Robert Earl Patterson

# Human Factors of Stereoscopic 3D Displays

Springer

# Human Factors of Stereoscopic 3D Displays

Robert Earl Patterson

# Human Factors of Stereoscopic 3D Displays

Springer

Robert Earl Patterson, Ph.D.
Dayton, OH, USA

ISBN 978-1-4471-6650-4 ISBN 978-1-4471-6651-1 (eBook)
DOI 10.1007/978-1-4471-6651-1

Library of Congress Control Number: 2014122930

Springer London Heidelberg New York Dordrecht


Printed on acid-free paper

Springer-Verlag London Ltd. is part of Springer Science+Business Media (www.springer.com)

*This book is dedicated to Alexxxis for, well, you know…*

# Acknowledgements

The author would like to thank Dr. John O. Merritt, Senior Consulting Scientist of The Merritt Group, for very helpful comments on a previous draft of this book and for helpful modifications to Figs. 1.1 and 1.2. The author would also like to thank Carl Douglass, of Argylebox, for helping create many of the figures in the book.

# Contents

# Chapter 1
# Introduction to Human Factors of Stereoscopic 3D Displays

This book presents a discussion of some of the fundamental human factors issues related to stereoscopic 3D displays. These issues determine how stereoscopic displays can be designed so they interface well with the human binocular visual system. In presenting this discussion, the idea embraced is that knowledge of the human binocular visual system is important when designing stereo 3D display systems.

People have extolled the virtues of stereoscopic 3D displays since the time of the Victorian era when Wheatstone (1838) discovered that binocular disparity was the cue for stereoscopic depth perception. Almost 100 years later, for example, Rowlands and Killian (1937) continued this praise for the scientific and entertainment virtues of stereoscopic displays. These virtues derive from the experience that depth perception induced by binocular disparity can be more compelling, robust, and immersive than depth perception induced by monocular cues (Patterson & Martin, 1992). The induction of strong immersive depth perception serves as the inspiration for creating these kinds of visual displays.

Not surprising, stereoscopic viewing may enhance performance on a number of tasks. Wickens, Todd, and Seidler (1989) suggested that stereoscopic viewing may enhance performance in domains such as air traffic control, telerobotics, computer-aided design, and meteorology. Muhlbach, Bocker, and Prussog (1995) reported that stereoscopic viewing can increase the feeling of telepresence (feeling of sharing space with a remote site). Servos, Goodale, and Jakobson (1992) reported that grasping movements were more accurate with binocular vision than with monocular vision. Sollenberger and Milgram (1993) found that stereoscopic or rotational display techniques increased accuracy on a 3D path-tracing task over two-dimensional (2D) displays. Performance was best when both techniques were combined. Ware and Mitchell (2008) reported graph comprehension was greater with a high resolution stereoscopic display. Kooi (2011) showed that stereoscopic depth can be used for perceptually segregating elements in a scene and reducing display clutter.

Interestingly, stereoscopic displays have been employed in medicine (Van Beurden, van Hoey, Hatzakis, & Ijsselsteijn, 2009). Falk, Mintz, Grunenfelder, Fann, and Burdon (2001) found that, with minimally invasive surgery, tasks



R.E. Patterson, *Human Factors of Stereoscopic 3D Displays*,
DOI 10.1007/978-1-4471-6651-1_1

involving certain motor skills were performed faster with binocular than monocular viewing. These authors suggested that future surgical systems should focus on 3D visualization because it promotes faster and more precise movement. In medical diagnosis, the advantage of stereoscopic displays for depicting complex spatial structures can enhance the discovery of abnormalities. Getty and Green (2007) proposed that the future use of stereoscopic digital imaging should improve the detection, diagnosis, and treatment of disease, and enhance the training of medical professionals. They mention that stereoscopic displays may be useful for breast cancer screening, diabetic retinopathy, and the teaching of anatomy. Votanopoulos, Brunicardi, Thornby, and Bellows (2008) reported that 3D imaging offers significant advantages when teaching laparoscopic skills to inexperienced individuals.

But not all studies have shown large gains in task performance from stereo viewing. Merritt (2011) provides an analysis of various factors involved in stereo display viewing which likely contributed to the reason why some studies have shown significant performance gains with stereo displays while other have not. According to Merritt, in many of the studies that showed only modest or no performance benefits for stereo viewing over 2D viewing, the potential gains were likely limited by adverse human factors issues (e.g., excessive mismatch between vergence and accommodation; unnecessary deviation from ortho-stereo camera/display geometry), by the use of visual tasks did not require good binocular depth perception, or by inappropriate experimental designs that limited statistical significance of results. In the studies that demonstrated significant performance gains for stereo displays, the gains were realized by well-designed stereo-image capture and display systems, by visually demanding tasks, and by use of appropriate experimental designs.

Similarly, McIntire, Havig, and Geiselman (2012) reported that stereoscopic displays can enhance performance on many depth-related tasks relative to 2D displays, such as judging distance, breaking camouflage, or manipulating objects through space. These authors suggested that, while stereoscopic displays can improve the spatial understanding of 3D scenes and environments, relatively simple tasks that do not require accurate depth perception may yield small or no benefits with stereoscopic viewing. More complicated tasks that require accurate depth perception may show large benefits with stereoscopic viewing.

In a subsequent review of the human performance literature, McIntire, Havig, and Geiselman (2014) reported that stereoscopic displays improved performance over 2D displays in 60 % of the studies they reviewed, showed a marginal benefit or unclear or mixed results in 15 % of the studies, and offered no benefit over 2D displays in 25 % of the studies. Stereoscopic displays were found to be most useful for tasks involving the manipulation of objects or the identification or classification of objects. Wickens, Hollands, Banbury, and Parasuraman (2013) suggested that the advantages of stereoscopic displays are greatest when visibility is degraded, there is high scene complexity, and only a few monocular cues are present.

Thus, stereo viewing can be very beneficial and enhance performance on a wide variety of tasks that are visually demanding and require accurate depth perception. But exactly what are stereo displays and what are their key features that distinguish them from other types of displays? We now turn to the specific features of stereo displays.

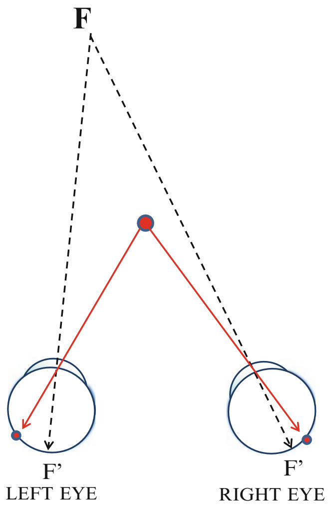


**Fig. 1.1** Schematic of binocular disparity. The drawing depicts a perspective view of the left and right eyes fixating stimulus F in the distance, which projects images to the fovea of the two eyes (F′). The *red disk*, representing a stimulus positioned in depth in front of stimulus F, projects images to the two eyes' retinal areas that are laterally shifted relative to each other (i.e., disparate retinal areas). The image from the disk falls to the left of the fovea (F′) in the left eye and to the right of the fovea (F′) in the right eye. This shift is binocular disparity (crossed disparity in this example)

Stereoscopic displays create depth perception via a cue called *binocular disparity* (some authors call this cue 'retinal disparity'). Binocular disparity refers to a lateral shift or difference between the spatial positions of corresponding left-eye and right-eye images, which is produced by viewing the relative z-axis depth among objects in the visual field (see the red disk and its left-eye and right-eye images in Fig. 1.1). Depth perception from binocular disparity is called stereopsis, or stereoscopic depth perception. This topic is discussed more fully in Chap. 2.

It has been recognized for some time that a stereopsis-like perception of depth can be experienced with single pictures, given certain viewing conditions. Such conditions include looking at a picture with one eye through a pinhole, or looking at a large picture from a great distance (Ames, 1925; see also Schwartz, 1971; Schlosberg, 1941). As discussed by Siegel, Tobinaga, and Akiya (1999), many of these kinds of effects likely have their origin in the suppression of the binocular sense that the picture is flat, which allows the brain to experience the depth based on an understanding of scene content. While interesting in their own right, I consider these kinds of effects to be relatively unique and they will not be discussed further in this book.

The lateral shift in the position of corresponding monocular images in the two eyes, binocular disparity, is created by binocular parallax. Binocular parallax refers to the apparent lateral displacement of the perceived position of objects when viewed along the left-eye versus the right-eye lines of sight, which is produced by the horizontal separation of the eyes. Figure 1.2 shows binocular parallax.

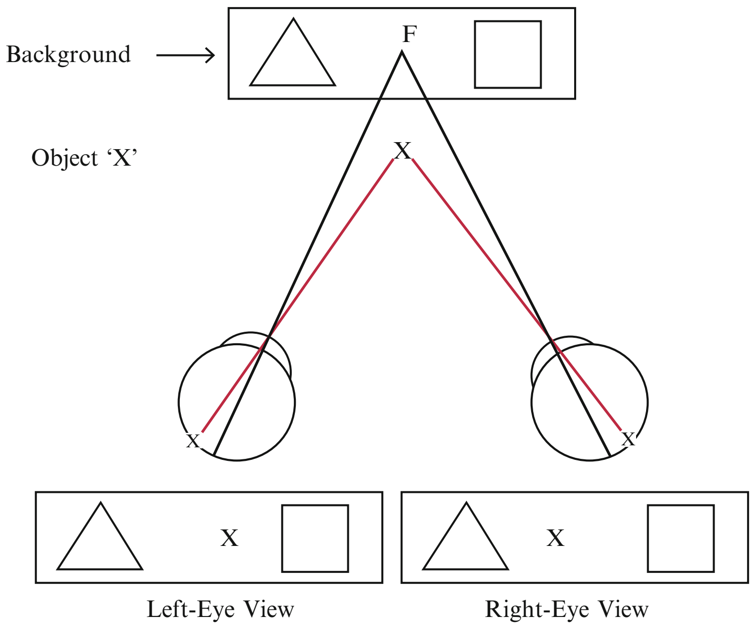


**Fig. 1.2** A depiction of binocular parallax. Top-down view of left-eye and right-eye viewing object X against a background of two shapes (*triangle* and *rectangle*). In the left eye's view, object X appears closer to the *rectangle*. In the right eye's view, object X appears closer to the *triangle*. This apparent displacement of the position of object X when viewed along the left-eye versus the right-eye line of sight, produced by the horizontal separation of the eyes, is binocular parallax

With stereoscopic displays, the purpose is to create left-eye and right-eye retinal images with a horizontal spatial shift between them (binocular disparity) that would normally occur with binocular parallax. Such spatial shifts can be simulated, for example, on a flat display surface. To keep the two eyes' views separate, the views can be presented with a spatial-multiplexing method, in which the two eyes' views are spatially interleaved. The two eyes' views can also be kept separate via a temporal multiplexing method (also called field sequential), in which the two eyes' views are temporally interleaved by presenting each eye's view on alternate frames of the display. Spatial multiplexing and temporal multiplexing are discussed briefly in Chap. 3. For a detailed discussion of the design of stereoscopic display systems, including spatial and temporal multiplexing, see Lueder (2012).

This book presents topics relating to the human factors of stereoscopic displays. Specific engineering systems such as head-worn displays, or specific types of imagery, or topics other than stereopsis (e.g., binocular versus monocular sensitivity) are not discussed. Certain designs of stereo display systems are discussed only to the point necessary for understanding a given human factors topic.

The ideas and discussions in the remainder of this book draw heavily upon the basic vision science literature and some cognition literature, yet many articles, papers, and books found in the applied literature are also covered. The book is broken into three major parts.

Part I provides the necessary background information, covering the basics of human stereopsis (Chap. 2) and the stimulus arrangement for creating stereoscopic displays (Chap. 3). This part forms a useful primer for those new to the field, and also serves as a refresher for readers with more background, and introduces various concepts and required terminology.

Part II forms the core of the book, introducing and discussing the various human factors issues that affect stereo depth perception: low-level factors (Chaps. 4 and 5); textual factors (Chaps. 6 and 7); and high-level factors (Chaps. 8 and 9).

In Part III a number of recommendations for stereoscopic display design are presented (Chap. 10). These recommendations will serve as a useful guide for those involved in the design, implementation and testing of stereoscopic display systems.

# Part I
# Background Information

# Chapter 2
# Basics of Human Binocular Vision

This chapter presents the basics of human binocular vision: the longitudinal horopter, horizontal binocular disparity, binocular disparity gradients, binocular rivalry, spatio-temporal frequency processing, and visual pathways. Vertical disparity will not be discussed; for discussion of vertical disparity, see papers by Tyler (1983) and Tyler and Scott (1979).

## 2.1 Horopter and Binocular Disparity

Figure 2.1 depicts a top-down view of two eyes looking out into the visual field and fixating on stimulus F. Now imagine an arc that passes through the fixation point called the "horopter". Positions along the length of the horopter define the locations of objects out in the visual field that give rise to pairs of left- and right-eye images that stimulate corresponding retinal points in the two eyes. Those locations possess zero binocular disparity. The horopter is important because it defines a set of baseline locations in space from which relative depth is judged. This is an important point: stereopsis is not perceived depth relative to the user; rather, stereopsis is perceived depth relative to the horopter (see Fig. 2.1).

The horopter defined empirically by psychophysical measurements is not the horopter defined geometrically (Ogle, 1964; Shipley & Rawlings, 1970). The geometric horopter is called the *Vieth-Muller* circle, which passes through the nodal points of the two eyes and the point of fixation (not shown in Fig. 2.1). The empirical horopter is based on several different criteria such as common perceived direction (nonius method) or the equidistant plane. The nonius horopter is the more appropriate measure from a physiological perspective (Shipley & Rawlings, 1970). Reference to the 'horopter' in this book will be referring to the nonius horopter.

