Virtual Reality for Vision Science

Accepted for publication in Christopher Maymon, Gina Grimshaw, Ying Choon Wu (eds.) 2023. Cham. Virtual Reality in Behavioral Neuroscience: New Insights and Methods. Springer.

Virtual Reality for Vision Science

Paul B. Hibbard

Abstract

Virtual reality (VR) allows us to create visual stimuli that are both immersive and reactive. VR provides many new opportunities in vision science. In particular, it allows us to present wide field-of-view, immersive visual stimuli; for observers to actively explore the environments that we create; and for us to understand how visual information is used in the control of behaviour. In contrast with traditional psychophysical experiments, VR provides much greater flexibility in creating environments and tasks that are more closely aligned with our everyday experience. These benefits of VR are of particular value in developing our theories of the behavioural goals of the visual system and explaining how visual information is processed to achieve these goals. The use of VR in vision science presents a number of technical challenges, relating to how the available software and hardware limit our ability to accurately specify the visual information that defines our virtual environments, and the interpretation of data gathered in experiments with a freely-moving observer in a responsive environment.

1. Introduction

This chapter discusses the significant opportunities for advancing our understanding of vision made possible by virtual reality (VR). It presents the benefits and drawbacks of using VR to study visual perception, focussing on the current state of the art, but also on the inherent differences between VR and other display technologies and methodological approaches. Since VR can mean many different things in different contexts, I begin with a working definition of VR for the purposes of this discussion. Lanier (2017) provides many thought-provoking definitions, some of which are especially relevant for our purposes. In particular, his fourth definition provides the most useful framing of VR for exploring its role as a tool for vision science:

> *The substitution of the interface between a person and the physical environment with an interface to a simulated environment*
>
> (Lanier, 2017, page 47).

In practice, I define the visual component of VR as a display that is both *immersive* and *responsive*. The typical example hardware is a head-mounted display, although many of the same principles and practical considerations apply to other implementations, such as a CAVE, in which the user is surrounded by large display screens at a distance of several metres (Cruz-Neira, Sandin, DeFanti, Kenyon, & Hart, 1992). A head-mounted display provides a binocular, wide field-of-view screen which displays the virtual environment, while excluding any visual input from the real physical environment. This is combined with 6 degrees-of-freedom motion tracking of the headset, so that its position and orientation in 3D space are known and can be used to update the images presented to the observer in response to their movement (figure 1). In addition, tracking of the observer's hand positions, or more extensive position tracking of their body, are required to produce an environment and display that update in response to the user's actions. Together, these properties can be used to create an environment that is immersive (providing the optic array of the virtual environment, while excluding the optic array of the physical environment) and responsive (changing in near-to-real time in response to the user's actions).

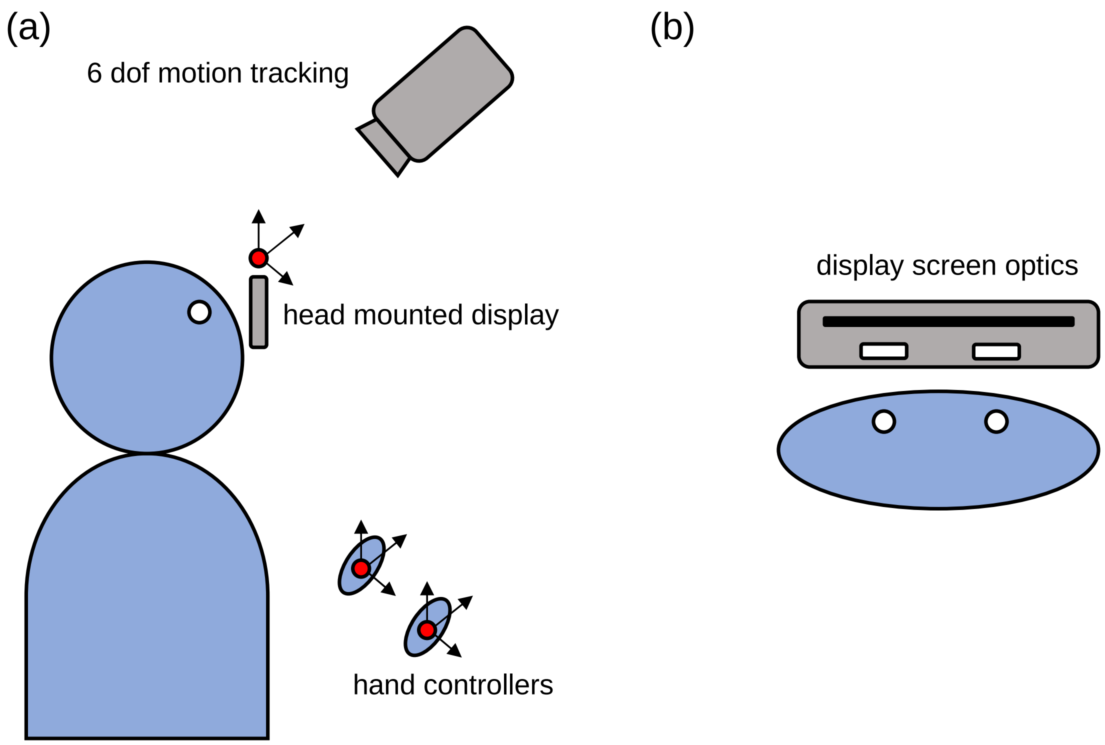

Figure 1. *The essential components of VR for visual display. (a) The observer wears a head-mounted display, and typically interacts with the environment via two hand-held controllers. The motion of the headset, and the hand controllers, is tracked to provide 6 degrees-of-freedom estimates of their location and orientation in 3D space. (b) Within the head-mounted display, the display screen is placed close to the observer's eyes. This creates a wide, binocular field of view, and occludes the view of the outside world. The screen optics ensure that the image is formed at a comfortable distance.*

Head-mounted VR systems achieve this immersive experience by combining a number of key components. First, the virtual environment must be created. This may be built in a games engine or other software platform, and defines the components of the environment and their behaviour. These include, for example, the objects and other characters in the environment, their behaviour in response to simulated physics, and possible interactions with the user. It also includes components such as lighting and atmospheric effects that determine the appearance of the environment for the user. As the user is a part of the environment, tracking of their own motion is also required so that their view of the environment can be updated, and to allow interactions such as picking up and moving objects. The observer's view of the environment is then rendered and presented on binocular display screens with a wide field of view. As these screens are placed very close to the user's eyes, the headset also includes optics that allow the displayed images to be clearly and comfortably perceived. This creates an effective image plane at a fixed distance of a metre or so from the observer.

With this definition in mind, we can explore the novel opportunities provided by VR for the scientific study of our sense of vision (Scarfe & Glennerster, 2015, 2019). In order to do this, I first outline vision as a sense, and vision science as a way of understanding this sense. I illustrate how VR provides opportunities to change how we can study vision, expanding both the types of experiments we can undertake, and the theoretical understanding that follows from these. I will illustrate this with examples of recent studies, outline the technical challenges and possibilities associated with VR, and provide my own outlook on the opportunities that it provides for a different kind of vision science.

2. What is vision, and how do we study it scientifically?

2.1 The ambient optic array

The starting point for understanding vision is the ambient optic array, the structured arrangement of light with respect to a point of observation (Gibson, 1979). This is determined by the physical structure of the environment: the visible objects, surfaces and materials, the light sources, and the location of the observation point (figure 2). For each direction from the observation point, the spectral intensity of light depends on the illumination and reflectance properties of the physical structures along that line of sight. As the observer moves, the ambient optic array is sampled from different locations. This definition of the starting point of vision does not depend on any properties of the observer, but is a definition of the information available from the world in the form of the structured array of light. In particular, the optics of the eye or camera, and the imaging surface, which together may be involved in creating a focussed image from the optic array, are not part of the definition (Rogers, 2021). As most people have two eyes, we sample the ambient optic array from two points simultaneously. As we move around the environment, the locations from which it is sampled change.

2.2 The retinal image

Each of our eyes contains a lens and a cornea, which together focus the light to create an image from the optic array on the retinal surface. This means that, at each moment in time, the two retinal images provide us with a partial sample of the ambient optic array from two locations. The information in each image can be defined as a function of visual direction, rather than the location on the retina to which that direction projects. The visual direction to each visible point is defined as the orientation of the line connecting that point to the centre of the eye. Since there is a one-to-one mapping between visual direction and retinal location, each point on the retina can be associated with a single visual direction (Koenderink, 1984; Rogers, 2021).

In creating a visual display for VR, our goal is to position a display screen so as to create a desired pair of retinal images, and thus a pair of samples of the ambient optic array that correspond not to the physical surroundings of the user, but to the virtual environment that we wish them to experience. VR thus provides us with the ability to create the experience of whatever virtual visual environment we want. From the perspective of vision science, it provides a means of experimentally controlling the ambient optic array of the observer. This control of the entirety of the visual input, in a way that mimics the ambient optic array in the physical world, and its ability to change in response to the movement and actions of the observer, differ fundamentally from the more restricted opportunities available on more traditional display screens with a fixed location relative to the observer.

