I. E. (1968). *A head-mounted three dimensional display.* Paper presented at the Proceedings of the American Federation of Information Processing Societies (AFIPS, 1968). New York.

Svarverud, E., Gilson, S., & Glennerster, A. (2012). A demonstration of 'broken' visual space. *PloS One*, *7*(3), e33782, doi:10.1371/journal.pone.0033782.

Svarverud, E., Gilson, S. J., & Glennerster, A. (2010). Cue combination for 3D location judgements. *Journal of Vision*, *10*(1):5, 1–13, doi:10.1167/10.1.5. [PubMed] [Article]

Tarr, M. J., & Warren, W. H. (2002). Virtual reality in behavioral neuroscience and beyond. *Nature Neuroscience, 5*(Suppl.), 1089–1092, doi:10.1038/nn948.

Tcheang, L., Gilson, S. J., & Glennerster, A. (2005). Systematic distortions of perceptual stability investigated using immersive virtual reality. *Vision Research*, *45*(16), 2177–2189.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.

Todd, J. T., Chen, L., & Norman, J. F. (1998). On the relative salience of Euclidean, affine, and topological structure for 3-D form discrimination. *Perception*, *27*(3), 273–282.

Todd, J. T., Christensen, J. C., & Guckes, K. M. (2010). Are discrimination thresholds a valid measure of variance for judgments of slant from texture? *Journal of Vision*, *10*(2):20, 1–18, doi:10.1167/10.2.20. [PubMed] [Article]

Todd, J. T., & Norman, J. F. (2003). The visual perception of 3-D shape from multiple cues: Are observers capable of perceiving metric structure? *Perception & Psychophysics*, *65*(1), 31–47.

Todd, J. T., Tittle, J. S., & Norman, J. F. (1995). Distortions of 3-dimensional space in the perceptual analysis of motion and stereo. *Perception*, *24*(1), 75–86.

Tse, B., Harwin, W., Barrow, A., Quinn, B., Diego, J. S., & Cox, M. (2010). Design and development of a haptic dental training system: hapTEL. *Haptics: Generating and Perceiving Tangible Sensations*, *6192*, 101–108.

van Boxtel, J. J., Wexler, M., & Droulez, J. (2003). Perception of plane orientation from self-generated and passively observed optic flow. *Journal of Vision*, *3*(5):1, 318–332, doi:10:1167/3.5.1. [PubMed] [Article]

Varela, F. J., Thompson, E., and Rosch, E. (1991). *The embodied mind*. Cambridge, MA: MIT Press.

Wann, J. P., Rushton, S., & Mon-Williams, M. (1995). Natural problems for stereoscopic depth perception in virtual environments. *Vision Research*, *35*(19), 2731–2736.

Ward Larson, G., & Shakespeare, R. (1998). *Rendering with radiance: The art and science of lighting visualization*. San Francisco: Morgan Kaufmann.

Watt, S. J., Akeley, K., Ernst, M. O., & Banks, M. S. (2005). Focus cues affect perceived depth. *Journal of Vision*, *5*(10):7, 834–862, doi:10.1167/5.10.7. [PubMed] [Article]

Watt, S. J., Akeley, K., Girshick, A. R., and Banks, M. S. (2005). Achieving near-correct focus cues in a 3-D display using multiple image planes. In B. E. Rogowitz, T. N. Pappas, & S. J. Daly (Eds.), *Proceedings of SPIE: Human vision and electronic Imaging* (pp. 5666–5673).

Wexler, M., Panerai, F., Lamouret, I., & Droulez, J. (2001). Self-motion and the perception of stationary objects. *Nature*, *409*(6816), 85–88, doi:10.1038/35051081.

Wexler, M., & van Boxtel, J. J. (2005). Depth perception by the active observer. *Trends in Cognitive Science*, *9*(9), 431–438, doi:10.1016/j.tics.2005.06.018.

Zabulis, X., & Backus, B. T. (2004). Starry night: A texture devoid of depth cues. *Journal of the Optical Society of America, A: Optics, Image Science, & Vision*, *21*(11), 2049–2060.