The horopter is a reference or baseline depth plane passing through fixation from which the depth of objects located in other depth planes is judged. If these other



R.E. Patterson, *Human Factors of Stereoscopic 3D Displays*,
DOI 10.1007/978-1-4471-6651-1_2

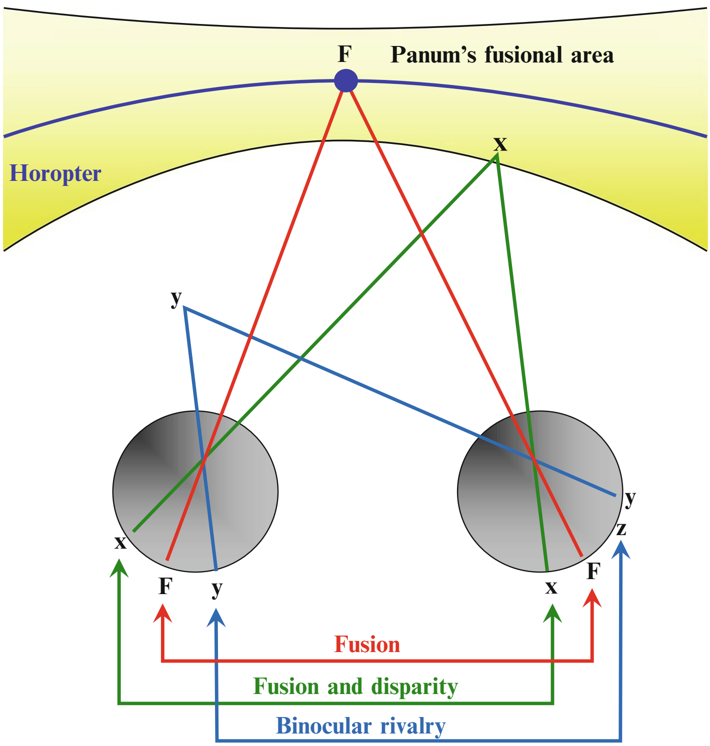


**Fig. 2.1** Drawing depicting the basics of stereoscopic viewing. The two *circles* represent a top down view of the two eyes, fixation point F, the horopter passing through the fixation point, Panum's fusional area, and object X and object Y. When fixation point F is fixated, as shown in the drawing, the images from F stimulate corresponding retinal points (foveae) in the two eyes and are fused. Object X is positioned in front of the horopter and thus carries a crossed disparity, but the images from X which stimulate non-corresponding (disparate) retinal points in the two eyes are fused because X is located within Panum's fusional area. Object Y is positioned farther in front of the horopter and also carries a crossed disparity, but the images from Y which also stimulate disparate retinal points in the two eyes are seen as diplopic (double) because Y is located outside Panum's fusional area. Because Y carries a large crossed disparity, and thus its two retinal images are shifted on very disparate retinal areas, the image of Y in the left eye may stimulate a retinal area that is corresponding to an area in the right eye which is stimulated by an image z from a different object in the visual field (not shown), which would provoke binocular rivalry. Reproduced from Figure 1 of Patterson (2007), Human factors of 3D displays, Journal of the Society for Information Display, 15, 861–871. Copyright Society for Information Display. By permission

objects are not positioned along the horopter, then they possess a non-zero magnitude of disparity. Objects positioned in depth in front of, or behind, the horopter give rise to pairs of images that stimulate non-corresponding, or disparate, retinal points in the two eyes. Objects positioned in depth in front of the horopter give rise to crossed disparity, and objects position in depth behind the horopter give rise to uncrossed disparity. These terms, crossed disparity and uncrossed disparity, are used as labels for the relative direction of the depth of objects from the horopter, either in front of (crossed) or behind (uncrossed) the horopter.

There is a spatial zone surrounding the horopter (yellow in Fig. 2.1) called Panum's fusional area. Objects positioned in Panum's fusional area will give rise to left- and right-eye images that are perceptually fused and seen as single objects. Objects positioned outside Panum's area will give rise to left- and right-eye images that cannot be perceptually fused and thus are seen as double images (diplopia). The ability to fuse disparate images depends upon disparity magnitude; the largest disparity at which fusion occurs is called the *disparity limit of fusion*. This limit is measured with the diplopia threshold—the threshold at which fusion is lost and double images are perceived. This can be achieved by having an observer report whether he or she perceives a briefly-exposed (duration = 160 ms) stereo stimulus as single and fused, or as two unfused images, as the disparity magnitude of the stimulus is systematically varied. The stereo stimulus would be presented briefly so that vergence eye movements could not alter the sign or magnitude of disparity—the latency of vergence eye movements is about 160 ms.

Many factors affect the disparity limit of fusion (see Arditi, 1986, for review). There are factors that affect this limit that display designers could manipulate or control, such as stimulus size and stimulus retinal eccentricity, as shown in Table 2.1.

The terms patent stereopsis and qualitative stereopsis are sometimes used (e.g., Ogle, 1964). Patent stereopsis refers to the interval of z-axis depth over which perceived depth increases monotonically with disparity magnitude, either away from the horopter with increasing crossed disparity (stimulus moving toward the observer), or away from the horopter with increasing uncrossed disparity (stimulus moving away from the observer). As the limit of patent stereopsis is approached, binocular fusion is lost and double images (diplopia) are seen—the stimulus is now outside Panum's fusional area (Fig. 2.2). Outside the range of patent stereopsis, depth perception with diplopia is called qualitative stereopsis and perceived depth becomes unreliable: further increases in crossed or uncrossed disparity continue to produce diplopia and perceived depth collapses inward toward the horopter. With diplopia, one of the two monocular images may be perceptually suppressed via a process called binocular rivalry (see below). Stimulus size and stimulus retinal eccentricity also affect the disparity limits of patent and qualitative stereopsis, again shown in Table 2.1.

**Table 2.1** Two stimulus sizes and two retinal eccentricities and their effects on the disparity limits for binocular fusion, patent stereopsis, and qualitative stereopsis

| | Small size (<15 arcmin) | | | Large size (1.0–6.6°) | | |
|---|---|---|---|---|---|---|
| | Fusion | Patent | Qualitative | Fusion | Patent | Qualitative |
| Foveal area | 10 arcmin | 20 arcmin | 2° | 20 arcmin | 2° | 8° |
| 6° Eccentricity | 20 arcmin | 2° | 3.5° | – | – | – |

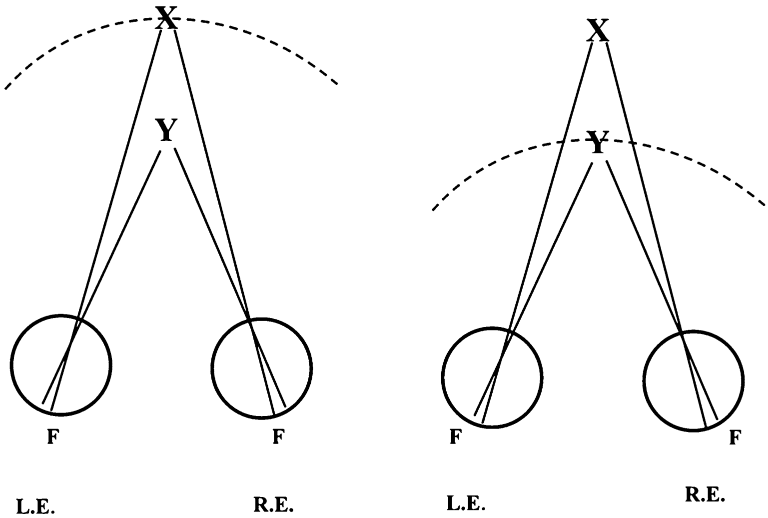


**Fig. 2.2** Depiction of relative versus absolute disparity. In the *left panel*, the observer is fixating on object X, whose images stimulate the fovea in the two eyes ('F') and thus object X projects a zero disparity value to the visual system (stimulation of corresponding retinal points). The *curved dashed line* shows the horopter going through object X. Object Y projects a crossed disparity to the visual system. In the *right panel*, the observer converges the eyes and shifts fixation to object Y, whose images now stimulate the fovea in the two eyes ('F') and thus object Y now projects a zero disparity value to the visual system. The *curved dashed line* shows the horopter going through object Y. Object X now projects an uncrossed disparity to the visual system. In both cases, the relative disparity between objects X and Y remains the same

When vergence eye movements are made, fixation and the horopter are shifted to various positions in the visual field. An object that initially stimulates the binocular visual system with crossed disparity may end up stimulating the visual system with uncrossed disparity, or vice versa. *In this case*, *the relative disparity between stationary objects in the visual field remains constant but their absolute disparity as projected to the visual system*, *which is the relevant cue for stereopsis* (Cumming & Parker, 1999), *will change whenever vergence eye movements are executed* (Patterson, 2007). See Fig. 2.2. This is an important point: if you want to know precisely the disparity sign and magnitude that is stimulating the visual system, then you must know where the observer is looking.

Vergence eye movements may serve to increase the disparity range over which reliable depth perception occurs. A mental representation of the visual field may be constructed over time by the integration of depth information across vergence eye movements (Patterson et al., 2006; Patterson & Martin, 1992).

Voluntary eye movements have been shown to increase the disparity limits of fusion, from a limit of about 24–27 arcmin without eye movements to a limit of several degrees with eye movements (Yeh & Silverstein, 1990). Voluntary eye movements have been shown to improve stereoscopic depth perception (Foley &

Richards, 1972). However, in other ways the effects of eye movements on stereopsis are complex: the longitudinal horopter is normal in the frontoparallel plane with symmetric convergence, but the horopter rotates horizontally with asymmetric convergence (i.e., fixation off the midsagittal plane; Ogle, 1964; Shipley & Rawlings, 1970), which would change the regions in the visual field that would support fusion and stereopsis.

It is commonly thought that vergence eye movements produce a conflict with accommodation when stereo displays are viewed: When viewing a stereo display, the stimuli for accommodation are images on the surface of the display. When a user changes vergence angle to converge to a virtual object appearing in depth in front of or behind the display, the vergence angle can be mismatched relative to the accommodative response. In Chap. 4, this issue is discussed in detail; such a conflict should occur only for short viewing distances and thus not be a general problem when vergence eye movements are made; a general remedy for the problem is given in Chap. 4.

The perception of relative depth from disparity is different from depth perception based on an observer making vergence eye movements. Although a change in vergence angle can be induced by variation in disparity, changes in vergence would provide only indirect information about relative depth. Relative depth in this case would be given by the sensing of a difference between two vergence positions via proprioception, not from disparity directly. Depth estimates from proprioception would be relatively imprecise compared with those from disparity (Patterson & Martin, 1992). Information from proprioception may augment the perception of depth, which is discussed in Chap. 9.

## 2.2 Binocular Disparity Gradient

A concept called the 'binocular disparity gradient' is important for achieving binocular fusion. The binocular disparity gradient can affect the ability of an individual to binocularly fuse and process multiple stimuli presented in stereoscopic depth. Given two objects that are laterally separated and also positioned in different depth planes, the binocular disparity gradient is defined as the difference in absolute disparity between the two objects divided by the mean angular separation between the combined images coming from one object versus the combined images coming from the other object (i.e., akin to the lateral separation between the objects). This concept is depicted in Fig. 2.3, which shows two viewing situations depicting a horizontal gradient of disparity.

In both panels of Fig. 2.3, the observer is fixating on object O and object X is positioned slightly to the left and behind object O. In the left panel, the two objects have a horizontal disparity gradient of less than 2, while in the right panel the two objects have a horizontal disparity gradient of 2. The critical disparity gradient is 1.0, a value above which the two disparities cannot be simultaneously fused (Burt & Julesz, 1980; Tyler, 1973). Burt and Julesz (1980) reported that when two objects

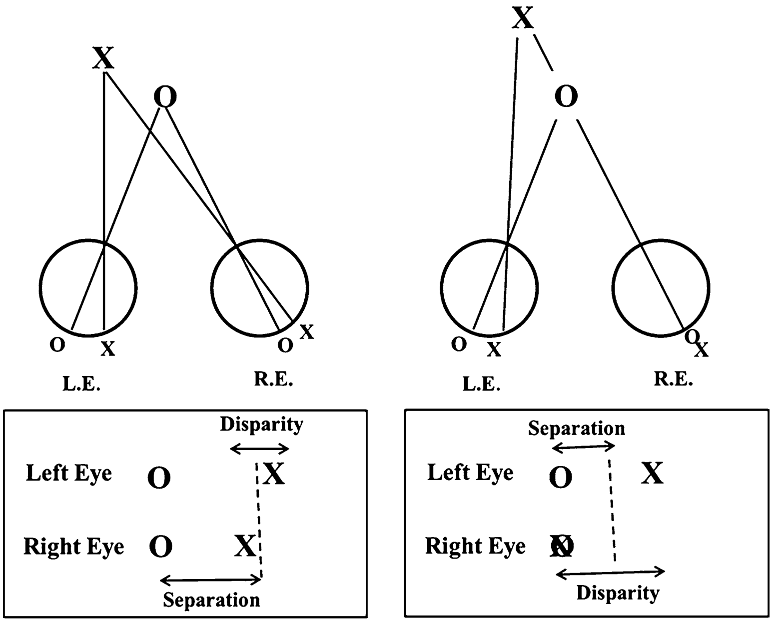


**Fig. 2.3** Depiction of the binocular disparity gradient. Top-down view of two eyes (L.E. = left eye; R.E. = right eye) viewing two objects in the visual field, Object 'O' (fixated) and Object 'X.' The two lower boxes give an analysis of the disparity and separation in each drawing. Disparity gradient is defined as the angular disparity between the images of two objects divided by the angular separation. Angular separation is defined as the angle between the mean direction of the images of one object and the mean direction of the images of the other object (mean direction is given by the *vertical dashed lines* in the lower boxes). The two objects O and X in the *left panel* have a disparity gradient of less than 2, while the two objects in the *right panel* have a disparity gradient of 2. Reproduced from Figure 2.7 of Howard and Rogers (1995), Binocular vision and stereopsis. Oxford, UK: Oxford University Press. By permission of Oxford University Press, USA

have a disparity gradient greater than 1 (i.e., the disparity exceeds the mean angular separation between the images of the objects), only one object can be perceptually fused at a time. The disparity of multiple objects located in different depth planes can be more easily fused if the objects have a sufficient horizontal and/or vertical separation between them when viewed from the observer's position.