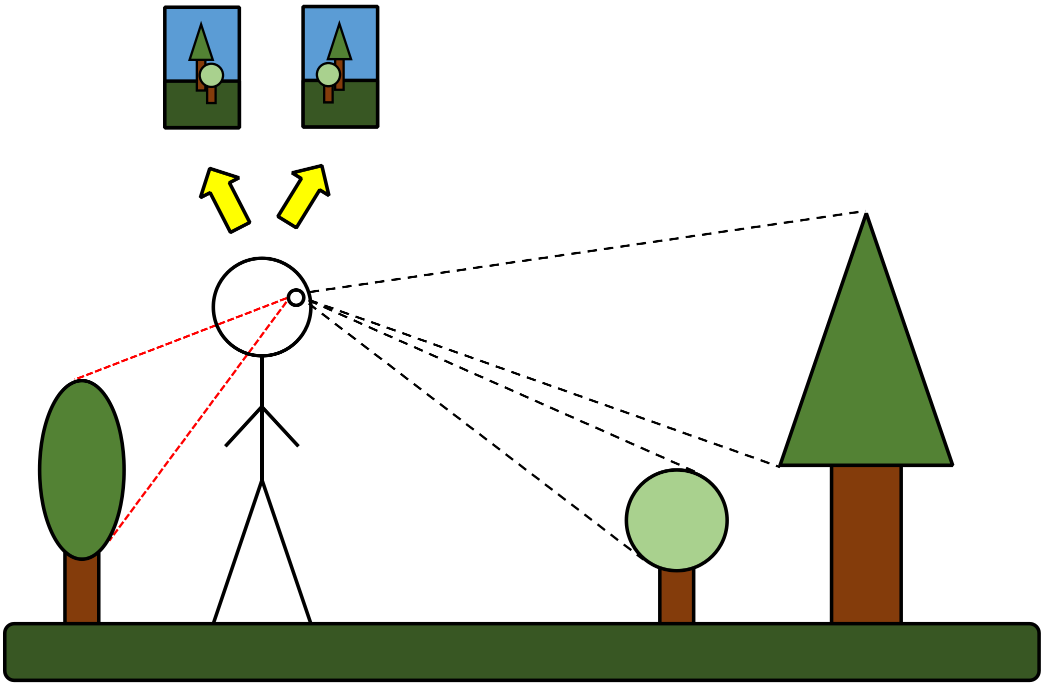

Figure 2. *The ambient optic array and the retinal images. The ambient optic array is defined at each point in space. In this figure the optic array is shown from the observer's point of view. This includes visual information arriving at this point from all locations. However, the viewer can only see those components lying within their field of view (such as the dashed black lines) but not points lying outside this (such as the dotted red lines). The projection of the optic array through the observer's eyes creates the two retinal images.*

2.3 The visual system

The human visual system consists of the eyes and a large number of cortical and subcortical areas in the brain. This system includes the photoreceptors in the retina which perform the task of transduction, creating a neural response to the incoming stimulus; the optic nerve, transmitting this signal to the brain; and the areas of the brain involved in visual processing. It is the task of vision science to understand the role of the visual system in making use of the information provided by the optic array.

2.4 What is vision science?

This is the scientific study of visual perception. It assesses the processes by which we use visual information, how vision influences and affects our behaviour and experience, and the means by which this is achieved. While this is a broad undertaking, requiring multiple, complementary approaches, in its simplest terms it provides three types of understanding (Koenderink, 2019). The first is the physiology of the way that the visual system responds to light. The second is how this information is used to control our behaviour. The third is the phenomenological experience of seeing.

2.4.1 Responses of the visual system to the ambient optic array

This endeavour covers topics such as the spectral response properties of retinal photoreceptors (Schnapf, Kraft, Nunn, & Baylor, 1988); the receptive fields of neurons in the lateral geniculate nucleus (Cleland, Dubin & Levick, 1971) and primary visual cortex (Hubel & Wiesel, 1962); the retinotopic maps formed in the classical visual areas (Zeki, 1978); nonlinear aspects of visual responses (Wilson, 1980; Heeger, 1992); the response properties of higher level cortical areas, such as the fusiform face area (Kanwisher and Yovel, 2006); and the organisation of visual areas into visual processing streams (Livingstone and Hubel, 1987; Milner and Goodale, 2006). It also covers our more abstract conceptualisations such as filters and channels (Braddick, Campbell and Atkinson, 1978; Graham, 1989; Zhaoping, 2014), and the relationship between the responses of populations of neurons and our perceptual judgments (Parker and Newsome, 1998).

2.4.2 The use of visual information to control behaviour

Vision provides us with information that allows us to make better decisions, and to execute actions with skill and dexterity. It allows us to recognise objects and people; to understand the emotional state of others, to recognise them as individuals, and to interact with them socially; to move around our environment safely, and to pick up and manipulate objects. To understand the role of vision therefore is to provide an account not only of how information is detected, encoded, and transmitted, but also of how it is employed in all these ways and more. These tasks and behaviours themselves need to be part of our theories.

2.4.3 The phenomenological experience of seeing

The most obvious fact of vision, to a sighted observer, is not the physiological responses of their visual system, nor the importance of these responses in enacting behaviour, but the immediate conscious experience of seeing that places them in the here and now. As with understanding the use of vision in the control of behaviour, virtual reality provides the opportunity for the controlled, measurable assessment of visual experience as it relates to our everyday behaviour. The potential for a traditional psychophysical experiment to provide generalisable insights on this topic are limited. The full reality of our perception, as individuals inhabiting and immersed in a three-dimensional world cannot be extrapolated from experiments in which immobilised observers make forced-choice responses about the nature of abstract patterns presented on a display screen.

3. The role of display technology in shaping vision science

As our available display technologies have developed, the kinds of empirical understanding that are possible from laboratory-based experiments have grown. In understanding the opportunities afforded by virtual reality, it is instructive to consider it alongside other technologies. It is also important to remember that VR does not make these other technologies obsolete; rather, it adds to the kinds of experiments that it is possible to undertake.

Prior to the ubiquitous adoption of computer generated, screen-presented stimuli in vision research, experiments typically used optical apparatus to create and present stimuli (Koenderink, 1999). These would include light sources, prisms, mirrors, beam-splitters, filters and such like. These would be arranged with great precision using optical benches and rails, to present the stimulus to an observer whose head position was held firmly in place using a bite-bar. Oscilloscopes, and then graphics cards and computer monitors, offered increased flexibility, complexity and ease with which images could be created and displayed. This means that, if an image can be specified as an array of numbers representing the intensity in the red, green, and blue channels, it can be displayed, subject to the resolution and dynamic range limits of the screen (Pelli,1997). This has been especially facilitated by the development of software platforms that take care of the interactions with the computer hardware (Brainard, 1997; Peirce, 2007), allowing the experimenter to concentrate on the specification of the stimuli. These software and hardware technologies have greatly facilitated the creation and display of stimuli. These stimuli can be ordered in terms of the degree to which they take account of the observer's point of view.

3.1. Simple visual patterns

Some visual stimuli are defined with no reference to a real or imagined object beyond the retinal image. They exist purely as visual patterns on the screen, and examples include random dot patterns (Julesz, 1971, Braddick, 1974), gratings (Braddick, 1981), plaids (Adelson & Movshon, 1982), and Gabor patches (Daugman, 1985). They are not intended to be interpreted as projections of physical objects, but are designed to test the way in which visual information is encoded and used for perceptual judgements. They are used to address questions such as our acuity to spatial alignment or orientation (Westheimer, 1972; Heeley, Buchanan-Smith, Cromwell , & Wright, 1997), whether shape and depth can be perceived purely on the basis of binocular disparities (Julesz, 1971), or how neural responses are interpreted to determine the perceived direction of motion of an image (Adelson & Movhson, 1982, Newsome, Britten, & Movshon, 1989).

3.2. Pictures of physical objects

The same display devices can also present stimuli that are images of real or fictional objects, scenes, or people. This can be achieved either through the use of photographs or videos captured using a camera, or through the rendering of stimuli intended to be images of objects and scenes with specified physical properties. Creating such stimuli requires an understanding of, and ability to implement, the properties of projective geometry, lighting, and image formation.

If we assume the location of the optical point of observation within the scene, then it is a simple process to determine the visual direction from that point to any 3D location in space. An example here could be a stereoscopically presented stimulus in which the left and right images are intended as the views of an object with a particular 3D location, shape, and size. Each point on the object is projected onto the appropriate location in the left and right images, according to its defined location relative to the optical centres of the left and right eye (Johnston, 1991; Johnston, Cumming & Parker, 1997).

In creating these images, we need to specify the location of the observer, and for them then to be presented so that the retinal image faithfully reflects the experience of viewing the scene from that position. This requires accurate positioning of the observer's eyes, and spatial calibration of the display screen. If this is not done, then we create a situation in which the observer is looking at a picture in which their point of observation and that used in creating the stimuli do not necessarily agree. It has been observed that much of vision science, through using stimuli presented on 2D display screens, is a science of how we see pictures, and not of how we see real objects in the three-dimensional world (Wade, 2013).

3.3. The ambient optic array sampled from a specific viewpoint

It is however possible to position the observer and the images so that they accurately recreate the retinal images that would be experienced from a particular 3D scene viewed from the observer's location. This can be achieved with great precision using a bite bar to keep the observer's head fixed, and a sighting device to ensure that this fixed location is in the correct position relative to the screen. Once this position is known, a mapping between visual directions from the optical centres and pixel locations on the screen can be determined, for example by positioning a physical grid in front of the displays. This mapping can then be used to accurately transform the images and position them on the display screens (Backus, Banks, van Ee & Crowell, 1999). If photographs are used as stimuli, it is also necessary to calibrate the camera images (Hibbard, 2008). If all the appropriate steps are followed, then it is possible to create stimuli that form known retinal images for the observer, replicating the images that would be formed in an equivalent three-dimensional scene. This approach is used for example in studies of how binocular information is used in the perception of 3D structure (Backus et al., 1999; Hillis, Watt, Landy & Banks, 2004; Watt, Akeley, Ernst & Banks, 2005).

3.4. The ambient optic array of a freely moving observer, interacting with a 3-dimensional world

The most direct way in which this can be achieved, of course, is through conducting experiments directly in the real world using three-dimensional scenes and objects as stimuli (Gibson, 1950). This approach is very uncommon, save for some notable and valuable exceptions. Studies of inattention blindness have for example used real, staged environments, in which actors may be swapped during interactions with participants, demonstrating a surprising lack of awareness of this change (Simons & Levin, 1998). Studies of the accuracy of distance perception have also been performed extensively outside of the laboratory (Plumert, Kearney, Cremer & Recker, 2005). Research in the natural environment using mobile eyetracking also provides an understanding of how observers sample the optic array when performing everyday tasks (Foulsham, Walker & Kingstone, 2011). Such studies are valuable since they provide evidence of how we see, and act, in the natural environment, rather than in response to pictures on a display screen (Koenderink, 1999; Wade, 2013).

Experiments in the natural environment present a variety of practical challenges, however. The first is that they are very labour intensive. The situation must be set up in the same way for each participant. This requires the actors, the props, and the interactions between these and the participant to be arranged and controlled as similarly as is possible on every repetition. Even if this can be achieved, there will always be a limit to the amount of experimental control that is possible. However carefully the scene is set, there will always be some differences which cannot be controlled. It can therefore be very difficult to replicate these studies with any precision. Another problem is that the data needed to fully describe the trial – the shape, size, and location of all objects in the scene, and the movement of the participant and other people throughout the trial, will typically not be recorded in fine detail, even if video and audio recordings of the scene are made.

In some experimental designs, it is possible to use quite complex natural visual stimuli, while varying only a single parameter of interest. This provides some of the advantages of research in natural environments, although without the ability to study natural behaviour. For example, in studies of the perception of distance it is possible to vary distance specified by convergence while keeping all other visual information unaltered (Tresilian, Mon-Williams and Kelly, 1999; Mon-Williams, Tresilian and Roberts, 2000). This has the advantage of providing precise experimental control while studying perception in a full-cue, natural environment rather than using typically sparse laboratory stimuli – thus at a natural ‘operating point’ (Koenderink, 1998). It is limited, however, in that while the visual stimuli may contain all the complexity of a typical scene, the tasks used in experiments are highly constrained.

3.5. The ambient optic array of a freely moving observer, interacting with a 3-dimensional world

Many of the advantages of performing experiments in the natural environment can be achieved in VR, whilst maintaining greater control over the stimuli presented and the data collected. VR thus provides an ideal environment in which to undertake experiments involving a freely moving observer interacting with a 3-dimensional world (Scarfe and Glennerster, 2015). In comparison with complex real-world stimuli, the major advantages of VR are control, automation, and data capture. The exact same environment can be presented on each trial, specified by the experimenter down to the precise location of every point on every object, the nature of the lighting, and the behaviour of all the components. Once created, no further input from the experimenter is required. In addition, a detailed description of the scene, and the user's movements, can be recorded. This recording can specify the virtual 3D scene, the projected 2D images, and the relationship between the two. The latter, for example, means that we have access to the ground truth associated with each pixel in the display, so that properties such as the 3D location to which that point corresponds and the object to which it belongs can all be recorded (Goutcher, Barrington, Hibbard and Graham, 2021).

Some of the challenges of conducting experiments in the real world remain when they are performed in virtual reality, however. For example, because the stimulus depends on the user's motion within - and interaction with - the environment, each trial will be different for every observer, and different on every repetition, even when the specified environment is identical.

4. New tasks for new stimuli

Experiments in vision science are defined not just by the kinds of stimuli that are presented, but also by the tasks set for the observer, and the behavioural data that are collected. Perhaps the simplest example is the forced choice discrimination task, in which the observer is presented with two stimuli and asked to judge whether they are the same or different. Brindley (1960) defined this as a 'type A' psychophysical task (Brindley, 1960). Type A tasks have the strong theoretical advantage of a clear linking hypothesis which allows us to make inferences about the physiological responses to the stimuli presented (Brindley, 1960; Morgan, Melmouth and Solomon, 2013). If an observer can discriminate reliably below two stimuli in a type A task, then there must be some difference in the way that the brain responds to them. Other psychophysical tasks ('type B' tasks) require the observer to report on the appearance of the stimulus (for example its orientation or direction of motion) and thus depend on the observer's judgement criteria as well as the neural encoding of the stimulus.

For type B tasks, there is often an agreed 'correct' answer. When using an abstract visual pattern, this will be a property of the image itself, for example the orientation of the grating, or the direction of motion of the random dot pattern. In other cases, whether or not a response is correct depends on both the observer's judgment and the experimenter's definition of the distal stimulus of which the proximal retinal images are projections. For example, the experimenter determines that a particular size and direction of binocular disparity is the projection onto the retinas of a particular depth separation (Johnston, 1991), that an image texture results from a particular slant and tilt of a planar surface (Hillis et al., 2004), or the 'veridical' 3D shape depicted in a single 2D image (Pizlo, 2010). In all these cases, the 'experimenter's share' as well as the beholder's share (Koenderink, van Doorn & Kappers, 2001) and the actual stimulus presented are all essential components determining whether or not a response is correct, or the way in which it is wrong.

Beyond forced-choice psychophysical experiments, researchers may also seek to understand how visual information is used in the control of behaviour, such as maintaining a desired direction of heading (Warren & Hannon, 1988; Rushton, Harris, Lloyd and Wann, 1998), or reaching to pick up an object (Servos, Goodale & Jakobson, 1992; Watt & Bradshaw, 2000; Bradshaw et al., 2004; Melmoth & Grant, 2006). In these cases, the experimental data are the measures of the observer's behaviour itself, and the interpretation of the data can depend on a number of theoretical assumptions. For example, in studies of reaching and grasping, the maximum speed of motion and the maximum size of the grip aperture are taken as measures of the apparent distance and size of the object to be picked up, respectively (Servos et al., 1992).

When interpreting such natural behavioural responses, as opposed to forced choice categorisations, it is important to appreciate that they are not direct measures of 'pure' perception, but are also determined by the requirements of the action to be completed. For example, peak speed of motion and peak grip aperture will incorporate margins of error, dependent on the participant's confidence in their perceptual estimates, so as not to bump into or knock over the target (Keefe, Suray, & Watt, 2019). There is thus rarely a simple one-to-one mapping that can be established between estimated perceptual parameters and behavioural measures. This greatly complicates the interpretation of data in comparison with forced choice tasks, as for example when comparing the degree to which visual illusions affect perception and action, since there is no single measure that can be meaningfully compared across different tasks (Franz, Gegenfurtner, Buelthoff & Fahle, 2000).

Virtual reality provides great opportunities in expanding the type of behavioural data that can be used in experiments, beyond simple forced choice tasks, while maintaining a high level of experimental control. The benefit of less constrained responses is the increased range of theoretical questions that can be posed, but this also presents much more challenging problems in the interpretation of data.

5. A vision science of natural environments and natural tasks

VR allows us to immerse a participant in the ambient optic array associated with a virtual environment; for this environment to be altered in response to the participant's behaviour; for complex naturalistic behaviour to be undertaken in this environment; and for full details of the structure of the environment, and the participant's actions, to be recorded. Together, these properties of VR expand the possibilities for experiments which address different areas of vision science, and different types of theoretical questions about vision, than traditional screen-based, forced choice experiments.

Traditional psychophysical techniques tend to be of most relevance to understanding how information is encoded, rather than how this is used in everyday vision. Typical models of visual encoding will include channels composed of perceptual filters or templates, non-linearities, sources of noise, and a final decision-making stage (Lu & Dosher, 1999). Because these models are generally used to account for performance in forced-choice categorisation tasks, they do not address how the information delivered by these channels might be used in the control of action, or in more complex cognitive tasks. Whether or not this presents a significant limitation depends on the nature of the visual processing under investigation.

Several rather different characterisations of the purpose of vision have been presented. One class, exemplified by the concept of 'inverse optics', sees the task of visual processing as reconstructing a representation of the three-dimensional environment on the basis of the information available from the 2D retinal images (Adelson and Pentland, 1996; Pizlo, 2001, Mamassian, Landy, & Maloney, 2002). In such models, the goal is the creation of a general-purpose visual representation, which is then accessed for all subsequent tasks (Marr, 1982). This might for example specify the distance to the seen point in each visual direction (Descartes, 1637/1965; Sedgwick, 2021), this distance plus the surface orientation at each point (Marr & Nishihara, 1978), or some other representation that specifies the three-dimensional structure of the environment, and our location within it.

Under this characterisation, it would be possible to account for how this process of inverse optics is achieved, without needing to concern ourselves with the nature of the tasks for which it is subsequently used. Conversely, if the encoding of visual information is not separable from the ways in which it is used, then it would be necessary to use complex, naturalistic tasks in order to understand perception.

Another important consideration is the way in which visual perception depends upon, and make assumptions about, the structure of the environment. Gibson's theory of ecological optics outlines a number of arguments for why typical everyday environments, and typical everyday activities within these environments, should be used in vision research.

The first consideration is that our environment is structured in ways that shape the nature of the visual information that is available, and in ways that determine how this is then used in perception. It is expected that the encoding of visual information will be tuned to these regularities, so as to be optimised for the typical natural environment (Olshausen & Field, 1996; Simoncelli & Olshausen, 2001). The distribution and material properties of objects and surfaces in our environment determine the statistical and structural properties of images. These include the luminance, (Field & Brady, 1997; Balboa, Tyler & Grzywacz, 2001; Rogers 2021), binocular disparity (Hibbard, 2007, 2008; Sprague, Cooper, Tošić & Banks, 2015), colour (Chiao, Cronin & Osorio, 2000) and motion (van Hateren, 1993; Dong & Atick, 1995a, 1995b) characteristics of typical natural scenes. The use of virtual reality to create naturalistic scenes as stimuli in psychophysical studies allow us to replicate these characteristics more fully than can be achieved using typical display screens with a smaller field of view in which stimuli are not updated to reflect the observer's movement.

One example is the way that different sources of information are used in the perception of distance and depth. Optimal cue combination models specify that cues should be combined through a weighted-averaging process, with the weights determined by the relative reliabilities of different cues (Landy, Maloney, Johnston & Young, 1995). Laboratory experiments with carefully specified stimuli have been successful in testing the predictions of these models (Ernst & Banks, 2002; Hillis, Watt, Landy, & Banks, 2004; Watt, Akeley, Ernst & Banks, 2005; Keefe, Hibbard & Watt, 2011). However, such models and experiments by themselves do not provide any information about the relative importance of cues in the natural environment. To do this requires some understanding of the reliabilities of these cues in typical everyday scenes (Nagata, 1991; Cutting & Vishton, 1995; Hibbard, 2021), and to test the predictions in this context requires us to perform experiments using such scenes. VR is valuable in this context in allowing us to perform psychophysical tests of the accuracy of 3D perception in typical everyday scenes, while manipulating the information that is available to observers (Hornsey, Scarfe and Hibbard, 2020, Hornsey & Hibbard, 2021; Hartle & Wilcox, 2021). Hornsey and Hibbard (2021) used this approach to quantify the contributions of various depth cues to the accuracy of distance and size judgements. They showed that binocular and pictorial cues both contribute to improved precision in size judgments, as predicted by theoretical models, and that binocular cues are important not only in near space, but also beyond distances of 10m.