The disparity gradient may affect the visual system's ability to fuse disparity information when stereo displays are viewed if objects with relatively large disparities are located too close together in the x/y-plane when viewed from the observer's location. This means that it may be prudent to keep objects with relatively large disparities sufficiently separated in the x/y-plane if possible. If not, then the observer may experience loss of fusion with one or more of the objects. This topic deserves to be investigated empirically with the kind of stereo displays employed in real-world applications to determine how serious the issue may be with real-world viewing.

## 2.3 Binocular Rivalry

There can be a potential problem when viewing a given stereo display whenever the images for the left and right eyes, coming from different display channels and/or optical systems, are misaligned or distorted to the point that observers cannot perceptually fuse portions of the two eyes' views. When an individual who views a stereo display cannot perceptually fuse portions of the two eyes' views, a visual process called binocular rivalry will be provoked. Binocular rivalry (Blake, 1989, 2001; Breese, 1899; Howard, 2002; Howard & Rogers, 1995; Levelt, 1965; Patterson, Winterbottom, Pierce, & Fox, 2007) refers to a state of competition between the eyes, such that one eye inhibits the visual processing of the other eye, when the two eyes view discordant stimuli. The visibility of the images in the two eyes fluctuates, with one eye's view being visible while the other eye's view is rendered invisible and suppressed, which reverses over time.

Binocular rivalry can be elicited by differences in attributes or characteristics between the images seen by the two eyes, such as differences in orientation, hue, luminance, contrast polarity, form, size, and/or motion velocity. Binocular rivalry can occur over a wide range of light levels throughout the visual field (Blake, 2001, pp. 8–9). The visual tolerance levels for interocular differences in stimulation are (Rash, Mozo, McEntire, & Licina, 1996; Tsou & Shenker, 2000): ±23 arcmin horizontal and ±11.5 arcmin vertical, horizontal or vertical differences in image size of up to 1.5 %, a rotational difference of up to ±10–12 arcmin, and deviation between centers of the two eyes' views of 0.18 prism diopters.

Patterson et al. (2006) also suggested that an interocular difference in luminance of greater than 30 % will likely provoke rivalry. Moreover, a given stimulus viewed by one eye will typically dominate a rival stimulus seen by the other eye if the former possesses greater contour density, higher contrast, a wider range of spatial frequencies, or faster motion. Practice over a number of days (e.g., 10 days) may help individuals control the rate of rivalry alternations (Lack, 1969). As an example, Fig. 2.4 depicts left-eye and right-eye views of stimuli that would provoke vigorous binocular rivalry due to differences in orientation and size (i.e., spatial frequency) of the bars making up the patterns.

When wavelengths are different enough to produce percepts of different hues, such differences can provoke rivalry (Hollins & Leung, 1978). This has implications for stereo displays that employ the anaglyph technique in which the two eyes' views are separated via the use of different bands of wavelengths (e.g., red images to one eye, blue or green images to the other eye). The anaglyph technique may be prone to inducing binocular rivalry. In this author's experience, as much as 15 % of individuals with normal stereo vision may experience chromatic rivalry when the anaglyph technique is used.

The inhibition provoked by binocular rivalry occurs at many levels of the visual system (Blake, 2001), and it can make visual processing unstable, unpredictable, and impair the ability of observers to visually guide and direct attention to targets in the visual field (Schall, Nawrot, Blake, & Yu, 1993). It is important to ensure that


**Fig. 2.4** Left-eye view and right-eye view of stimuli that provoke binocular rivalry due to differences in orientation and size (i.e., spatial frequency) of the bars making up the patterns. To induce rivalry, set up a viewing arrangement (try a pocket mirror—and see Fig. 3.2) in which the left eye's view is presented only to the left eye, and the right eye's view is presented only to the right eye. The visibility of the images in your two eyes will fluctuate, with one eye's view being visible while the other eye's view is invisible and suppressed, which reverses over time. Note that rivalry can be induced with more subtle differences between the images seen by the two eyes; see text for details

the two eyes' views of a stereo display are fusable and that no rivalry is present owing to misalignment or distortion of the two display channels and/or optical systems.

## 2.4 Spatio-Temporal Frequency Processing

When discussing certain aspects of binocular vision and stereoscopic depth perception, it is necessary to cover the topic of the visual processing of spatial frequency and temporal frequency of luminance modulation. One of the basic visual abilities is the detection of luminance contrast in space and in time. The visual processing of spatial and temporal frequency of luminance modulation (contrast) is fundamental and is placed within the context of 'frequency filtering', the idea that the early visual system performs a kind of filtering operation on the spatial and temporal distribution of luminance within the visual field (e.g., modeled as Fourier analysis). Early visual processing can be modeled as a filtering operation that filters the retinal image based on the spatial frequency and temporal frequency content of stimulation, where frequency is defined as the rate of *luminance modulation*. At a higher stage of processing, the visual system is thought to integrate the frequency information into a composite that represents various objects and their movements.

The human visual system can process spatial frequencies within the range of about 30 cycles per degree (cyc/deg) of visual angle (20/20 vision) or higher at the high end, and down to about 0.1 cyc/deg at the low end, depending on conditions (Campbell & Robson, 1968; De Valois & De Valois, 1988). This range of spatial

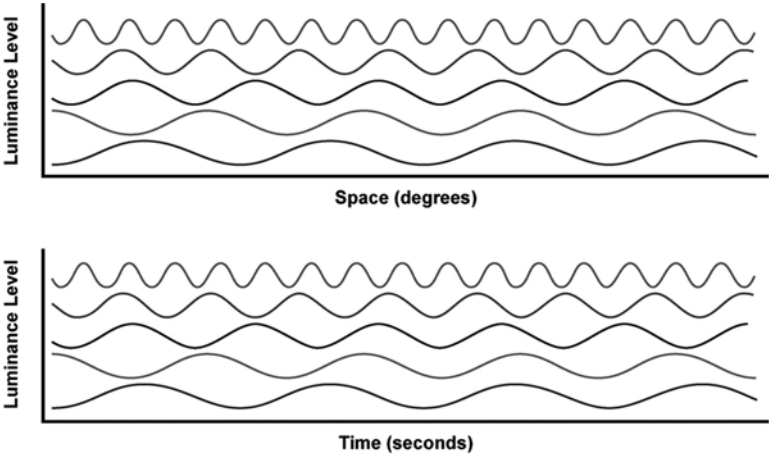


**Fig. 2.5** Depiction of different spatial frequencies (*upper panel*) and temporal frequencies (*bottom panel*). In the figure, the x-axis represents space (degrees of visual angle; *top panel*) or time (seconds; *bottom panel*) and the y-axis represents relative luminance level (i.e., absolute position along the y-axis is to be discounted). The *top panel* depicts different rates of luminance modulation across space, or different spatial frequencies (in units of cycles per degree, or cyc/deg), and the *bottom panel* depicts different rates of luminance modulation in time, or different temporal frequencies (in units of cycles per second, or Hz). In each panel, a range of frequencies is depicted, from a low frequency positioned at the bottom of each panel to a high frequency positioned at the top of each panel; the frequencies are offset from one another along the y-axis arbitrarily

frequencies is visually processed by different sets of neurons, each set of which responds to a given smaller band of frequencies (i.e., a spatial-frequency visual 'channel'). Some sets of neurons, high spatial frequency channels, respond to high spatial frequencies, which correspond to fine spatial detail. Such neurons would possess high spatial acuity (respond to the upper-most waveforms in the top panel of Fig. 2.5). Other sets of neurons, low spatial frequency channels, respond to low spatial frequencies, which correspond to coarse spatial detail. Such neurons would possess poor spatial acuity (respond to the lower most waveforms in the top panel of Fig. 2.5). The collection of channels represents neural responding to the entire range of spatial frequencies, responding which is integrated into various composite representations of objects and elements in the visual field at higher stages of processing.

The visual system can process temporal frequencies within the range of about 50–60 cycles per second (Hz) at the high end, and down to 0 Hz (i.e., a steady-state stimulus) at the low end, depending on conditions (De Lange, 1952, 1954; Kelly, 1971). Some sets of neurons (high temporal frequency channels) respond to high rates of temporal variation in luminance, which corresponds to high temporal acuity (respond to the upper waveforms in the bottom panel of Fig. 2.5). Other sets of neurons (low temporal frequency channels) respond to lower rates of temporal variation, which corresponds to poor temporal acuity (respond to the lower waveforms in the bottom panel of Fig. 2.5). At higher stages of processing, the visual system integrates temporal frequency information into various composite representations of movement and temporal structure.

Across the spatial-temporal frequency spectrum, sensitivity to high spatial frequencies is typically associated with sensitivity to lower temporal frequencies, and sensitivity to low spatial frequencies is associated with sensitivity to higher temporal frequencies. High spatial acuity/low temporal acuity are characteristics of a pathway that projects from the central retina to higher visual cortical areas (ventral cortical stream, or VCS) and which detects small binocular disparities (small disparity limit of fusion, fine depth discrimination). Low spatial acuity/moderate or high temporal acuity are characteristics of a pathway that projects from the central and peripheral retina to different higher cortical areas (dorsal cortical stream, or DCS) and which detects large disparities (large disparity limit of fusion, poor depth discrimination). These two pathways, VCS and DCS, are discussed in more detail below.

The spatial frequency-temporal frequency content of displayed information will determine the range of available disparities that can be fused and processed by the binocular visual system. This topic is discussed more fully in Chap. 6.

## 2.5 Visual Pathways

This section briefly covers the visual pathways in primate vision with a particular focus on stereo processing. The functional significance of these visual pathways for human factors issues will be discussed in subsequent chapters. To anticipate, we will learn, for example, that the high spatial acuity/low temporal acuity pathway (ventral cortical stream) that detects small binocular disparities, and therefore supports performance on tasks such as fine stereo depth discrimination, may be impaired by spatial multiplexing methods that entail decreased display spatial resolution. On the other hand, the low spatial acuity/moderate or high temporal acuity pathway (dorsal cortical stream) that detects large disparities, and therefore supports performance on tasks such as heading control, may be impaired by temporal multiplexing (field sequential) methods that involve decreased display temporal resolution.

In a basic sketch of the primate visual system (Fig. 2.6; see Blake and Sekuler, 2005, for overview), visual processing begins in the retina with light being transduced into neural signals by the rods and cones. From the retina, signaling projects to layers in the thalamus in an area called the lateral geniculate nucleus, or LGN. In this projection, cells from the inner half of the retina of the left eye (called the nasal hemi-retina) cross the midline of the body at the optic chiasma and combine with cells from the outer half of the retina from the right eye (called the temporal hemi-retina) and project to the LGN on the right side of the body (in the right hemisphere of the brain). Because the left half of the visual field projects onto the nasal hemi-retina of the left eye and the temporal hemi-retina of the right eye, both of which project to the LGN on the right side, *information located in the left visual field projects to the right side of the brain*. Cells from the nasal hemi-retina of the right eye cross at the optic chiasma and combine with cells from the temporal hemi-retina

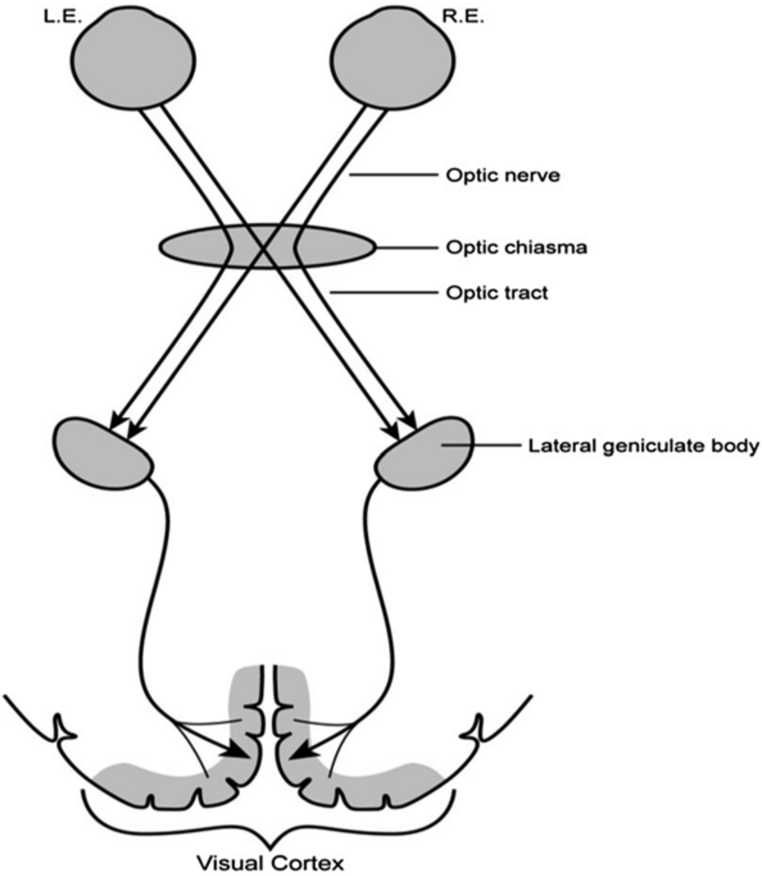


**Fig. 2.6** Basic sketch of a top-down view of the primate visual system. Visual processing begins in the retina (L.E. = left eye; R.E. = right eye). From the retina, signals project to layers in the thalamus in an area called the lateral geniculate nucleus (or body), or LGN. In this projection, cells from the inner half of the retina of the left eye (nasal hemi-retina) cross the midline of the body at the optic chiasma and combine with cells from the outer half of the retina from the right eye (temporal hemi-retina) and project to the LGN in the right hemisphere of the brain. Because the left half of the visual field projects onto the nasal hemi-retina of the left eye and the temporal hemi-retina of the right eye, both of which project to the LGN on the right side, information located in the left visual field projects to the right side of the brain. Cells from the nasal hemi-retina of the right eye cross at the optic chiasma and combine with cells from the temporal hemi-retina of the left eye and project to the LGN on the left side of the brain. The same organizational scheme thus applies to the right half of the visual field: the right half of the visual field projects to the LGN on the left side, thus information located in the right visual field projects to the left side of the brain. From the LGN, signals project into the occipital lobe of the visual cortex in area V1 in both hemispheres

of the left eye and project to the LGN on the left side of the brain. Because the right half of the visual field projects onto the nasal hemi-retina of the right eye and the temporal hemi-retina of the left eye, both of which project to the LGN on the left side, *information located in the right visual field projects to the left side of the brain*.