Gibson (1950) proposed that the structure of the environment shapes the way that we make use of information in perception. He proposed a 'grounded' theory of perception, reflecting the fact that our natural environment is composed of surfaces rather than unconnected points. These surfaces create structure in the visual input, and may also determine the goals of perception; for example, we may judge the distance of objects relative to their background context, rather than their distance from the observer through empty space (Glennerster & McKee, 1999; Glennerster, McKee & Birch, 2002; Petrov & Glennerster, 2004; He & Ooi, 2000; Sedgwick, 2021). For these influences of surfaces, or other features of the environment on our perception to be captured, it is important that they are included in the stimuli used in experimental studies. The full extent of the structure of the natural environment can readily be incorporated when conducting experiments outdoors (Gibson,1950) or in VR, but not using traditional screen displays.

Gibson also emphasised the importance of thinking of perception as a process of active exploration, rather than passive observation. That is, we create the visual input that we experience by the way that we move in our environment and sample the ambient optic array. Again, this process of active exploration can only be incorporated into empirical studies by conducting them in the physical environment or in VR. Active exploration of the world creates optic flow (Lee, 1980; Rogers, 2021). As we move from one point to another, we experience different samples of the ambient optic array. The structure of surfaces in the environment means that these samples are related in predictable ways, creating higher-level patterns of motion which provide reliable information about the structure of the environment. Not only does this redundancy allow us to encode information efficiently (Olshausen & Field, 1996; Simoncelli & Olshausen, 2001), it can inform us directly about the structure of the environment, and our actions, without this needing to be mediated by simpler representations. Examples of this type of structure that have been proposed include gradients of motion and the orientation of surfaces; dynamic occlusion and the presence and depth order of surfaces; and the relationship between the focus of expansion of motion and the observer's direction of heading (Rogers, 2021).

The way in which we make use of visual information in everyday tasks also needs to be considered to properly understand how visual information is processed. Gibson proposed that we directly access the information that supports our actions in the world (Gibson, 1979; Sedgwick, 2021). Other formulations of the way in which visual information is used by observers include van Uexkuell's sensorimotor loop (Koenderink, 2019), O'Regan and Noe's sensorimotor account of vision (O'Regan & Noe, 2001) and the interface theory of perception (Hoffman, Singh & Prakesh, 2015), in which the goal of perception is not to create veridical representations, but to guide adaptive behaviours. Empirical testing of these theories therefore depends on the appropriate use of typical everyday tasks, an endeavour for which VR is ideally suited (Tarr & Warren, 2002; Bhargava, Lucaites, Hartman, Solini, Bertrand, Robb, Pagano & Babu, 2020). This has been applied, for example, in understanding important behaviours such as reaching and grasping (Hibbard & Bradshaw, 2003, Klinghammer, Schütz, Blohm, & Fiehler, 2016, Kopiske, Bozzacchi, Volcic, & Domini, 2019) and navigation (Tarr & Warren, 2002; Muryy & Glennerster, 2021).

VR has also been used to empirically evaluate the nature of our perception of space, notably addressing the question of whether we maintain a stable 3D representation of the visual world while we navigate within it. Glennerster and colleagues have, for example, shown that we can fail to notice large changes in the scale of our environment (Glennerster, Tcheang, Gilson, Fitzgibbon & Parker, 2006); that judgments of distances between multiple pairs of points are not necessarily mutually consistent within this environment (Svarverud, Gilson & Glennerster, 2012); that our ability to point to previously seen targets is not consistent with a single, stable 3D representation (Vuong, Fitzgibbon & Glennerster, 2019; Scarfe & Glennerster, 2021), and that we successfully use navigation strategies that are inconsistent with a Euclidean representation of space (Muryy & Glennerster, 2021). This programme of research has shown how it is possible to make use of VR to develop and test theoretical models of the essential nature of visual perception (Glennerster, 2016).

Virtual reality also has potential applications in areas of visual cognition beyond perception of the geometric structure of the environment. A critical example of this is the visual information used for social interactions. We are readily able to infer information about others' emotions from their facial expressions and movement. Most research on this topic makes use of stimuli that are artificial, in that they are static, grey-scale, monochrome photographs of actors portraying emotions, which are typically manipulated so as to control for psychophysically important properties such as luminance and contrast, rather than preserving systematic variations of these properties (Gray, Adams, Hedger & Newton, 2013; Menzel, Redies & Hayn-Leichsenring, 2018; Webb, Hibbard & O'Gorman, 2020; Webb, Asher & Hibbard, 2022). These highly impoverished stimuli are then used in simple forced-choice tasks, such as the detection of the presence of stimuli, or their categorisation as prespecified expressions. While this level of control may be necessary to isolate the relevant properties of the stimuli under investigation, it is important that the methodology is expanded to assess the relevance of such research in understanding how we use visual information in everyday social situations.

This can firstly be achieved by using more natural stimuli, particularly by including dynamic information (Jack & Schyns, 2015). Another important extension is to use environments in which virtual actors react to our own behaviours (Geraets, Tuente, Lestestuiver, van Beilen, Nijman, Marsman & Veling, 2021), and to use appropriate measures of real-world social understanding. These characteristics - more natural stimuli, reactivity, and the study of real-world behaviours - are all key features of virtual reality, emphasising the importance of the technology in this area of vision science. An important focus for research in this field is to ensure that social agents in VR are believable and accurately convey information about emotions and other social cues.

The focus of the current chapter is on the use of VR in vision research. It should be noted however that a powerful characteristic of VR is that it is a multisensory technology, allowing us to include spatial audio and haptic cues. The provision of multiple cues, and in particular the congruence between cues in a multisensory environment, is an important factor in creating an immersive experience (Servotte, Goosse, Campbell, Dardenne, Pilote, Simoneau, Guiillaume, Bragard & Ghuysen, 2020). Since VR allows for the independent manipulation of individual cues, it is ideally suited for exploring the role of multisensory information in the perception of our 3D environment.

6. Hardware and software characteristics of virtual reality

Lanier's fifth definition of VR gives a useful description of what we are trying to achieve, technically, in creating a virtual reality visual display:

> *A mirror image of a person's sensory and motor organs, or if you like, an inversion of a person*
>
> (Lanier, 2017, page 47)

As the user moves around and interacts with the environment, the goal is to create the ambient optic array that would be experienced in the simulated environment. This reframes the inverse problem not in terms of how the user recreates the 3D structure of the environment, but of how we can use VR to understand their experience and sensory-motor loops. As outlined in the introduction, this is achieved via hardware and software requirements that include (1) creation of the virtual environment, (2) tracking the motion of the user, (3) rendering the images of the environment to be presented, (4) displaying them on the two screens, and ensuring that the displayed images are clearly and comfortably perceived through (5) the use of appropriate head-set optics and (6) correct positioning of the display screens and optics relative to the observer. Each of these steps presents technical challenges. While these have all been addressed in current consumer reality to the extent that it now provides a comfortable, immersive, and convincing *feeling* of realism for most users, it is important in vision science to understand how this has been achieved and that differences remain between VR and the real world. In the following section, I highlight issues at each step in this pipeline that are relevant for vision scientists.

6.1 The environment

Technological advances mean that it is possible to move beyond VR environments composed of simple geometric forms to create complex, naturalistic scenes. This can include, for example, 3D scans of natural objects (Goutcher et al., 2021), landscapes (Liang, Liu & Zhou, 2014; Jeong, Lee & Kim, 2021), and human body movement and facial expressions captured from actors (Nonis, Dagnes, Marcolin & Vezzetti, 2019). This creates a great deal of scope when designing virtual environments for experiments in vision science, requiring decisions about the level of complexity that is most appropriate for a given research context. It has been argued that, when creating visual displays in general, we should avoid the temptation to increase complexity (Smallman & John, 2005). However, in the case of vision science, our goals include understanding the way in which particular aspects of our visual environment influence perception. The potential to create highly realistic, immersive environments opens opportunities for understanding the important characteristics of the ambient optic array, in the same way that digital photographs have done for static and dynamic images (Simoncelli & Olshausen, 2001). Decisions about the complexity and realism of the environment are the same in both cases. While simplified artificial stimuli run the risk of excluding important information, the complexity of naturalistic stimuli may make experimental designs intractable (Rust & Movshon, 2005).

As environments become more seemingly realistic, there is also the danger that any artefactual differences between the real and virtual environments may go unnoticed (Koenderink, 1999). In vision science, there is a risk that our findings may reflect these artefacts. In the case of photographs, for example, the framing and composition of the image may create global statistics that are not representative of mundane, everyday images, while the rectangular pixel lattice will influence local statistical properties (van Hateren & van der Schaaf, 1998; Hertzmann, 2022). In the case of VR, where we determine the shape, material properties, position, movement, and behaviour of all objects, we increase the scope that any of these properties may diverge from what is characteristic of typical natural scenes, whilst still appearing highly realistic to observers. For example, the rendering of scenes in VR requires models of complex natural phenomena including the movement of, and collisions between, objects and materials, which must be simulated to approximate the real world.

6.2 Motion tracking and latency

Virtual reality requires us to track the user’s position so that the display can be updated in as close to real-time as possible. Any latencies in this process will have negative effects on the control of action and the experience of agency, ownership, presence, and immersion (Waltemate, Senna, Hülsmann, Rohde, Kopp, Ernst & Botsch, 2016), while also contributing to feelings of sickness. While latency has improved in recent VR systems, with a 90Hz update rate being typical, latencies as low as 7ms are detectable by observers (Scarfe & Glennerster, 2019). As with all aspects of the virtual environment, the deliberate manipulation of these parameters, for example to include high latencies to evaluate its effect on multisensory processing and perceptual experience (van Dam & Stephens, 2018), or incorrect positioning of the observer to understand the role of eye-height in the perception of distance (Leyrer, Linkenauger, Bülthoff, Kloos, & Mohle, 2011; Kim & Interrante, 2017), demonstrates how VR provides flexible control of the environment that can be harnessed in experiments.

In addition to ensuring the accurate presentation of the environment, tracking of head and eye-movements provides rich behavioural data that allows us to understand how people navigate and interact with virtual environments (Alcañiz, Chicchi-Giglioli, Carrasco-Ribelles, Marín-Morales, Minissi, Teruel-García, Sirera & Abad, 2022; Gulhan, Durant, & Zanker, 2022).

6.3 Rendering the images

Early VR used wireframes or simply shaded objects, so that it was obvious to the user that the environment was artificial. Just as our ability to create more complex 3D environments has improved, so has our ability to render these to create images that have a highly realistic appearance (Zibrek, Martin & McDonnell, 2019). As this level of realism improves, it becomes increasingly difficult to tell natural from artificial environments. While this is of course highly beneficial for applications of VR, when used in vision science there is a risk that our results are influenced by artefacts which may not be noticeable to the participants or experimenters. These will occur in particular due to the technical challenge of accurately rendering lighting (Koenderink, 1999).

Some of the complexities here come from the facts that each point on a surface is lit not just by a light source from a single direction, and that once light has hit a matte surface, it will be reflected in all directions, not just towards the observer. This means that light will bounce between surfaces, and objects will cast shadows on one another as they occlude portions of the ambient light array. All of this can be simulated with ray-tracing (Todd, Egan & Kallie, 2015; Todd, 2020), however this is very computationally intensive and therefore slow to render. Currently, even with high-powered cluster computing, a single frame can take minutes to create rather than the milliseconds available to create stimuli for real-time display. For vision science, it is therefore important either to use physically-accurate lighting models, or to understand the details and consequences of departures from accuracy for a given experiment design.

6.4 The display screen

For each eye, the image is displayed on a screen with a particular spatial resolution, field of view, and dynamic range. The best resolution of the human eye, in the fovea, is around 1 arc min, meaning that we can discriminate between one and two points if they are presented farther apart than this. This means that we can see pixilation if pixels are larger than this separation. Currently pixels in VR headsets are about 3 arc min in size. Screen resolution thus still limits our ability to present very fine-detailed information, and to specify the precise location of features through anti-aliasing. This has implications for studying aspects of visual acuity in VR, and can create artefactual differences in the statistical properties of scenes in comparison with the natural environment.

Human vision operates over a high dynamic range, sensitive to illumination from as low as 10^{-6} cd m^{-2} to as high as 10^{8} cd m^{-2} (Banterle, Artusi, Debattista & Chalmers, 2017). This is much greater than the range that is achievable with high dynamic range displays, thus limiting the range of illumination over which vision science can operate, for VR and other displays alike. This limited dynamic range is also a consideration in determining the sense of realism in VR (Vangorp, Bazyluk, Myszkowski, Mantiuk, Watt & Siedel, 2014).

The human binocular field of view is approximately 200 degrees horizontally by 100 degrees vertically, with a binocular overlap of 120 degrees (Spector, 1990). In current VR, the horizontal and vertical extents of the field of view are around 90 degrees and 36 degrees respectively, with a 50 degree horizontal binocular overlap. Thus, while VR provides a very wide field of view in comparison with a traditional display screen, it is still limited in the extent to which it can be used to study peripheral vision (Strasburger, Rentschler & Jüttner, 2011). The reduced binocular field of view also means that the contributions of binocular vision, for example to the perception of distance and the observer’s movement, may be underestimated when studied using VR (figure 3).

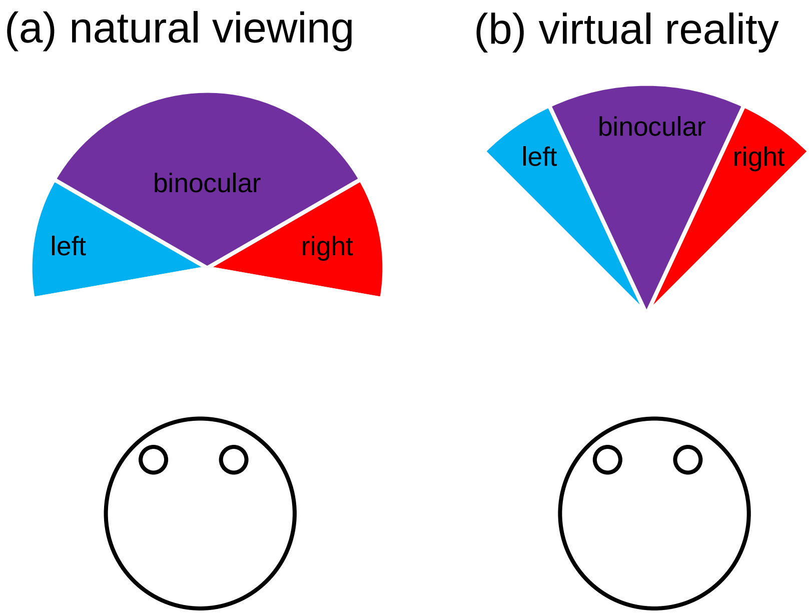

Figure 3. *(a) The human binocular field extends to 200 degrees, with 120 degrees of binocular overlap. (b) Both the overall field of view, and the binocular overlap, are smaller than this in VR.*

6.5 Headset optics

The display screen in an HMD is positioned just a few millimetres from the viewer. In order to accommodate a screen at this distance, powerful lenses are placed in the headset, creating an image plane that is at some distance from the user (figure 4). This image plane appears at a constant, set distance in VR. This distance is determined by the lenses, regardless of the distance to the object in the scene, as specified by binocular convergence and other cues. This creates cue-conflict, and difficulties for the coupling of accommodation and convergence responses, which are known to be a source of discomfort in VR and other 3D displays (Shibata, Kim, Hoffman & Banks, 2011). This could theoretically be addressed by the use of multifocal displays, which use multiple overlapping screens at different distances. This technique allows the effective image distance of each point to be varied independently, by distributing it across these multiple screens. This in turn allows the accommodative load to vary with object distance, providing a more natural relationship between convergence and accommodation (Zhong, Jindal, Yöntem, Hanji, Watt & Mantiuk, 2021). Multifocal displays are not however currently in use in HMDs.

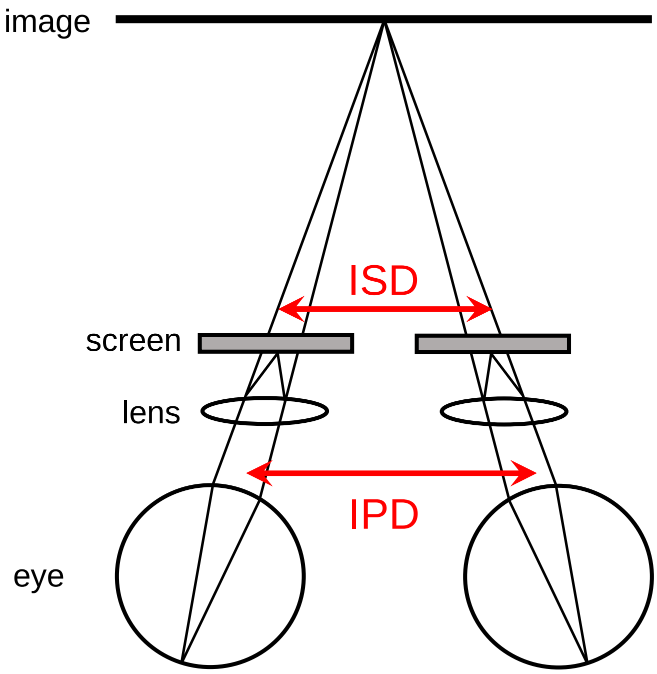

Figure 4. *Within the headset, the optics focus the light from the screen so that the image appears at a comfortable distance. The separation of the lenses, the inter-screen distance (ISD), and the effective interpupillary distance (IPD) used in rendering stimuli, should all ideally be matched to the observer's IPD.*

In natural viewing, objects that are not at the accommodation distance will be blurred. In VR, this optical blurring will not occur, since in this case the amount of blur depends on the distance to the image plane, rather than to the object in the rendered scene. This can be compensated for by including an appropriate blur into the displayed image (Held, Cooper & Banks, 2012). Again, by allowing control over the image distance, multifocal displays would remove these incorrect focus cues, introducing natural, optically created defocus to the retinal image.

6.6 Positioning of the display screens, and lenses, relative to the observer.

When we render the stimuli, we need to specify the optical centres of the two eyes in the environment. Motion tracking allows us to measure the position of the headset. To then determine the positions of the optical centres relative to the headset requires us to specify the interpupillary distance (IPD) – the distance between the user's two eyes (figure 4). If there is a mismatch between the observer's IPD and that assumed in rendering and presenting the stimuli, then the binocular disparities in the images will not accurately reflect those that would be experienced when viewing the intended scene (Scarfe & Glennerster, 2019). This is an important consideration because IPD varies considerably between individuals. For adults, the overall mean IPD is 63.4 m, with a standard deviation of 3.8mm (Dodgson, 2004). This represents a mean of 63.4 mm for women, and 64.7 mm for men. It also increases with age from a mean value of 50 mm at the age of 5, reaching the final adult value at an age of 19 (MacLachlan & Howland, 2002). There are a number of important ways in which a mismatch between the effective IPD in the headset, and the user's IPD, can negatively affect the viewing experience (Hibbard, van Dam and Scarfe, 2020).