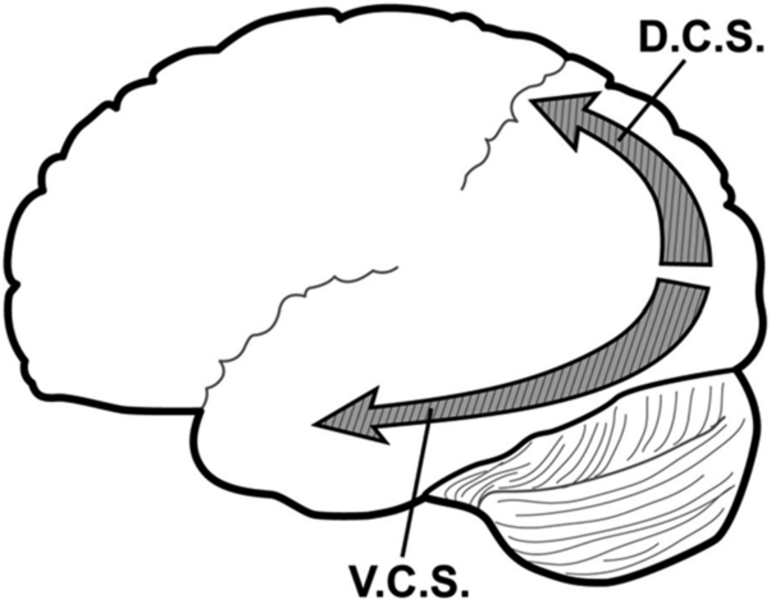


**Fig. 2.7** Drawing showing the left side of the human brain. The rightmost region of the drawing (near the origin of the two *arrows*) is the occipital cortex and area V1, from which two functionally distinct pathways emerge. The ventral cortical stream (V.C.S. in the figure) projects into the temporal lobe and is thought to be involved in the functional analysis of spatial pattern information and object identification. The dorsal cortical stream (D.C.S. in the figure) projects into the parietal lobe and is thought to be involved in the functional analysis of motion information, heading control during locomotion, and action priming. See text for detail

From the LGN, signals project into the occipital lobe of the cortex in area V1 in both hemispheres. (Other pathways projecting to subcortical structures will not be discussed.) From V1, signals project into area V2, also located in the occipital lobe (not shown). From V2, signals project to various areas in parietal and temporal cortex, such as areas V3, V4, and V5 (also not shown). The neural projections from the retina into visual cortex are thought to comprise two parallel pathways, the ventral cortical stream and the dorsal cortical stream, as shown in Fig. 2.7 (Livingstone & Hubel, 1988; Milner & Goodale, 1995; Schiller, Logothetis, & Charles, 1990; Ungerleider & Mishkin, 1982).

## 2.6 Parallel Pathways

The ventral cortical stream (VCS, or V.C.S. in Fig. 2.7) draws connections mainly from the central retina and projects to areas in visual cortex, such as areas V1, V2, and V4. Cortical areas in the ventral stream functionally analyze spatial pattern information. The neurons in this pathway have a sluggish and sustained response, high spatial acuity, poor temporal acuity, and chromatic sensitivity. The VCS may be involved in the functional analysis of spatial pattern information, object identification, and conscious perception (Milner & Goodale, 1995).

The dorsal cortical stream (DCS or D.C.S. in Fig. 2.7) draws connections from both the central and peripheral retina and projects to areas in visual cortex, such as V1, V2, V5, and MST. Cortical areas in the dorsal cortical stream process optic flow information for heading control (Peuskens, Sunaert, Dupont, Van Hecke, & Orban, 2001), biological motion (Grossman & Blake, 2001; Grossman et al., 2000), and integrate vision with action (e.g., T'so & Roe, 1995; Van Essen & DeYoe, 1995; Yabuta, Sawatari, & Callaway, 2001). The neurons in this pathway have a transient response, high temporal acuity, poor spatial acuity, and lack chromatic sensitivity. The dorsal cortical stream is thought to be involved in the functional analysis of motion information, heading control during locomotion, and action priming (Milner & Goodale, 1995).

Farivar (2009) suggested that the ventral and dorsal cortical streams are not fully dissociated, anatomically or functionally. The dorsal stream processes dynamic 3D shape cues, which would suggest that the dorsal stream plays a role in object recognition, a role originally thought to belong exclusively to the ventral stream. Farivar concluded that the dorsal stream extracts 3-D shape information from certain dynamic cues and relates those representations to cue-invariant and view-invariant representations of objects in the ventral stream, an interpathway interaction (see also Hegde & Felleman, 2007).

Both pathways in the primate visual system, ventral and dorsal, contain areas responsive to binocular disparity (e.g., Backus, Fleet, Parker, & Heeger, 2001; Burkhalter & Van Essen, 1986; Georgieva, Peeters, Kolster, Todd, & Orban, 2009; Patten & Murphy, 2012; Tsao et al., 2003). For example, in a functional magnetic resonance imaging (fMRI) study, Likova and Tyler (2007) reported that a region of the dorsal stream in human cortex was activated by a pure disparity-defined stimulus moving in the z-axis.

As mentioned above, the functional significance of disparity processing in both the ventral and dorsal streams for human factors issues will be discussed in subsequent chapters. We now turn to a discussion of the stimulus arrangement for creating stereoscopic displays.

# Chapter 3
# Stimulus Arrangements for Creating Stereoscopic Displays

Stereoscopic displays present binocular disparity on a display surface that is used for getting images projected onto the left- and right-eye retinas so that depth perception can be induced. To do so, slightly different images must be presented to the two eyes. There are several ways to present different images to the two eyes. The *spatial-multiplexing* approach involves presenting images to the two eyes so that they are spatially interleaved. This can involve simultaneous left-eye and right-eye presentations by using two regions on one 2D display, with each region seen by only one eye, or by using wavelength-multiplexing techniques. The *time-multiplexing*, or *field-sequential*, approach entails presenting images to the left and right eyes so that they are temporally interleaved. This involves having the left-eye and right-eye presentations on alternate frames of the display. The basics of these two approaches are discussed below; for more thorough discussion, see Lueder (2012).

## 3.1 Spatial-Multiplexing

One can grasp the basics of how stereo depth is simulated on a 2D display by considering Fig. 3.1. In Fig. 3.1, top panel (A), the basic stimulus arrangement for creating a stereoscopic display is shown. Here, the left-eye image (called half-image) and the right-eye image (also called half-image) are projected to the eyes of an observer who is fixating, in this example, the center bar of each eye's view. Notice that the separation between the two bars is greater in the right eye's view than in the left eye's view, creating a crossed disparity (i.e., depth in front of fixation) for the non-fixated bar. In Fig. 3.1, bottom panel (B), the perceptual result from this stimulus arrangement is shown. In this figure, the two half-images are stacked one on top of the other, which simulates a spatial multiplexing technique for stereo viewing (Patterson, 2007). When the half-images in the two eyes are binocularly fused, the non-fixated bar will appear in depth in front of the display with crossed



R.E. Patterson, *Human Factors of Stereoscopic 3D Displays*,
DOI 10.1007/978-1-4471-6651-1_3

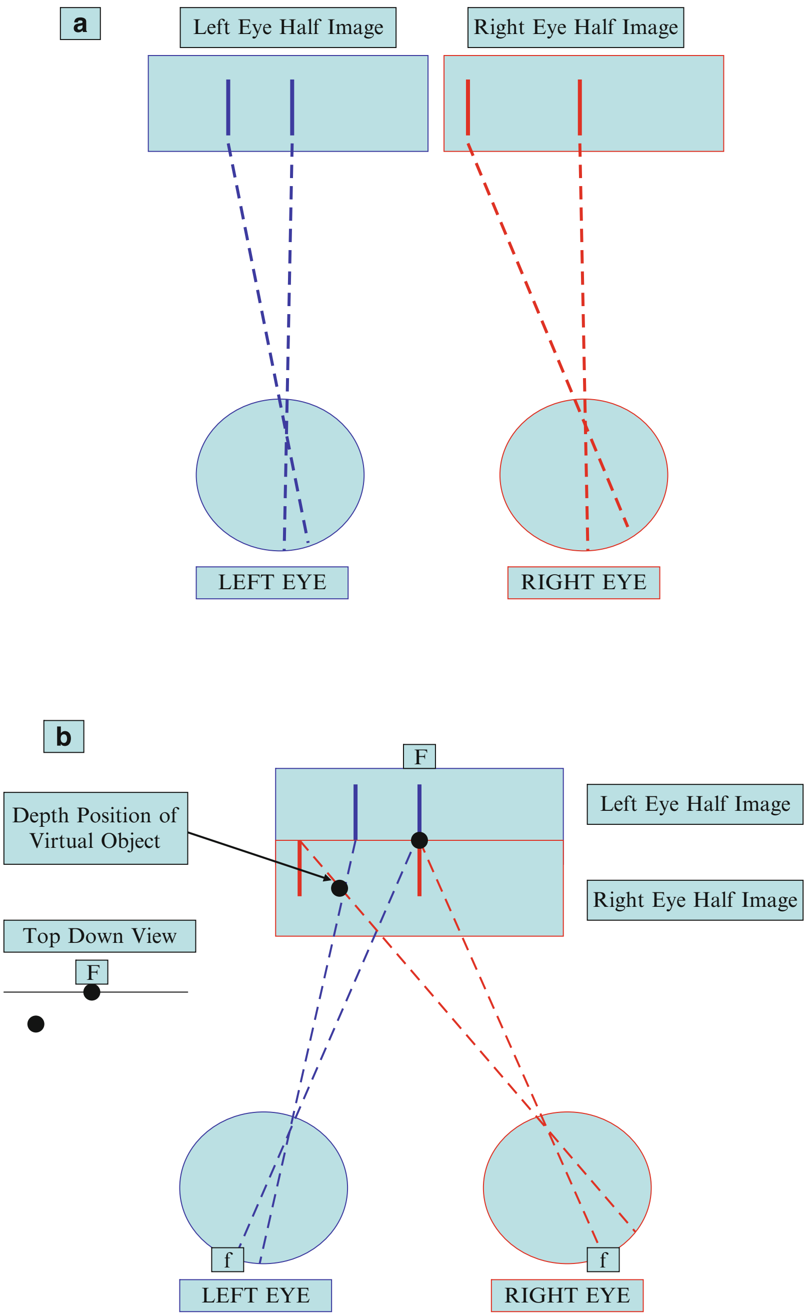


**Fig. 3.1** (**a**) Basic stimulus arrangement for creating a stereoscopic display; top down view. Left-eye image (called 'half-image' in the figure) and right-eye image (also called 'half-image') are projected to the eyes of an observer who fixates, in this example, the center bar in each eye's view. (The term 'half-image' is typically used in basic research on human stereopsis.)

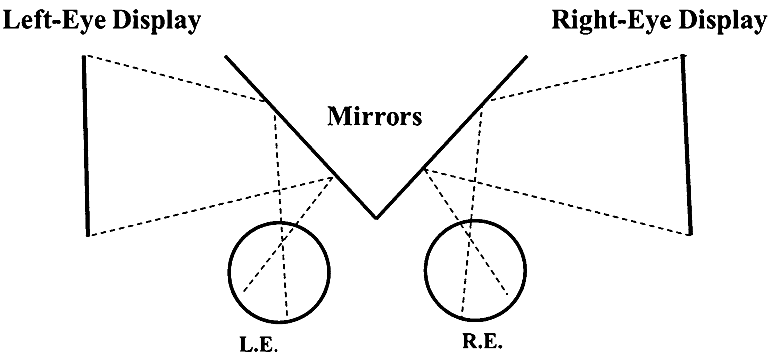


**Fig. 3.2** Top-down view depicting a mirror stereoscope (a spatial multiplexing technique). 'Left-eye display' and 'right-eye display' show the position of the two displays. 'Mirrors' shows the position of the two mirrors that direct the light rays to the two eyes

disparity. Different techniques are used for spatially separating the two eyes' views, such as lenticular displays or parallax barrier-type displays. Special eye glasses may have to be worn by the user, such as polarized lenses or chromatic lenses (the latter called the anaglyph technique), so that the left eye sees only the left-eye's view and the right eye sees only the right eye's view. As discussed by Lueder (2012), in many approaches used for spatial multiplexing, each view can exhibit only half the resolution of a 2D display.

As another example of spatial multiplexing, consider Fig. 3.2 which depicts a mirror stereoscope arrangement (note that, in this method, each eye's view has full resolution because the method does not involve spatial multiplexing in the usual sense). Here, each eye's view appears on a visual display, which is directed to the corresponding eye by way of a mirror. The vergence angle adopted by the eyes as they view the images reflected by the mirrors should be consistent with the distance that the light rays have to travel so that the vergence and accommodative demand are matched.

Problems with the spatial multiplexing approach include the need to maintain good spatial registration of the two eyes' views and in some cases (but not with a mirror stereoscope) the potential loss of spatial resolution.