The first problem is the potential for binocular divergence (figure 5a). In natural viewing, when we fixate an object in the distance, the directions of the gaze of the two eyes are parallel; as the object moves closer, the eyes will converge, with the angle of convergence increasing with decreasing distance. Divergence of the eyes is therefore never required in order to fixate a target in natural viewing, although in practice this can occur due to imperfect fixation (Darko-Takyi, Khan & Nirghin, 2016). If the observer's IPD is smaller than that used in creating and rendering the stimuli, divergent viewing will be required to fixate distant stimuli. Our ability to make such divergent eye-movements is limited, leading to a loss of binocular fusion, and viewing discomfort (IJsselsteijn, de Ridder & Vliegen, 2000; Hoffman, Girshick, Akeley & Banks, 2008; Lambooij, IJsselsteijn, Heynderickx, 2011; Shibata, Kim, Hoffman & Banks, 2011).

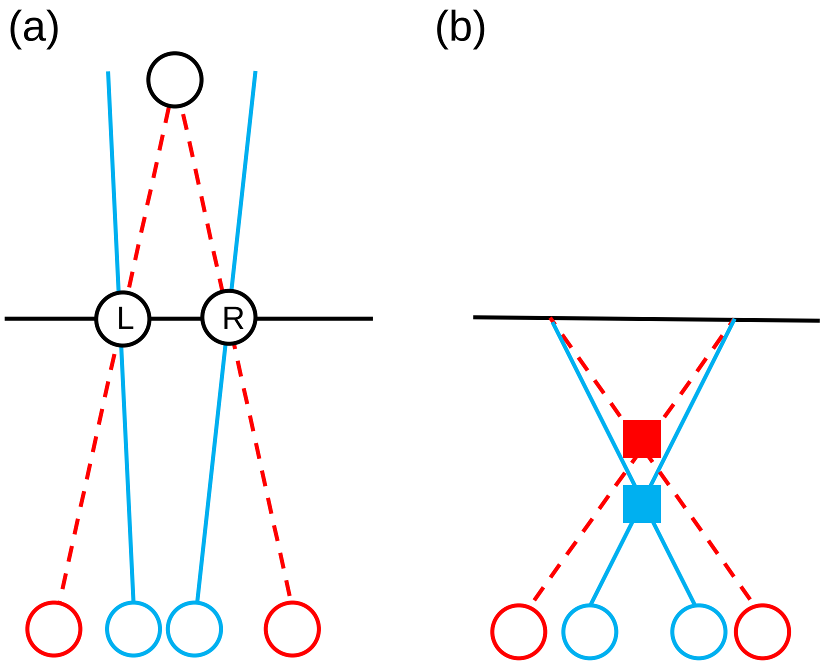

Figure 5. *(a) A bird's eye view showing the positioning of a point in the left and right eyes' images, in order to create the appearance of an object (the unfilled circle) at a distance beyond the image plane. The assumed locations of the observer's eyes used to position the points are shown by the red circles. If the observer's eyes (depicted by the blue circles) are closer together than assumed, this can create divergent viewing. (b) An incorrectly assumed IPD will also lead to an incorrect distance specified by binocular cues. In this case, the target will appear closer than intended.*

The second problem is that when the assumed and actual IPD are not matched, the perceived distance, shape, and size of objects specified by binocular cues will be incorrect (figure 5b). The binocular disparity of a point is determined by the fixation distance, the position of the point relative to fixation, and the viewer's IPD. This means that two viewers with different IPDs will experience different binocular disparities when viewing the same scene. The rendering and display of stimuli with an incorrect IPD will thus present incorrect binocular information to the viewer. Specifically, if the IPD used is too large, the specified distance will be closer than intended, while if it is too small, if will be larger than intended. The predictions for the misperception of distance and depth from these calculations are a worst-case scenario, however, assuming that distance is perceived purely on the basis of binocular cues. When the influence of other cues such as perspective and motion parallax are taken into account, the effect of an incorrect IPD will be much reduced (Hibbard et al., 2020; Hibbard, 2021).

The maximum binocular disparities that can be presented on a display screen are limited by the conflict between convergence and accommodation when objects are presented away from the image plane. On 3D display screens, including in VR, there is a maximum binocular disparity of up to 2 degrees than can be presented without creating discomfort (Hoffman, Girshick, Akeley & Banks, 2008; Lambooij, IJsselsteijn, Bouwhuis &Heynderickx, 2011; Shibata, Kim, Hoffman & Banks, 2011). From this limit, we can calculate the range of distances that can be comfortably presented to the viewer. Since the relationship between distance and disparity depends on IPD, this comfortable range will also vary with the IPD used to render stimuli, with the comfortable range decreasing with increasing IPD. Regardless of the viewer's IPD, a larger range of depths can be presented by decreasing the IPD used to render and display the stimuli (Siegel & Nagata, 2000). This however comes at the cost of the problems of divergence, misperception of distance and off-axis viewing discussed elsewhere in this chapter (McIntire, Havig, Harrington, Wright, Watamaniuk & Heft, 2018).

Even if the images are rendered and positioned appropriately for the observers' IPD, there may still be a misalignment between the positions of the lenses and the eye, creating off-axis viewing. This introduces a number of optical artefacts (Howarth, 1999), including prismatic effects, which shift the images away from their intended locations (Peli, 1995) and chromatic aberration (Beams, Kim & Badano, 2019). In order to minimise the problems associated with in an incorrect IPD, therefore, it would be necessary to set both the lens separation, and the IPD used in software rendering, to match the observer. Although this is not practical in most current research pipelines and practices, it may become possible with future technological development.

7. Conclusions

VR presents many new opportunities for vision sciences. Notably, it allows us to create virtual environments that are immersive and that react to our movements and actions. This in turn provides us with stimuli that more fully reflect our natural environment, and affords a broader range of interactions with them, than is generally possible with a typical display screen. Firstly, it allows us to create wide field-of-view environments that reflect many of the statistical properties of our natural environment. Secondly, it allows the user to actively explore the environment. This is a particularly important characteristic, since our visual inputs are determined not just by the structure of the environment, but by our own actions within it.

VR also provides many opportunities to extend our theoretical understanding of the ways in which visual processing may be influenced by, or directly linked to, the control of behaviour. In comparison with traditional, forced-choice psychophysical experiments, this introduces a degree of complexity to the data that we gather. The control of visual stimuli, the manner and order in which they are presented, and the responses available to the participant are fundamental to the power of the psychophysical approach. In contrast, the unconstrained, open-world exploration of and interaction with the environment that are possible in VR mean that the data that our experiments generate have the potential to be much less structured. It is therefore important, in designing experiments in VR, to balance the ecological validity that it provides with the need for data that can be interpreted in relation to our theoretical research questions.

VR presents some technical considerations related to the nature of the visual display, and the extent to which both software and hardware restrictions allow us to accurately create our intended visual environments. It is important therefore for us to remember that VR is not a substitute for other display technologies and psychophysical techniques. Rather, these different approaches should be seen as complementary. Gibson, for example, argued that theories of vision need to account for our experience of the visual field, or of the visual world giving rise to this visual field (Gibson, 1950, Sedgwick, 2021). VR is particularly well-suited to understanding our interactions with the visual world.

References

Adelson, E. H., & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. Nature, 300(5892), 523-525.

Adelson, E. H., & Pentland, A. P. (1996). The perception of shading and reflectance. Perception as Bayesian inference, 409, 423.

Alcañiz, M., Chicchi-Giglioli, I. A., Carrasco-Ribelles, L. A., Marín-Morales, J., Minissi, M. E., Teruel-García, G., Sirera, M. & Abad, L. (2022). Eye gaze as a biomarker in the recognition of autism spectrum disorder using virtual reality and machine learning: A proof of concept for diagnosis. *Autism Research*, *15*(1), 131-145.

Backus, B. T., Banks, M. S., Van Ee, R., & Crowell, J. A. (1999). Horizontal and vertical disparity, eye position, and stereoscopic slant perception. Vision Research, 39(6), 1143-1170.

Balboa, R. M., Tyler, C. W., & Grzywacz, N. M. (2001). Occlusions contribute to scaling in natural images. Vision Research, 41(7), 955-964.

Banterle, F., Artusi, A., Debattista, K., & Chalmers, A. (2017). Advanced high dynamic range imaging. AK Peters/CRC Press.

Beams, R., Kim, A. S., & Badano, A. (2019). Transverse chromatic aberration in virtual reality head-mounted displays. Optics Express, 27(18), 24877-24884.

Bhargava, A., Lucaites, K. M., Hartman, L. S., Solini, H., Bertrand, J. W., Robb, A. C., Pagano, C.C. & Babu, S. V. (2020). Revisiting affordance perception in contemporary virtual reality. Virtual Reality, 24(4), 713-724.

Braddick, O. J. (1974) A short-range process in apparent motion. Vision Research, 14, 519-527.

Braddick, O., Campbell, F. W., & Atkinson, J. (1978). Channels in vision: Basic aspects. In Perception (pp. 3-38). Springer, Berlin, Heidelberg.

Braddick, O. (1981). Spatial frequency analysis in vision. Nature, 291(5810), 9-11.

Bradshaw, M. F., Elliott, K. M., Watt, S. J., Hibbard, P. B., Davies, I. R., & Simpson, P. J. (2004). Binocular cues and the control of prehension. Spatial Vision, 17(1-2), 95-110.

Brainard, D. H. (1997). The psychophysics toolbox. Spatial Vision, 10(4), 433-436.

Brindley, G.S. (1960). Physiology of the Retina and Visual Pathway. London: Edward Arnold.

Chiao, C. C., Cronin, T. W., & Osorio, D. (2000). Color signals in natural scenes: characteristics of reflectance spectra and effects of natural illuminants. Journal of the Optical Society of America A, 17(2), 218-224.

Cleland, B. G., Dubin, M. W., & Levick, W. R. (1971). Sustained and transient neurones in the cat's retina and lateral geniculate nucleus. The Journal of Physiology, 217(2), 473-496.

Cruz-Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. V., & Hart, J. C. (1992). The CAVE: audio visual experience automatic virtual environment. Communications of the ACM, 35(6), 64-73.

Cutting, J. E., & Vishton, P. M. (1995). Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In Perception of space and motion (pp. 69-117). Academic Press.

Darko-Takyi, C., Khan, N. E., & Nirghin, U. (2016). A review of the classification of nonstrabismic binocular vision anomalies. Optometry Reports, 5(1).

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. Journal of the Optical Society of America A, 2(7), 1160-1169.

Descartes, R. (1637/1965). Discourse on method, optics, geometry, and meteorology (P. J. Olscamp,Trans.). Bobbs-Merrill.

Dodgson, N. A. (2004, May). Variation and extrema of human interpupillary distance. In Stereoscopic displays and virtual reality systems XI (Vol. 5291, pp. 36-46). International Society for Optics and Photonics.

Dong, D. W., & Atick, J. J. (1995a). Statistics of natural time-varying images. Network: Computation in Neural Systems, 6(3), 345.

Dong, D. W., & Atick, J. J. (1995b). Temporal decorrelation: a theory of lagged and nonlagged responses in the lateral geniculate nucleus. Network: Computation in Neural Systems, 6(2), 159-178.

Dosher, B. A., & Lu, Z. L. (1999). Mechanisms of perceptual learning. Vision Research, 39(19), 3197-3221.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. Nature, 415(6870), 429-433.

Field, D. J., & Brady, N. (1997). Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes. Vision Research, 37(23), 3367-3383.

Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research*, *51*(17), 1920-1931.

Franz, V. H., Gegenfurtner, K. R., Buelthoff, H. H., & Fahle, M. (2000). Grasping visual illusions: No evidence for a dissociation between perception and action. Psychological Science, 11(1), 20-25.

Geraets, C. N. W., Tuente, S. K., Lestestuiver, B. P., van Beilen, M., Nijman, S. A., Marsman, J. B. C., & Veling, W. (2021). Virtual reality facial emotion recognition in social environments: An eye-tracking study. Internet Interventions, 25, 100432.

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

Gibson, J. J. (1979). The ecological approach to visual perception. Houghton Mifflin Company.

Glennerster, A. (2016). A moving observer in a three-dimensional world. Philosophical Transactions of the Royal Society B: Biological Sciences, 371(1697), 20150265.

Glennerster, A., & McKee, S. P. (1999). Bias and sensitivity of stereo judgements in the presence of a slanted reference plane. Vision Research, 39(18), 3057-3069.

Glennerster, A., McKee, S. P., & Birch, M. D. (2002). Evidence for surface-based processing of binocular disparity. Current Biology, 12(10), 825-828.

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.

Goutcher, R., Barrington, C., Hibbard, P. B., & Graham, B. (2021). Binocular vision supports the development of scene segmentation capabilities: Evidence from a deep learning model. Journal of vision, 21(7), 13-13

Gulhan, D., Durant, S., & Zanker, J. M. (2022). Aesthetic judgments of 3D arts in virtual reality and online settings. *Virtual Reality*, 1-17.

Gray, K. L., Adams, W. J., Hedger, N., Newton, K. E., & Garner, M. (2013). Faces and awareness: low-level, not emotional factors determine perceptual dominance. Emotion, 13(3), 537.

Hartle, B., & Wilcox, L. M. (2021). Cue vetoing in depth estimation: Physical and virtual stimuli. Vision Research, 188, 51-64.

He, Z. J., & Ooi, T. L. (2000). Perceiving binocular depth with reference to a common surface. Perception, 29(11), 1313-1334.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. Visual Neuroscience, 9(2), 181-197.

Heeley, D. W., Buchanan-Smith, H. M., Cromwell, J. A., & Wright, J. S. (1997). The oblique effect in orientation acuity. Vision Research, 37(2), 235-242.

Held, R. T., Cooper, E. A., & Banks, M. S. (2012). Blur and disparity are complementary cues to depth. Current Biology, 22(5), 426-431.

Hertzmann, A. (2022). The choices hidden in photography. *Journal of Vision*, *22*(11).

Hibbard, P. (2007). A statistical model of binocular disparity. Visual Cognition, 15(2), 149-165.

Hibbard, P. B. (2008). Binocular energy responses to natural images. Vision Research, 48(12), 1427-1439.

Hibbard, P. B. (2021 in press) Estimating the contributions of pictorial, motion and binocular cues to the perception of distance. ECVP.

Hibbard, P. B., & Bradshaw, M. F. (2003). Reaching for Virtual Objects: Binocular Disparity and the Control of Prehension. Experimental Brain Research, 148(2), 196-201.

Hibbard, P. B., van Dam, L. C., & Scarfe, P. (2020). The implications of interpupillary distance variability for virtual reality. In 2020 International Conference on 3D Immersion (IC3D) (pp. 1-7).

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. Journal of Vision, 4(12), 1-1.

Hoffman, D. D., Singh, M., & Prakash, C. (2015). The interface theory of perception. Psychonomic Bulletin & Review, 22(6), 1480-1506.

Hoffman DM, Girshick AR, Akeley K & Banks MS (2008) Vergence–accommodation conflicts hinder visual performance and cause visual fatigue, Journal of Vision, 8(3):33–33.

Hornsey, R. L., & Hibbard, P. B. (2021). Contributions of pictorial and binocular cues to the perception of distance in virtual reality. Virtual Reality, 1-17.

Hornsey, R. L., Hibbard, P. B., & Scarfe, P. (2020). Size and shape constancy in consumer virtual reality. Behavior Research Methods, 52(4), 1587.

Howarth, P. A. (1999). Oculomotor changes within virtual environments. Applied Ergonomics, 30(1), 59-67.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. The Journal of Physiology, 160(1), 106-154.

IJsselsteijn WA, de Ridder H & Vliegen J (2000) . Subjective evaluation of stereoscopic images: effects of camera parameters and display duration. IEEE Transactions on Circuits and Systems for Video Technology, 10(2):225–233.

Jack, R. E., & Schyns, P. G. (2015). The human face as a dynamic tool for social communication. Current Biology, 25(14), R621-R634.

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

Johnston, E. B., Cumming, B. G., & Parker, A. J. (1993). Integration of depth modules: Stereopsis and texture. Vision Research, 33(5-6), 813-826.

Julesz, B. (1971). Foundations of cyclopean perception. Chicago: University of Chicago Press.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. Philosophical Transactions of the Royal Society B: Biological Sciences, 361(1476), 2109-2128.

Keefe, B. D., Suray, P. A., & Watt, S. J. (2019). A margin for error in grasping: hand pre-shaping takes into account task-dependent changes in the probability of errors. Experimental Brain Research, 237(4), 1063-1075

Keefe, B. D., Hibbard, P. B., & Watt, S. J. (2011). Depth-cue integration in grasp programming: no evidence for a binocular specialism. Neuropsychologia, 49(5), 1246-1257.

Kim, J., & Interrante, V. (2017). Dwarf or giant: the influence of interpupillary distance and eye height on size perception in virtual environments. In 27th International Conference on

Artificial Reality and Telexistence, ICAT 2017 and the 22nd Eurographics Symposium on Virtual Environments, EGVE 2017 (pp. 153-160). Eurographics Association.
Klinghammer, M., Schütz, I., Blohm, G., & Fiehler, K. (2016). Allocentric information is used for memory-guided reaching in depth: A virtual reality study. Vision Research, 129, 13-24.
Koenderink, J. J. (1984). The concept of local sign. In: van Doorn A. J., van de Grind W. A., & Koenderink J. J. (Eds.), Limits in Perception. Utrecht: VNU Science Press.
Koenderink, J.J. (1999) Virtual Psychophysics, Perception, 28, 669-674.
Koenderink, J. J. (1998). Pictorial relief. Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 356(1740), 1071-1086.
Koenderink, J.J. (2019) Vision, an optical user interface, Perception, 48(7), 545-601.
Koenderink, J. J., van Doorn, A. J., Kappers, A. M., & Todd, J. T. (2001). Ambiguity and the 'mental eye'in pictorial relief. Perception, 30(4), 431-448.
Kopiske, K. K., Bozzacchi, C., Volcic, R., & Domini, F. (2019). Multiple distance cues do not prevent systematic biases in reach to grasp movements. Psychological research, 83(1), 147-158.
Lambooij, M., IJsselsteijn, W. A., & Heynderickx, I. (2011). Visual discomfort of 3D TV: Assessment methods and modeling. Displays, 32(4), 209-218.
Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. Vision Research, 35(3), 389-412.
Lanier, J. (2017). Dawn of the new everything: A journey through virtual reality. Random House.
Lee, D. N. (1980). The optic flow field: The foundation of vision. Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 290(1038), 169-179.
Leyrer, M., Linkenauger, S. A., Bülthoff, H. H., Kloos, U., & Mohler, B. (2011, August). The influence of eye height and avatars on egocentric distance estimates in immersive virtual environments. In Proceedings of the ACM SIGGRAPH symposium on applied perception in graphics and visualization (pp. 67-74).
Liang, Z., Liu, D., & Zhou, M. (2014, December). Research on large scale 3D terrain generation. In 2014 IEEE 17th International Conference on Computational Science and Engineering (pp. 1827-1832).
Livingstone, M. S., & Hubel, D. H. (1987). Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. Journal of Neuroscience, 7(11), 3416-3468.
MacLachlan, C., & Howland, H. C. (2002). Normal values and standard deviations for pupil diameter and interpupillary distance in subjects aged 1 month to 19 years. Ophthalmic and Physiological Optics, 22(3), 175-182.
Mamassian, P., Landy, M., & Maloney, L. T. (2002). Bayesian modelling of visual perception. Probabilistic models of the brain, 13-36.
Marr, D (1982) Vision. A Computational Investigation into the Human Representation and Processing of Visual Information (New York: W H Freeman)
Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. Proceedings of the Royal Society of London. Series B. Biological Sciences, 200(1140), 269-294.
McIntire, J.P., Havig, P.R., Harrington, L.K., Wright, S.T., Watamaniuk, S.N.J. & Heft, E. (2018) Microstereopsis is good, but orthostereopsis is better: Precision alignment task performance and viewer discomfort with a stereoscopic 3d display. In Three- Dimensional Imaging, Visualization, and Display
Melmoth, D. R., & Grant, S. (2006). Advantages of binocular vision for the control of reaching and grasping. Experimental Brain Research, 171(3), 371-388.
Menzel, C., Redies, C., & Hayn-Leichsenring, G. U. (2018). Low-level image properties in facial expressions. Acta psychologica, 188, 74-83.
Milner, D., & Goodale, M. (2006). The visual brain in action, OUP Oxford.