**Fig. 3.1** (continued) The separation between the two bars is greater in the right eye's view than in the left eye's view, creating a crossed disparity (front depth) for the non-fixated bar. (**b**) Perceptual result from the stimulus arrangement shown in Panel A. Here, the two half-images are stacked one on top of the other, which simulates a spatial multiplexing technique for stereo viewing. The "F" shows that the center bar is fixated, and the "f" indicates the fovea in the two eyes. When the half-images in the two eyes are binocularly fused, the non-fixated bar will appear in depth in front of the display screen with crossed disparity (see insert labeled "Top Down View" on the left side of the diagram). Reproduced from Figure 1 of Patterson and Silzars (2009), Immersive stereo displays, intuitive reasoning, and cognitive engineering. Journal of the Society for Information Display, 17, 443–448. Copyright Society for Information Display. By permission

## 3.2 Temporal-Multiplexing

The time-multiplexing, or field-sequential, approach entails presenting images to the left and right eyes in temporal alternation. This is shown in Fig. 3.3 where LE depicts the left-eye's view and RE depicts the right eye's view. One can see that when one eye's view is turned 'on', the other eye's view is turned 'off', and vice versa. Special eye glasses will usually have to be worn by the user, such as shutter glasses that open and close in synchronization with the alternation of the display, so that the left eye sees only the left-eye's view and so on for the right eye. In the time-multiplexing approach, images can be presented to the left and right eyes in temporal alternation in a way that allows the full resolution of each eye's view to be maintained, but the addressing of the display must work at twice the speed of a typical 2D display (Lueder, 2012).

The temporal alternation can produce flicker that is visible to each eye. As explained by Banks, Read, Allison, and Watt (2012), there are several ways to present the alternating images in time. In a single-flash protocol, each image is

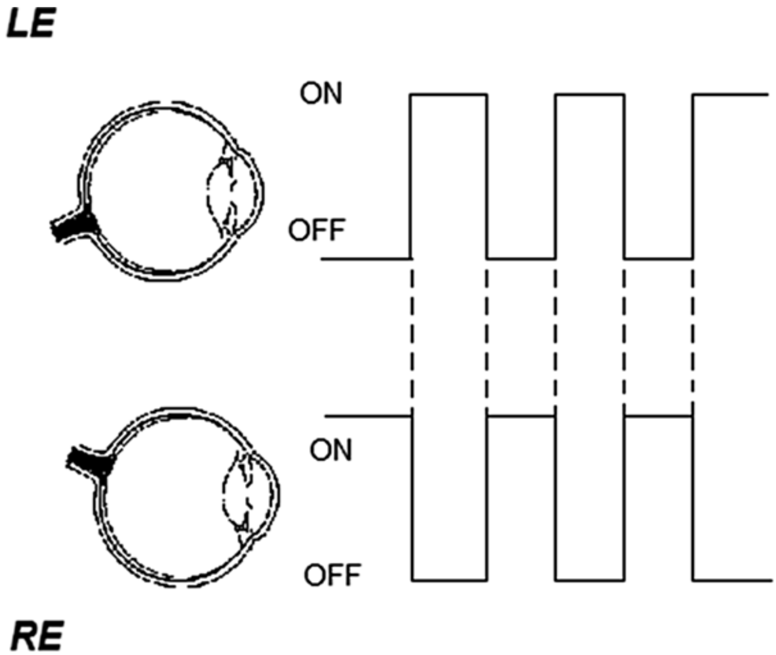


**Fig. 3.3** Drawing depicting the time-multiplexing (i.e., field sequential) method of stereoscopic presentation. LE = left eye; RE = right eye. Each square-wave trace shows the left eye's (LE) and right eye's (RE) on-off cycles of frames of a stereoscopic display; for each eye, the stimulus approximates square-wave flicker. The on-off cycles for the two eyes are in anti-phase such that when the frame for one eye is turned on, the frame for the other eye is turned off. Reproduced from Figure 2 of Patterson (2007), Human factors of 3D displays, Journal of the Society for Information Display, 15, 861–871. Copyright Society for Information Display. By permission

presented once. In the double-flash protocol, each image is presented twice before updating. In a triple-flash protocol, each image is presented thrice before updating. The double- and triple-flash protocols are used to reduce the visibility of flicker.

Banks et al. (2012) discussed how various temporal presentation methods affect the viewer's perception of flicker and motion artifacts with stereo displays. For flicker, presentation rate, not image update rate, should determine flicker visibility. Flicker visibility should be less for a fixed image update rate by using the multiflash protocol. For motion artifacts, image update rate, not presentation rate, should determine the visibility of motion artifacts. The visibility of motion artifacts should increase with increasing stimulus speed and decrease with increasing image update rate. Banks et al. also suggested that, for a fixed image update rate, multiflash protocols should produce more visible motion artifacts. These suggestions were supported in a study by Hoffman, Karasev, and Banks (2011).

Flicker and motion artifacts are issues attendant to visual displays in general; these issues are not peculiar to stereo display viewing. For stereo display viewing, visual temporal processing of stereo information is very sluggish, with the limit of temporal resolution of disparity information being about 8 Hz (Patterson, Ricker, McGary, & Rose, 1992). The visual system would be insensitive to rapid temporal change in disparity even if temporal change in luminance information was visible as flicker or motion artifact.

Banks et al. (2012) discussed how a temporal delay to one eye's input can cause a moving object to appear displaced in depth. This phenomenon is related to the Pulfrich effect (Pulfrich, 1922; Lit, 1949) that occurs when a laterally-moving object is binocularly viewed with a darkened filter over one eye: the object's lateral motion appears to have a depth component due to the relative difference in signal latency between the two eyes brought about by their difference in illumination level. The eye with the darkened filter perceives the object at a delayed position relative to the other eye. The difference in timing is interpreted as a spatial disparity between the two eyes' images, hence the perceived depth effect. (Binocular night vision goggles have this issue, which creates the need for the left-eye and right-eye tubes to be adjusted to within 10 % to avoid the spurious depth effects with moving targets.)

In the present case, the relative difference in signal latency between the two eyes is produced by the field-sequential technique. Objects moving in one direction can be perceived as being closer than they should be and objects moving in the opposite direction can be perceived as being farther. According to Banks et al. (2012), the largest distortion in perceived depth should occur with single-flash protocols and the smallest distortion with triple-flash protocols. The magnitude of the distortion can equal several arcmin in the worst cases.

For the time-multiplexing technique to work, visual signals in each of the two monocular pathways must persist long enough to bridge the temporal gaps in stimulation produced by having each eye stimulated intermittently. Each eye's response must persist through the 'dark' interval so as to blend with the other eye's response when its display is turned on, and vice versa, so as to adequately stimulate the stereo-processing visual system. Given that the visual system is composed of pathways with differing temporal integration properties (recall ventral versus dorsal cortical streams, last chapter), then the responding of these pathways may show a differing ability to persist and bridge the temporal gaps. This, in turn, would consign the time-multiplexing method to be used with those pathways whose responding can bridge such temporal gaps. For example, for certain displays, it may be that stereo processing within the faster responding, and less enduring (shorter visual persistence) DCS may be impaired by the intermittent stimulation, while stereo processing within the slower responding and more enduring (longer visual persistence) VCS remains unaffected.

Moreover, there are three forms of ‘visual’ persistence (Coltheart, 1980): (1) enduring neural activity following stimulus offset (neural persistence); (2) enduring visibility of a stimulus following its offset (visual persistence); and (3) enduring information about a stimulus following its offset (information persistence). It is not clear which of these forms of visual persistence would be relevant to the question of temporal integration of stereo information and time multiplexing posed here.

In short, it is unclear whether different rates of time-multiplexing differentially impact the pathways that subserve different stereo abilities, such as fine depth discrimination or the binocular fusion of large disparities. The effects of different rates of time-multiplexing on different stereoscopic abilities needs to be studied.

Now that the necessary background information has been covered (Chap. 2, human stereopsis, and Chap. 3, stimulus arrangement for creating stereoscopic displays), a discussion of the factors that affect depth perception in stereo displays will now be presented. Six low level factors will be covered (Chaps. 4 and 5), as will three contextual factors (Chaps. 6 and 7) and six high-levels factors (Chaps. 8 and 9). The book ends with concluding remarks, which present recommendations for stereoscopic display design (Chap. 10). A discussion of the low-level factors that affect the viewing of stereo displays is discussed next, which is in Part II of this book.

# Part II
# Factors That Affect Stereo Depth Perception in Stereo Displays

# Chapter 4
# Low-Level Factors

This chapter covers three low-level factors that affect stereo viewing: interocular crosstalk, accommodation-vergence conflict, and Percival's Zone of Comfort. These factors are considered 'low level' because they involve peripheral stages of visual processing: crosstalk entails image leakage on the retina, while accommodation-vergence conflict and Percival's Zone of Comfort involve oculomotor responding.

Before proceeding to a discussion of these three factors, it must first be mentioned that the most obvious and basic human factors guideline, that the left-eye's image should be delivered to the left eye, and the right-eye's image delivered to the right eye, is violated surprisingly often. For example, the two eyes' view may sometimes be inadvertently swapped at conferences, or in military labs testing subjects to determine the efficacy of 3D versus 2D for teleoperation. It is sometimes the case in such situations that many people who set up the stereo displays for viewing cannot readily perceive that the binocular depth is reversed (i.e., in the uncrossed rather than the crossed direction, or vice versa) because the monocular depth cues are so strong.

This issue can be readily fixed by having the presenter or experimenter set up the stereo viewing by simply looking at the display with the stereo glasses as designed, and then again with the two eyes' views swapped, to see which situation yields the best stereo depth perception. According to John Merritt (personal communication, July 28, 2014, see also Merritt, 2011), it is surprising that swapped left-eye and right-eye images occur in perhaps 25 % of these kinds of situations. In some cases special precautions and/or checklists could be used to avoid showing stereo content with the two eyes' images swapped, and warning practitioners that they might not notice such inadvertent swapping of the information delivered by the two eyes' channels without thorough testing of the viewing arrangement. We now turn to the issue of crosstalk.



R.E. Patterson, *Human Factors of Stereoscopic 3D Displays*,
DOI 10.1007/978-1-4471-6651-1_4

## 4.1 Interocular Crosstalk

Interocular crosstalk is created by image artifacts from one eye's view leaking into the other eye's view. Crosstalk is sometimes referred to as 'ghosting' (i.e., faint images from one eye's view seen in the other eye's view). Crosstalk (*ct*) can be defined as a proportion of the desired luminance (Lueder, 2012):

$$ct = \text{luminance of light from undesired image} / \text{luminance of light from desired image.}$$

This is probably the most serious human factors issue because crosstalk serves to introduce a form of binocular, or interocular, noise into the visual system (Yeh & Silverstein, 1990) that degrades stereopsis in all of its respects. Interocular crosstalk can occur with both spatial multiplexing and temporal multiplexing (i.e., field sequential) techniques.

### *4.1.1 Spatial Multiplexing*

Interocular crosstalk can occur with spatial multiplexing when the technology permits light from one eye's image to leak into the partner eye (Lueder, 2012), such as with chromatic aberration with lenticular displays, or diffraction with parallax barrier-type displays. There can be interocular crosstalk with autostereoscopic displays if an observer is positioned at an incorrect viewing distance. The remedy for these problems is to limit chromatic aberration or diffraction and position the observer at the correct viewing distance.

### *4.1.2 Temporal Multiplexing*

Interocular crosstalk can occur with temporal multiplexing (field sequential) techniques when there is significant display persistence (Lueder, 2012; Yeh & Silverstein, 1990). Significant display persistence allows portions of the image in each eye's view to persist past the termination of each frame and leak into the other eye when its view is exposed. From Fig. 3.3 of the previous chapter, recall that an approximation of the onset and offset of the two eyes' view can be depicted by square wave profiles in anti-phase. When one eye's image is turned on, the other eye's image is turned off. Any leakage from one eye's view persisting past its intended termination and into the other eye's view will create crosstalk. One remedy for this problem is to limit display persistence to very brief durations, ideally below 1 ms. Alternatively, one could attempt to measure the crosstalk and subtract it from each eye's image via custom software or hardware. Lueder (2012) discusses several techniques for minimizing crosstalk with the temporal multiplexing method that are beyond the scope of this chapter.

Yeh and Silverstein (1990) reported that, for time-multiplexing techniques, as little as 2–7 % of interocular crosstalk can degrade image quality and significantly reduce the limits of binocular fusion. Consistent with these values, Kooi and Toet (2004) found that as little as 5 % of crosstalk can produce viewing discomfort.

Wang et al. (2011) investigated the effects of contrast (ranging from a contrast ratio of 1:1 to 100:1) and binocular disparity magnitude (ranging from 8 to 40 arc-min) of a simple stimulus on crosstalk visibility and acceptability thresholds for two time-multiplexed stereoscopic displays. One display involved active shutter glasses and the other display involved passive glasses. Crosstalk annoyance increased with increasing contrast and increasing disparity magnitude equally for the two types of stereo displays. The maximum level of allowable crosstalk ranged from 9 to 17 % for very low contrast values and was reduced to 1 to 6 % for a contrast ratio of 100:1. Wang et al. argued that these values were possibly too strict for natural image content; however Hanazato, Okui, and Yuyama (2000) suggested that crosstalk levels should be 1–2 % in order to obtain adequate quality stereoscopic images.

Wang et al. (2012) investigated the effect of motion blur on the perception of crosstalk, and the latter's influence on image quality, when dynamic image sequences were rendered on two types of stereoscopic displays. One display was a hold-type liquid-crystal display with patterned retarder and circular polarizing glasses. The other display was an impulse-type digital light-processing display with shutter glasses. Wang et al. found that perceivable motion blur was induced with the liquid-crystal display but not with the digital light-processing display. Crosstalk was less visible and more acceptable with images moving with a higher speed on the hold-type LCD because the crosstalk-induced quality degradation was partly masked by motion blur. Crosstalk degraded the quality of moving images less than it degraded the quality of still images; subjects were more sensitive to crosstalk in the still images (Wang et al., 2012).

A number of factors can influence crosstalk and its effects on display quality, such as the type of stereo display technology, stimulus contrast, disparity magnitude, still versus moving images, and the speed of moving images. Generally, it appears that for high-quality stereo viewing, crosstalk should be less than 2 %. Thus, the recommendation here is that crosstalk should be kept to less than 2 %. Woods (2010) reviewed many of the factors that produce crosstalk in stereoscopic displays, including the time-sequential method implemented on LCDs and the autostereoscopic method. Woods discussed crosstalk reduction and crosstalk cancellation along with methods for measuring and characterising crosstalk.

## 4.2 Accommodation-Vergence Conflict

When viewing a stereo display, the stimulus for accommodation would be the images on the surface of the display. However, when an observer converges to a virtual object appearing in depth in front of or behind the display, the vergence angle can be in conflict with accommodation (Wann, Ruston, & Mon-Williams 1995; see Fig. 4.1).

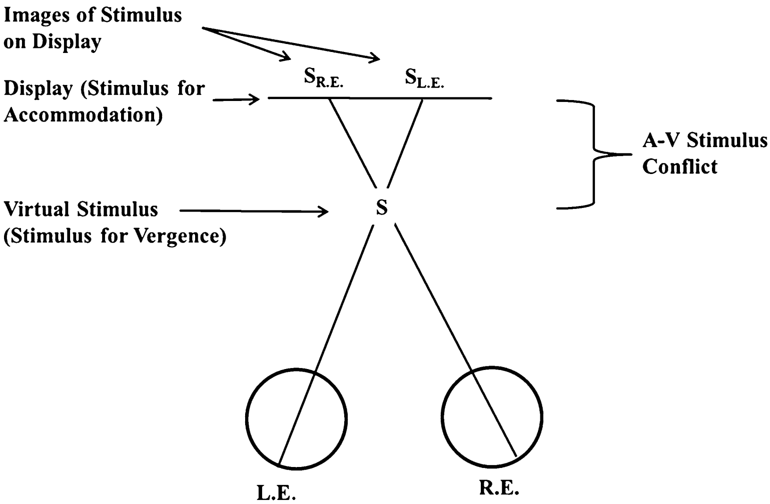


**Fig. 4.1** Diagram depicting accommodation-vergence stimulus conflict. The figure shows a top down view of two eyes (L.E. = left eye; R.E. = right eye) viewing a visual display. The two eyes are converging upon a virtual stimulus 'S' presented with stereoscopic depth. The displayed left-eye image of the virtual stimulus is $S_{L.E.}$ and the right-eye image is $S_{R.E.}$. Because the eyes are converged at S, the horopter will go through S and S will project a zero disparity to the visual system; thus the display itself will present an uncrossed disparity to the visual system. Also shown is the accommodation-vergence stimulus conflict ('A-V stimulus conflict'), which is defined as the z-axis distance between the fixated virtual stimulus (to which vergence is directed) and the visual display on which the images are presented and presumably where accommodation is driven

Due to the synergy between accommodation and vergence (Toats, 1972, 1974), it is believed that converging or diverging to a depth plane that is different from the display surface will pull the accommodative response to that depth plane, thereby making the images on the display surface become blurred. This, in turn, would tend to drive the accommodative response back to the display surface, and a conflict between accommodative responding and vergence eye movements would ensue. It is believed that an accommodation-vergence conflict can create eyestrain, visual discomfort, and visual fatigue (Hoffman, Girshick, Akeley, & Banks, 2008; Shibata, Kim, Hoffman, & Banks, 2011; Velger, 1998; Wann et al., 1995). Lambooij, Ijsselstein, Fortuin, and Heynderickx (2009) suggested that rapid changes in the demand placed on the accommodative-vergence linkage can create visual discomfort. However, these authors also suggested that the classical notion of accommodation-vergence conflict appears to be of minor importance when disparity values do not exceed 1 arcdeg when viewing stereoscopic television. Interestingly, Lambooij, Fortuin, IJsselsteijn, and Heynderickx (2009) suggested that a combination of fusion range measurements and self-report is appropriate for evaluating visual fatigue related to stereo displays.

The logic underlying the concept of accommodation-vergence conflict is centered on oculomotor *responding*. A conflict presumably occurs because converging to a depth plane different from the display surface would pull the accommodative *response* to that depth plane, thereby making the images on the display surface go out of focus. Blurred images would then drive the accommodative *response* back to the display, which would pull vergence (an eye movement *response*) back to the display, and so on. However, many studies have conflated a conflict in *stimulation* (shown in Fig. 4.1) with this assumed conflict in responding. But when one examines the situation more closely, it is unlikely that a conflict in responding will always occur.

Patterson (2009) discussed how an accommodation-vergence conflict should be less of a problem than commonly assumed and occur only for short viewing distances owing to the *depth of field* of the human eye. The depth of field refers to the range of distances in object space within which an image appears in sharp focus, and is specified in meters (m). Depth of field is calculated from the depth of focus using the formula $DOF = 1/F$, where DOF is the depth of field in meters, and F is the depth of focus in diopters (D). The depth of focus refers to the range of distances in image space within which an image appears in sharp focus.

The depth of field refers to the interval in depth over which a stimulus remains in focus and the accommodative response would not be differentially stimulated in a direct way. The depth of field varies according to fixation distance: the eye can tolerate much larger intervals of depth when those intervals are viewed from a far distance than when they are viewed from a near distance before an image goes out of focus. Converging or diverging away from the display surface may pull accommodation to that position in depth, but if that position is within the depth of field, then the *images of the stimulus on the display surface will still be in focus and the accommodative response would not be driven back to the display*. A conflict between accommodative and vergence responses should not occur if the images on the display surface remain within the observer's depth of field. In this case, the accommodative response would be free to follow vergence without any conflict in responding.

It has been suggested previously (Ogle & Schwartz, 1959) that the total depth of focus is on the order of 0.66 D for a 1-arcmin acuity target. However, a more recent estimate of the total depth of focus comes from a comprehensive review of the literature by Wang and Ciuffreda (2006). Wang and Ciuffreda concluded that the average total depth of focus is on the order of 1.0 D (or, equivalently, 0.5 D in front of fixation and 0.5 D behind fixation).

Let us now calculate the depth of field (recall in meters, m) from the depth of focus (recall in diopters, D) in more specific terms. First, for a given viewing distance under consideration, the viewing distance in meters ($D_m$) should be recast in diopters ($D_D$): $D_D = 1/D_m$. Next, calculate the estimated closest point of the depth of field (in meters) from the observer: $Closest_{DOF} = 1/(D_D + 0.5\ D)$. Now, the estimated distance of the depth of field (in meters) *in front of fixation* would be calculated as: $DOF_{front} = |Closest_{DOF} - D_m|$. The estimated farthest point of the depth of field (in meters) from the observer would be: $Farthest_{DOF} = 1/(D_D - 0.5\ D)$. And the estimated distance of the depth of field (in meters) *behind fixation* would be calculated as: $DOF_{behind} = |Farthest_{DOF} - D_m|$.

Based on Wang and Ciuffreda's estimate of the total depth of focus of 1.0 D (0.5 D in front of fixation, 0.5 D behind fixation), and the depth of field calculations shown above (see also Patterson, 2009), we have the following: for a fixation distance of 0.5 m, the total depth of field would range from a distance of about 0.1 m in front of fixation to about 0.17 m behind fixation. For a fixation distance of 1 m, the total depth of field would range from a distance of about 0.33 m in front of fixation to about 1.0 m behind fixation. For a fixation distance of 2 m, the total depth of field would range from a distance of about 1 m in front of fixation to an infinite distance behind fixation.

Patterson (2013) added two other fixation distances to these estimates: For a fixation distance of 3 m, which is close to the recommended viewing distance for TVs (Shibata et al., 2011), the total depth of field would range from 1.8 m in front of fixation to an infinite distance behind fixation. And for a fixation distance of 20 m, which is the recommended distance for viewing 3D cinema (Shibata et al., 2011), the total depth of field would range from about 18 m in front of fixation to an infinite distance behind fixation. Thus, for 3D cinema, almost the entire viewing distance—from a couple meters in front of the user to an infinite distance away—represents the usable depth interval for which accommodation-vergence conflict should not occur. These values are similar to estimates of Percival's zone of comfort as shown in Figure 2 of Hoffman et al. (2008) and in Figure 4 of Shibata et al. (2011).

The depth of focus (and by implication the depth of field) is affected by several factors, which include the luminance of the displayed imagery, which in turn affects pupil size, and the level of resolution. Increases in luminance level produce a smaller pupil; pupil diameter decreases linearly with logarithmic increases in luminance (Loewenfeld, 1993; Reeves, 1920). Thus, for a luminance level of 0.03 cd/m$^2$, pupil diameter will be slightly over 6 mm, while for a luminance level of 300 cd/m$^2$, pupil diameter will be close to 2 mm. For displays whose luminance level is 30 cd/m$^2$ or higher, pupil diameter should be in the range of 2–3 mm. A smaller pupil diameter leads to a larger depth of focus and therefore a larger depth of field. For each millimeter of decrease in pupil size, depth of focus increases by about 0.12 diopters (Ogle & Schwartz, 1959). Thus, display brightness may be an important human-factors comfort issue with smaller, closer stereo displays, such as those used in some 3D gaming systems and desktop 3D computer displays used for molecular modeling. With regard to level of resolution, Ogle and Schwartz (1959) showed that the total depth of focus increased by approximately 0.35 diopters per 0.25 arcmin increase in angular target size (i.e., decreased resolution).

As noted above, the total depth of field can be very large—from 18 m in front of fixation to infinity behind fixation—when viewing 3D cinema at a 20-m viewing distance, for example. This calls into serious question the idea that the problem with 3D movies is due to accommodation-vergence conflict (Engber, 2009). More generally, this suggests that accommodation plays only a minor role in stereo depth perception except for close viewing distances where the depth of field can be quite small. Indeed, a recent study by Hoffman et al. (2008) showed that the presence of accommodation-vergence conflict can impair visual performance and cause fatigue for short viewing distances. The viewing (focal) distances employed in that study were 31.1, 39.4, and 53.6 cm, each one of which would fall outside the depth of field of the other distances. Thus it is not surprising that an accommodation-vergence conflict impaired performance and caused fatigue in that investigation. The viewing/focal distances used in the Hoffman et al. (2008) study were relatively short compared to the distances recommended for viewing TVs or 3D cinema.

*Generally, the remedy for accommodation-vergence conflict is to present the stereo depth of the virtual object(s) such that when the latter is fixated its images on the display surface remain within the depth of field of the human eye.* To do so, one can estimate the stereo depth of the virtual objects by using the calculations given below in the section on "Distance Scaling of Disparity Information". After the stereo depth is calculated, one can then estimate the depth of field of the eye when the user fixates the virtual object and determine whether its images on the display would likely fall within the depth of field. This is the most important recommendation I can give for minimizing the potential for accommodation-vergence conflict.

It is important to note that even if the virtual object appears outside the depth of field, the depth of field will still play a role in reducing the potential accommodation-vergence conflict, as shown in Fig. 4.2. Each panel of Fig. 4.2 depicts a top down view of two eyes (L.E. = left eye; R.E. = right eye) viewing a visual display that is presenting a pair of stereo images, the left-eye image ($S_{L.E.}$) and the right-eye image ($S_{R.E.}$), of a stereoscopic virtual stimulus 'S'.

In panel A of Fig. 4.2, which is a reproduction of Fig. 4.1, the two eyes are converging upon the virtual stimulus S. Also shown is the *accommodation-vergence stimulus conflict* (called 'A-V stimulus conflict'), which is defined as the z-axis distance between the fixated virtual stimulus (to which vergence is directed) and the visual display on which the images are presented and presumably to which accommodation is driven. The z-axis distance between the images on the display (stimuli for accommodation) and the virtual stimulus (stimulus for vergence) represents the accommodation-vergence conflict in *stimulus* terms. As discussed above (and it is worth repeating again here), this accommodation-vergence stimulus conflict has been conflated with an accommodation-vergence conflict in responding throughout the literature.

Now consider the depth of field shown in panel B of Fig. 4.2. The depth of field is depicted as surrounding the fixated virtual stimulus (the actual size of the depth of field would depend upon a number of factors, as discussed above). In panel C of Fig. 4.2, the *effective* accommodation-vergence conflict ('Effective A-V Conflict') is depicted, which is taken to be the z-axis distance between the far boundary of the depth of field and the visual display. This is the approximate distance that would need to be traversed in order for the images on the display to be placed within the depth of field of the eye.