Mon-Williams, M., Tresilian, J. R., & Roberts, A. (2000). Vergence provides veridical depth perception from horizontal retinal image disparities. Experimental brain research, 133(3), 407-413.

Morgan, M. J., Melmoth, D., & Solomon, J. A. (2013). Linking hypotheses underlying Class A and Class B methods. Visual Neuroscience, 30(5-6), 197-206.

Muryy, A., & Glennerster, A. (2021). Route selection in non-Euclidean virtual environments. PloS one, 16(4), e0247818.

Nagata, S. (1991). How to reinforce perception of depth in single two-dimensional pictures. In Pictorial Communication In Real And Virtual Environments (pp. 553-573). CRC Press.

Newsome, W. T., Britten, K. H., & Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. Nature, 341(6237), 52-54.

Nonis, F., Dagnes, N., Marcolin, F., & Vezzetti, E. (2019). 3D Approaches and challenges in facial expression recognition algorithms—a literature review. Applied Sciences, 9(18), 3904.

Olshausen, B. A., & Field, D. J. (1996). Natural image statistics and efficient coding. Network: Computation in Neural Systems, 7(2), 333.

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. Behavioral and Brain Sciences, 24(5), 939-973.

Parker, A. J., & Newsome, W. T. (1998). Sense and the single neuron: probing the physiology of perception. Annual Review of Neuroscience, 21(1), 227-277.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. Journal of Neuroscience Methods, 162(1-2), 8-13.

Peli, E. (1995). Real vision & virtual reality. Optics and Photonics News, 6(7), 28.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision, 10, 437-442.

Petrov, Y., & Glennerster, A. (2004). The role of a local reference in stereoscopic detection of depth relief. Vision Research, 44(4), 367-376.

Pizlo, Z. (2001). Perception viewed as an inverse problem. Vision Research, 41(24), 3145-3161.

Pizlo, Z. (2010). 3D shape: Its unique place in visual perception. MIT Press.

Plumert, J. M., Kearney, J. K., Cremer, J. F., & Recker, K. (2005). Distance perception in real and virtual environments. ACM Transactions on Applied Perception (TAP), 2(3), 216-233.

Rogers, B. (2021). Optic Flow: Perceiving and Acting in a 3-D World. i-Perception, 12(1), 2041669520987257.

Rushton, S. K., Harris, J. M., Lloyd, M. R., & Wann, J. P. (1998). Guidance of locomotion on foot uses perceived target location rather than optic flow. Current Biology, 8(21), 1191-1194.

Rust, N. C., & Movshon, J. A. (2005). In praise of artifice. Nature Neuroscience, 8(12), 1647-1650.

Scarfe, P., & Glennerster, A. (2015). Using high-fidelity virtual reality to study perception in freely moving observers. Journal of Vision, 15(9), 3-3.

Scarfe, P., & Glennerster, A. (2019). The science behind virtual reality displays. Annual Review of Vision Science, 5, 529-547.

Scarfe, P., & Glennerster, A. (2021). Combining cues to judge distance and direction in an immersive virtual reality environment. Journal of Vision, 21(4), 10-10.

Schnapf, J. L., Kraft, T. W., Nunn, B. J., & Baylor, D. A. (1988). Spectral sensitivity of primate photoreceptors. Visual neuroscience, 1(3), 255-261.

Servos, P., Goodale, M. A., & Jakobson, L. S. (1992). The role of binocular vision in prehension: a kinematic analysis. Vision Research, 32(8), 1513-1

Sedgwick, H. A. (2021). JJ Gibson's "Ground Theory of Space Perception". i-Perception, 12(3), 20416695211021111

Servotte, J. C., Goosse, M., Campbell, S. H., Dardenne, N., Pilote, B., Simoneau, I. L, Guillaume, M., Bragard, I. & Ghuysen, A. (2020). Virtual reality experience: Immersion, sense of presence, and cybersickness. *Clinical Simulation in Nursing*, *38*, 35-43.

Shibata T, Kim J, Hoffman DM & Banks, MS (2011) The zone of comfort: Predicting visual discomfort with stereo displays. Journal of Vision, 11(8):11–11

Siegel, M. & Nagata S. (2000) . Just enough reality: comfortable 3-d viewing via microstereopsis. IEEE Transactions on Circuits and Systems for Video Technology, 10(3):387–396, 2000.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. Annual Review of Neuroscience, 24(1), 1193-1216

Simons, D. J., & Levin, D. T. (1998). Failure to detect changes to people during a real-world interaction. Psychonomic Bulletin & Review, 5(4), 644-649.

Smallman, H. S., & John, M. S. (2005). Naïve realism: Misplaced faith in realistic displays. Ergonomics in Design, 13(3), 6-13.

Spector RH. 1990. Visual fields. In Clinical Methods: The History, Physical, and Laboratory Examinations, ed. HK Walker, WD Hall, JW Hurst, pp. 565–72. Boston: Butterworths.

Sprague, W. W., Cooper, E. A., Tošić, I., & Banks, M. S. (2015). Stereopsis is adaptive for the natural environment. Science advances, 1(4), e1400254.

Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern recognition: A review. Journal of vision, 11(5), 13-13.

Svarverud, E., Gilson, S., & Glennerster, A. (2012). A demonstration of 'broken'visual space. PLoS One, 7(3), e33782.

Tarr, M. J., & Warren, W. H. (2002). Virtual reality in behavioral neuroscience and beyond. Nature neuroscience, 5(11), 1089-1092.

Todd, J. T. (2020). On the Ambient Optic Array: James Gibson's Insights About the Phenomenon of Chiaroscuro. i-Perception, 11(5), 2041669520952097.

Todd, J. T., Egan, E. J., & Kallie, C. S. (2015). The darker-is-deeper heuristic for the perception of 3D shape from shading: Is it perceptually or ecologically valid?. Journal of vision, 15(15), 2-2.

Tresilian, J. R., Mon-Williams, M., & Kelly, B. M. (1999). Increasing confidence in vergence as a cue to distance. Proceedings of the Royal Society of London. Series B: Biological Sciences, 266(1414), 39-44.

van Dam, L. C., & Stephens, J. R. (2018). Effects of prolonged exposure to feedback delay on the qualitative subjective experience of virtual reality. PloS One, 13(10), e0205145.

Van Hateren, J. H. (1993). Spatiotemporal contrast sensitivity of early vision. Vision Research, 33(2), 257-267.

Vangorp, P., Mantiuk, R. K., Bazyluk, B., Myszkowski, K., Mantiuk, R., Watt, S. J., & Seidel, H. P. (2014, August). Depth from HDR: depth induction or increased realism?. In Proceedings of the ACM Symposium on Applied Perception (pp. 71-78).

Van Hateren, J. H., & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. Proceedings of the Royal Society of London. Series B: Biological Sciences, 265(1394), 359-366.

Vuong, J., Fitzgibbon, A. W., & Glennerster, A. (2019). No single, stable 3D representation can explain pointing biases in a spatial updating task. Scientific Reports, 9(1), 1-13.

Wade, N. J. (2013). Deceiving the brain: Pictures and visual perception. Progress in Brain Research, 204, 115-134.

Waltemate, T., Senna, I., Hülsmann, F., Rohde, M., Kopp, S., Ernst, M., & Botsch, M. (2016). The impact of latency on perceptual judgments and motor performance in closed-loop interaction in virtual reality. In Proceedings of the 22nd ACM conference on virtual reality software and technology (pp. 27-35).

Warren, W. H., & Hannon, D. J. (1988). Direction of self-motion is perceived from optical flow. Nature, 336(6195), 162-163.

Watt, S. J., Akeley, K., Ernst, M. O., & Banks, M. S. (2005). Focus cues affect perceived depth. Journal of Vision, 5(10), 7-7.

Watt, S. J., & Bradshaw, M. F. (2000). Binocular cues are important in controlling the grasp but not the reach in natural prehension movements. Neuropsychologia, 38(11), 1473-1481.

Webb, A. L., Asher, J. M., & Hibbard, P. B. (2022). Saccadic eye movements are deployed faster for salient facial stimuli, but are relatively indifferent to their emotional content. *Vision Research*, *198*, 108054.

Webb, A. L., Hibbard, P. B., & O'Gorman, R. (2020). Contrast normalisation masks natural expression-related differences and artificially enhances the perceived salience of fear expressions. PloS one, 15(6), e0234513.

Westheimer, G. (1972). Visual acuity and spatial modulation thresholds. In Visual psychophysics (pp. 170-187). Springer, Berlin, Heidelberg.

Wilson, H. R. (1980). A transducer function for threshold and suprathreshold human vision. Biological Cybernetics, 38(3), 171-178.

Zeki, S. M. (1978). Functional specialisation in the visual cortex of the rhesus monkey. Nature, 274(5670), 423-428.

Zhaoping, L., & Li, Z. (2014). Understanding vision: theory, models, and data. Oxford University Press, USA.

Zhong, F., Jindal, A, Yöntem, Ö Hanji, P , Watt, S & Mantiuk R (2021) Reproducing reality with a high-dynamic-range multi-focal stereo display, ACM Transactions on Graphics, 40(6), 241.

Zibrek, K., Martin, S., & McDonnell, R. (2019). Is photorealism important for perception of expressive virtual humans in virtual reality?. ACM Transactions on Applied Perception (TAP), 16(3), 1-19.