Due to the existence of the depth of field, the user can simply shift convergence to a position slightly behind the perceived position of the virtual stimulus until the images on the display become positioned within the depth of field and blur is eliminated. This is shown in panel D of Fig. 4.2, where the eyes slightly diverge to a fixation position F which makes the virtual stimulus S now carry a small crossed disparity. The effect of the slight divergence to F is to place the far boundary of the

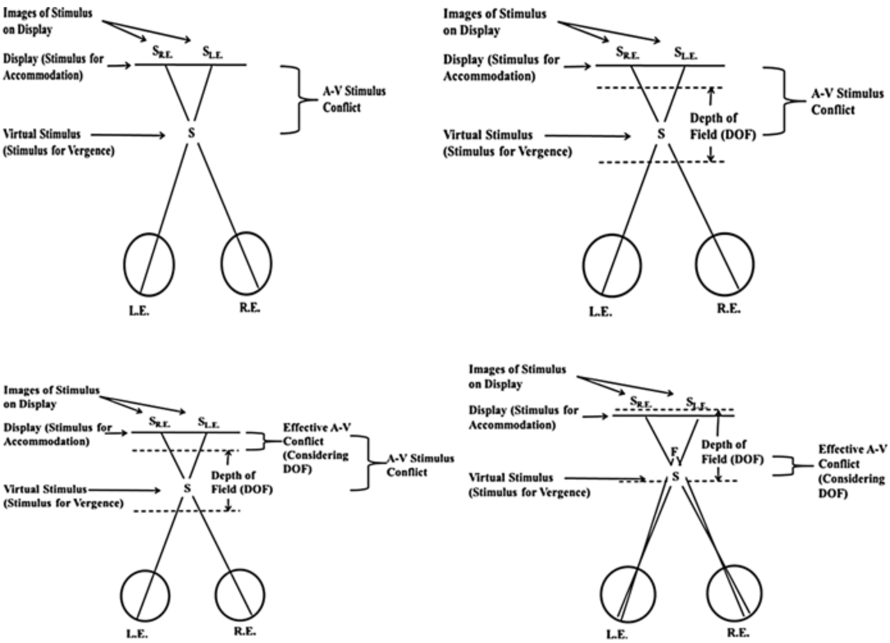


**Fig. 4.2** Diagram depicting different aspects of accommodation-vergence conflict. Each *panel* in the figure shows a top down view of two eyes (L.E. = left eye; R.E. = right eye) viewing a visual display. In the *upper left panel* (same as Fig. 4.1), the two eyes are converging upon a virtual stimulus 'S' presented with stereoscopic depth. The displayed left-eye image of the virtual stimulus is $S_{L.E.}$ and the right-eye image is $S_{R.E.}$ Because the eyes are converged at S, the horopter will go through S and S will project a zero disparity to the visual system; thus the display itself will present an uncrossed disparity to the visual system. Also shown is the accommodation-vergence stimulus conflict ('A-V stimulus conflict'), which is defined as the z-axis distance between the fixated virtual stimulus (to which vergence is directed) and the visual display on which the images are presented (and presumably where accommodation is driven). In the *upper right panel*, the hypothetical depth of field surrounding the fixated virtual stimulus is shown. In the *lower left panel*, the effective accommodation-vergence conflict ('Effective A-V Conflict') is depicted, which in this case is taken to be the z-axis distance between the far boundary of the depth of field and the visual display. In the *lower right panel*, the eyes slightly diverge to a fixation position F which makes the virtual stimulus S now carry a small crossed disparity. The effect of the slight divergence to F is to place the far boundary of the depth of field on the visual display which now enables the images of the virtual stimulus ($S_{R.E.}$ and $S_{L.E.}$) to be in focus. The z-axis distance between the perceived position of the virtual stimulus S and the fixation position F can also be taken to be the effective A-V conflict—that is, the amount of vergence change needed to make the images on the visual display be in focus. The effective A-V conflict will typically be much smaller than the A-V stimulus conflict because the latter does not take into account the depth of field

depth of field on the visual display which now enables the images of the virtual stimulus ($S_{R.E.}$ and $S_{L.E.}$) to be in focus. (If the virtual stimulus appears in depth behind the display, the user may need to shift convergence to be slightly in front of the virtual stimulus to place the displayed images within the near boundary of the depth of field.) The magnitude of the effective accommodation-vergence conflict can also be taken to be the z-axis distance between the virtual stimulus S and the shift in convergence (to F) necessary to position the displayed images within the user's depth of field.

Of course, the user may still have to re-converge to the virtual stimulus to look at it which makes the images on the display become slightly blurred; and the blurred images will again drive accommodation to refocus the images. But the images should become refocused once they fall within the depth of field. In short, the effective conflict will be between the boundary of the depth of field that is closest to the display, which is linked to vergence, and the images on the display. The *effective* accommodation-vergence conflict will typically be much less than the accommodation-vergence *stimulus* conflict because the latter does not take into account the depth of field.

These considerations suggest that the actual (effective) accommodation-vergence conflict present in the Hoffman et al. (2008) study, mentioned above, and in the Shibata et al. (2011) investigation, discussed below, was much less than what was assumed. More generally, the common idea that vergence-accommodation conflict is present in all traditional stereo displays is an over-generalization because an accommodation-vergence conflict in terms of motor responding would depend upon the magnitude of the disparity of the virtual stimulus (which determines its depth position) together with the observer's viewing distance to the display (which determines the depth of field). An accommodation-vergence conflict in motor responding should not be present when an observer convergences to a depth plane that places the displayed images within the depth of field, which would be true under many circumstances.

In the Shibata et al. study, the viewing/focal distances were 27 cm, 40 cm, 77 cm, and 10 m, each one of which would likely fall outside the depth of field of the other distances (their Experiment 1); or pairs of distances: 40 cm versus 30 cm, 40 cm versus 59 cm, 77 cm versus 48 cm, 77 cm versus 2.0 m, 10 m versus 1.11 m, and 10 m versus > infinity, each member of each pair of which would likely fall outside the depth of field of the other member of the pair (their Experiment 2). These authors found no evidence of impairment in performance on their visual discrimination task under the accommodation-vergence conflict conditions of Experiments 1 and 2 relative to the no-conflict conditions. Moreover, the results of Shibata et al.'s Experiment 1 showed that reported eye tiredness, blurriness of vision, and eye strain were only mildly present. This is likely to be due to the fact that the *effective* accommodation-vergence (motor) conflict would have been much less than the stimulus conflict due to the presence of the observers' depth of field, as discussed above.

It is also important to mention that, in the Shibata et al. study, the space-average luminance of the display was 0.13 $cd/m^2$, a relatively low value. This low luminance value has important implications for accommodation because it is well-known that displays with low values of luminance will cause accommodation to drift toward a resting distance whose average is about 66 cm in front of the observer, rather than to a distance corresponding to the 'fixation' stimulus. This phenomenon is called *dark focus*. There is also an analogous phenomenon called 'Dark Vergence'. Dark focus refers to the tendency of accommodation to drift toward a resting distance of approximately 1.52 diopters, or 66 cm, in front of the observer under degraded stimulus conditions or in darkness, although there are large individual differences (Johnson, 1976; Leibowitz & Owens, 1975a, 1975b, 1978; Owens, 1979).

For example, Leibowitz and Owens (1975b) found that accommodative responding was very close to the dark focus value when a building in an outdoor scene was viewed under very reduced luminance conditions, which would equal a luminance value on the order of about 3 cd/m$^2$. And in a study by Andre and Owens (1999), accommodative responding was close to the dark focus value when the display luminance was set at 0.61 cd/m$^2$. In short, under degraded stimulus conditions, such as low luminance values (i.e., below about 7 cd/m$^2$ according to Johnson, 1976), observers' accommodative response will not correspond to the focal distance of the stimulus, but instead accommodation will be pulled toward the dark focus position (unless of course the stimulus is already positioned at each person's dark focus value).

This means that, in the Shibata et al. (2011) study, given that the space-average luminance of the display was 0.13 cd/m$^2$, the average accommodative response to the nominally closer focal distances of 2.5 diopters (40 cm) and 3.7 diopters (27 cm) would probably have been farther than assumed and tending toward the average dark focus position (and thus enlarging the depth of field slightly for those distances), while the average accommodative response to the nominally farther focal distances of 0.1 diopters (10 m) and 1.3 diopters (77 cm) would probably have been closer than assumed (and thus shortening the depth of field slightly for those distances).

Moreover, in the studies by Banks and colleagues (Hoffman et al., 2008; Shibata et al., 2011), the stimuli to accommodation and vergence were manipulated but the accommodative and vergence *responses* themselves were not measured. Thus, we cannot be certain where the observers were focusing when the stereo displays were viewed in those studies—it was assumed that the observers were focusing on the display surface while verging to a different depth plane, which induced the supposed conflict.

However, studies by Hasegawa et al. (2011), Miyao et al. (2012), and Shiomi et al. (2013) measured both accommodation and vergence when their observers fixated a stimulus that oscillated between closer and farther depth positions on a stereo display (the depth positions were apparently within the observers' depth of field). These authors reported that accommodation and vergence together closely followed the depth of the virtual images and no conflict in oculomotor responding was present (accommodation following vergence was likely due to 'vergence-accommodation', the indirect effects on accommodation of changes in vergence). These results showed that the presence of an accommodation-vergence *stimulus* conflict does not necessarily produce a conflict between accommodative and vergence *responses*. And it is the latter that has been assumed to be at the heart of viewer discomfort when traditional stereo displays are viewed (Hoffman et al., 2008; Shibata et al., 2011; Velger, 1998; Wann et al., 1995).

In summary, it is concluded that: (1) there should little or no conflict between accommodation and vergence *responses* when a user converges to fixate a virtual object whose depth position allows its displayed images to be placed within the users' depth of field (below I show how to calculate the perceived depth of a stereo stimulus); (2) the depth of field is a reasonable yet simple measure of the zone of comfort within which virtual objects can be fixated without visual discomfort from accommodation-vergence (response) conflict; (3) when users converge to depth intervals that place the displayed images outside the depth of field, the reported discomfort is relatively mild and visual performance is unimpaired (Shibata et al., 2011), except for short viewing distances for which the depth of field is small (e.g., viewing distances of 31–54 cm; Hoffman et al., 2008); (4) given the long viewing distances involved in 3D cinema (e.g., 20 m), the depth of field is so large—from a couple meters in front of the viewer to an infinite distance behind the screen—that accommodation-vergence conflict cannot be the reason for visual discomfort when 3D movies are viewed; (5) when accommodation and vergence responses are measured while observers view a stereoscopic display depicting a virtual stimulus moving in z-axis depth (within viewers' depth of field), both accommodation and vergence follow the virtual stimulus and no conflict is evident (Hasegawa et al., 2011; Miyao et al., 2012; Shiomi et al., 2013). This suggests that accommodation-vergence conflict may have been overstated as a reason for visual discomfort when conventional stereo displays are viewed.

It should also be mentioned that accommodation and vergence can be 'uncoupled' when free viewing a pair of left-eye and right-eye images of a stereo picture, positioned side-by-side. To do so, an observer can converge the eyes and fixate in front of the pair of images (fixating the tip of a finger held up halfway between the eyes and the images helps) so that one image goes to one eye and the other image goes to the other eye. The pair of images will perceptually fuse and the virtual stereo stimulus will be seen floating in depth in front of the images (or the observer can try to diverge the eyes behind the pair of images to get an equivalent effect, which is more difficult to do). It is this author's opinion that this kind of free viewing of stereo pictures can be successful provided that the point of convergence places the pair of images within the observer's depth of field.

It should also be mentioned that presbyopia, a progressively reduced ability to focus on near objects, develops with aging. Its symptoms first appear usually between the ages of 40 and 50 years. This suggests that, for older stereo display users, the declining range of ocular focus may prevent the blur signal from entering the feedback response loop (i.e., if an older observer cannot adjust focus well, then perhaps the visual system will not send a signal to do so). This, in turn, may lessen the tendency for accommodation-vergence conflict in older stereo display users.

## 4.3 Percival's Zone of Comfort

Percival's Zone of Comfort (Percival, 1892) refers to the range of vergence and accommodation responses for which the user can fuse images without discomfort in binocular viewing (see Fig. 4.3). Banks and colleagues (e.g., Hoffman et al., 2008; Shibata et al., 2011) have suggested that this concept should be considered when

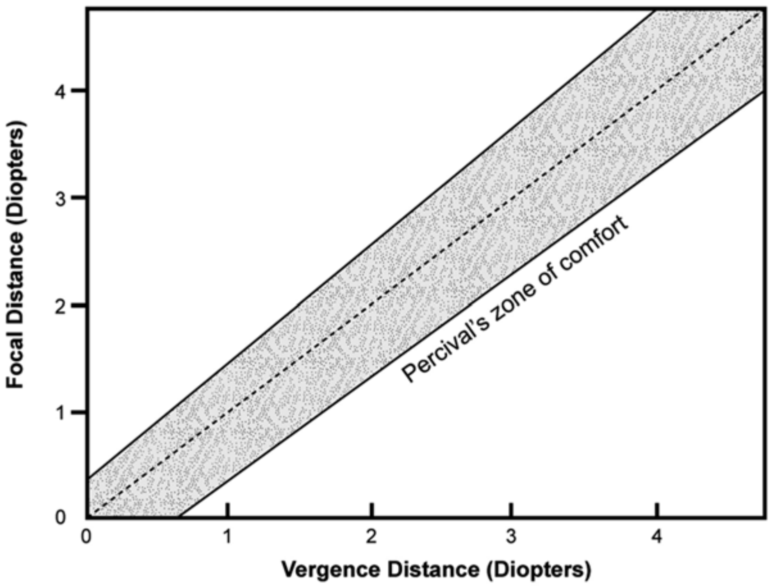


**Fig. 4.3** Percival's Zone of Comfort. The *grey-colored* zone in the figure is Percival's Zone of Comfort, which is the range of vergence and accommodation responses within which the user can fuse images without discomfort in binocular viewing. Reproduced from Figure 4 of Shibata, T., Kim, J., Hoffman, D.M. & Banks, M.S. (2011). The zone of comfort: Predicting visual discomfort with stereo displays. Journal of Vision, 11, 1–29. Copyright Association for Research in Vision and Ophthalmology. By permission

discussing the issue of accommodation-vergence conflict in stereo display viewing. However, an alternative concept that could be considered is the depth of field (and by implication, the depth of focus) of the eye: the observer should be able to make vergence eye movements such that the display remains within the observers' depth of field, which should prevent discomfort because the images on the display remain in focus.

This concludes Chap. 4. The next chapter covers other low-level factors that affect stereo viewing.

# Chapter 5
# Low-Level Factors, Continued

This chapter covers three low-level factors that affect stereo viewing: interocular differences in luminance, interocular differences in contrast, and stereoanomaly. These factors are considered 'low level' because they occur at peripheral stages of visual processing: interocular differences in luminance and contrast entail differences in image characteristics, and stereoanomaly is a condition that likely involves a decrease in sensitivity to stereo depth information at a disparity-detection stage of processing.

## 5.1 Interocular Differences in Luminance

Human stereo depth perception is robust despite significant differences in luminance level between the eyes (called interocular differences) as long as the images are static. (See Chap. 3, Sect. 3.2, for previous discussion of the Pulfrich effect, a spurious depth effect, that occurs with interocular differences in luminance of over 10 % and moving images.) Boydstun, Rogers, Tripp, and Patterson (2009) reported that stereoscopic depth perception (indexed by perceived depth measurements and depth discrimination thresholds) was relatively unaffected by interocular luminance differences of up to 60 %. Visual discomfort was studied by Kooi and Toet (2004) and they found that only slight discomfort was reported with interocular luminance differences of up to 25 %; above 25 % discomfort was greater. The recommendation here is that interocular differences in luminance level should be less than 25 %, ideally less than 5 % (the latter of which would also be good when you have moving images).



R.E. Patterson, *Human Factors of Stereoscopic 3D Displays*,
DOI 10.1007/978-1-4471-6651-1_5

## 5.2 Interocular Differences in Contrast

Human stereo depth perception is robust despite significant interocular differences in stimulus contrast. It has been reported (Hess, Liu, & Wang, 2003) that stereo depth perception, as indexed by depth discrimination performance, was largely unaffected by interocular contrast differences of up to 83 %. It has been found (Kooi & Toet, 2004) that the threshold for discomfort with interocular contrast differences was between 25 and 50 %. The recommendation here is that interocular differences in contrast should be kept to less than 25 %, ideally less than 5 %.

## 5.3 Stereoanomaly

As discussed in Chap. 2, objects located in a depth plane in front of fixation (and the horopter) will project images with “crossed disparity”, whereas objects located in a depth plane behind fixation (horopter) will project images with “uncrossed disparity”. There exists in the visual system different classes of cortical neurons that signal crossed versus uncrossed disparity, called ‘near depth’ cells (or channel) and ‘far depth’ cells (or channel), respectively, as well as neurons that are either excited by, or inhibited by, zero disparity (Cumming & DeAngelis, 2001; Poggio, 1995; Poggio, Motter, Squatrito, & Trotter, 1985). These classes of neuron form the neural basis of human stereo depth perception (see Fig. 5.1). LeVay and Voight (1988) and Freeman and Ohzawa (1990) have suggested that the response types of disparity-tuned cells may belong to a continuum, not to discrete categories. This concept would not change our analysis below; if a continuum exists, then the tuning curves shown in Fig. 5.1 would represent pooled responses from a number of crossed-disparity cells, uncrossed-disparity cells, and so on.

For some individuals, depth perception in one disparity direction, crossed or uncrossed, is unreliable even though depth perception in the other direction is reliable. Such individuals experience reversed depth perception in the unreliable direction; crossed disparity may be perceived as back depth, or uncrossed disparity may be perceived as front depth. This condition is referred to as ‘stereoanomaly’ (Richards, 1970, 1971) and it can occur in about 20–30 % of individuals under degraded stimulus conditions such as brief stimulus exposure (Patterson & Fox, 1984). The stereoanomalous individuals should be distinguished from a different group of individuals who possess a medical condition called strabismus and who are stereoblind, which represents approximately 6–8 % of the population.

One possible cause of stereoanomaly is that one of the classes of cortical neuron that encode crossed or uncrossed disparity (Cumming & DeAngelis, 2001; Poggio, 1995; Poggio et al., 1985) lacks normal sensitivity to disparity information. Because perceiving stereo depth is thought to involve a pooling of responses from many cortical neurons, it is possible that the neurons encoding the other disparity sign which possess normal sensitivity dominate in the neural response, producing the perception of reversed depth in these individuals (Patterson & Fox, 1984; Patterson

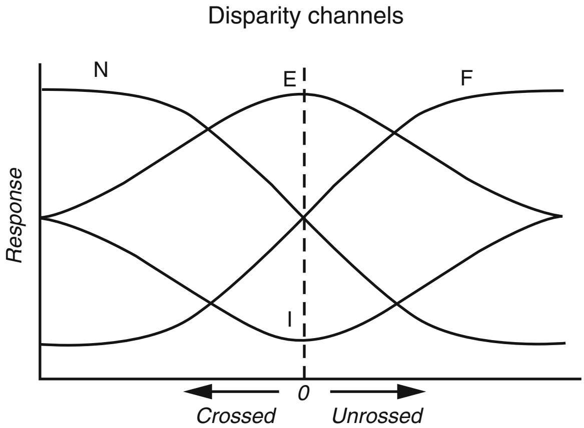


**Fig. 5.1** Sketch of the response (ordinate) of hypothetical binocular disparity-tuned neurons to crossed, uncrossed, or zero disparity (abscissa). 'N' indicates a near-depth channel (i.e., pooled response from neurons tuned to crossed disparity) that is excited by crossed disparity and inhibited by uncrossed disparity; 'F' depicts a far-depth channel (i.e., pooled response from neurons tuned to uncrossed disparity) that is excited by uncrossed disparity and inhibited by crossed disparity; 'E' indicates a zero-depth channel (i.e., pooled response from neurons) that is excited by zero disparity; 'I' depicts a zero-depth channel (i.e., pooled response from neurons) that is inhibited by zero disparity. Depth perception is believed to be derived from the pooled responses from all such channels. Reproduced from Figure 4 of Patterson (2007), Human factors of 3D displays, Journal of the Society for Information Display, 15, 861–871. Copyright Society for Information Display. By permission

et al., 1995). Individuals with strabismus and who are stereoblind presumably lack, or are deficient in, this neural substrate in their visual system.

Stereoanomaly and stereoblindness may limit the number of individuals who have the capability to use stereoscopic displays in certain applications. It may be important to screen for stereoanomaly and stereoblindness (van Ee & Richards, 2002). With respect to stereoanomaly, the likely remedy is to present disparity information under non-degraded conditions or bolster the disparity information with other depth or distance cues (van den Enden & Spekreijse, 1989). However, the precise way in which degraded viewing conditions (e.g., brief stimulus exposures) can contribute to the number of individuals reporting stereoanomaly has not been systematically studied. With regard to stereoblindness, there appears to be no known remedy for inducing normal stereopsis in these individuals.

## 5.4 Summary of Low-Level Factors

For clear and comfortable stereo viewing, crosstalk should be less than 2 % and interocular differences in luminance level and contrast should be less than 5 % each. To minimize accommodation–vergence conflict, the stereo depth of each fixated virtual object should enable its displayed images to be placed within the depth of field of the human eye. Finally, it may be important to screen for stereoanomaly and stereoblindness in certain applications (e.g., when the stereo information will be presented under degraded conditions or strong monocular cues are lacking).

Now that we have covered the low-level human factors, we now turn to a discussion of the contextual factors that affect the viewing of stereo displays.

# Chapter 6
# Contextual Factors

This chapter covers two contextual factors that affect stereo viewing: spatio-temporal frequency, and distance scaling of disparity. These factors are considered 'contextual' because they partly define the conditions that exist when someone views a stereo display: spatio-temporal frequency involves the distribution of display luminance, and distance scaling of disparity involves visual information about viewing distance and disparity within a cue integration process.

## 6.1 Spatio-Temporal Frequency Effects

The range of disparity magnitudes (in seconds, minutes, or degrees of arc) that yields reliable (patent) stereopsis depends upon the luminance spatial-frequency and the luminance temporal-frequency content of the displayed imagery or scene (Schor & Wood, 1983). Recall, as briefly discussed in Chap. 2 (see Fig. 2.5), that spatial frequency refers to the rate of modulation of luminance across space (in units of cycles per degree, or cyc/deg, of visual angle on the retina) and temporal frequency refers to the rate of modulation of luminance in time (in units of cycles per second, or Hz).

Schor and Wood showed that the smallest disparity that could be reliably discriminated, which was called stereoacuity or 'Dmin' (for disparity minimum) or lower disparity limit (LDL), was about 20 arcsec in the spatial frequency range of about 2–20 cyc/deg (the highest spatial frequency tested). This meant that the precise discrimination of small differences in stereo depth (which could be very fine) required the presence of fine spatial detail in the images on the stereo display. These authors also showed that the maximum disparity that can be reliably discriminated, called 'Dmax' (for disparity maximum) or upper disparity limit (UDL), was slightly over 40 arcmin within this spatial frequency range (i.e., fine spatial detail). Below a spatial frequency of about 2 cyc/deg (moderately coarse spatial details in an image), both Dmin (LDL) and Dmax (UDL) increased with decreasing spatial frequency in



R.E. Patterson, *Human Factors of Stereoscopic 3D Displays*,
DOI 10.1007/978-1-4471-6651-1_6

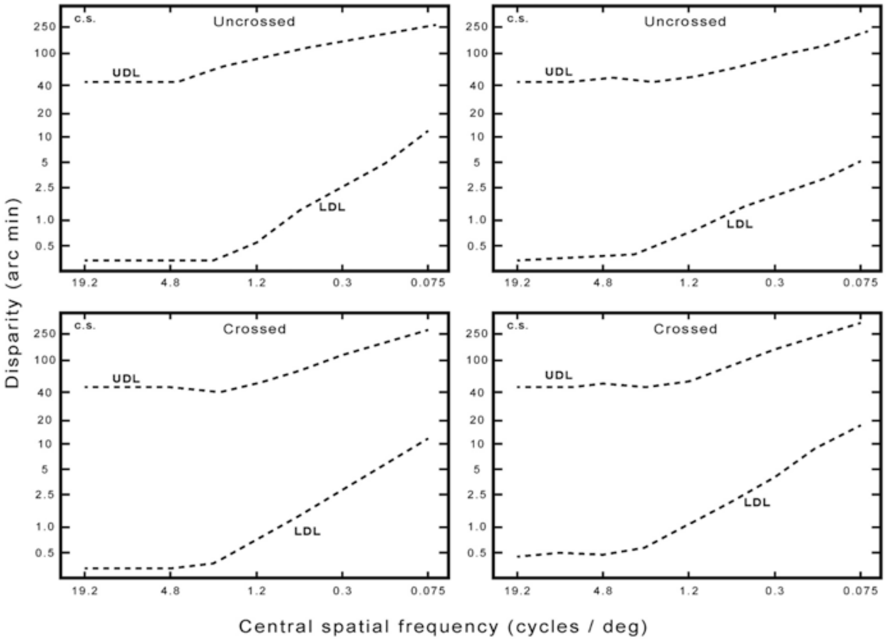


**Fig. 6.1** Binocular disparity thresholds for different central spatial frequencies. The ordinate shows threshold disparity in minutes of arc and the abscissa depicts spatial frequency of luminance modulation in cyc/deg. These data reveal a trend in which high spatial frequency (i.e., fine spatial detail in a stereo display) was associated with an ability to discriminate small disparity values (fine stereoacuity, or the lower disparity limit, LDL) near the horopter as well as moderate disparity values away from the horopter (upper disparity limit, or UDL). Low spatial frequency (i.e., coarse spatial detail in a stereo display) was associated with an ability to discriminate moderate disparity values (LDL) near the horopter and larger disparity values (UDL) away from the horopter. In a sense, the envelope of space, out in the visual field, within which depth can be discriminated—i.e., the range of usable depth perception, from LDL to UDL—shifts to larger disparity values from small disparity values as the spatial detail in an image on a stereo display changes to coarse from fine. Reproduced from Figure 2 of Schor, C.M. & Wood, I. (1983). Disparity range for local stereopsis as a function of luminance spatial frequency. Vision Research, 23, 1649. Copyright Elsevier. By permission of Elsevier

a linear fashion on log–log axes. At a low spatial frequency of about 0.1 cyc/deg, Dmin was about 5 arcmin (poor stereoacuity) and Dmax was about 4 arc degrees. The range of usable depth perception shifted from small disparity values to larger disparity values as the spatial details in the images on a stereo display changed from fine to coarse. These values applied equally well in both the crossed and uncrossed directions from the horopter. See Fig. 6.1.

Since these values apply to both the crossed and uncrossed ranges of disparity, we can add these values together to get the total disparity range of patent (reliable) stereopsis: it is about 80 arcmin with medium to high spatial frequencies (2–20 cyc/deg) represented in the display, and as large as 8 arc degrees with low spatial frequencies (0.1 cyc/deg) in the display.

Schor, Wood, and Ogawa (1984) found that, with low spatial frequencies of about 0.1–2 cyc/deg, sensitivity to disparity information was greater for dynamic (1 Hz) than for static disparities. With high spatial frequencies in the range of about 2–20 cyc/deg, sensitivity to disparity was equal for dynamic and static disparities. Patterson (1990) found that stereoacuity was on the order of 10–15 arcsec, which corresponded to a small Dmin, for a spatial frequency of 8 cyc/deg, and stereoacuity declined (thresholds increased) with decreasing spatial frequency. This trend followed that of Schor and Wood (1983). Patterson (1990) also reported that stereoacuity declined (higher thresholds) with the high spatial-frequency pattern when temporal frequency was increased across the range from zero to 20 Hz. However, stereoacuity improved (lower threshold of about 20 arcsec) with lower spatial-frequency patterns in the range of 0.6–4 cyc/deg when temporal frequency was moderate (1–5 Hz). Fine sensitivity to small disparity differences (i.e., fine stereoacuity) requires high spatial frequencies (fine spatial details in an image) and low temporal frequencies (e.g., static images), or low spatial frequencies (coarse spatial details) and moderate temporal frequencies. See Fig. 6.2.

Several authors (Edwards, Pope, & Schor, 1999) have reported that stereoscopic depth can be processed from briefly-presented (140-ms exposure) targets with very

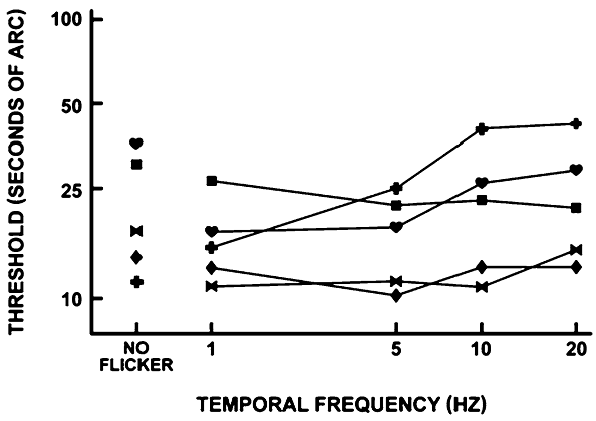


**Fig. 6.2** Stereoacuity thresholds for different temporal frequencies. The ordinate shows threshold disparity magnitude in seconds of arc and the abscissa depicts temporal frequency of luminance modulation. Each curve in the plot represents a different spatial frequency. These data revealed an interaction between temporal frequency and spatial frequency: high spatial frequencies (i.e., fine spatial detail), depicted with the *plus* symbol, were associated with an ability to discriminate small disparity values (fine stereoacuity) provided that temporal frequency was zero ('no flicker'). As temporal frequency increased from zero, stereoacuity was poor with the high spatial frequency. Middle spatial frequencies (i.e., moderately coarse spatial detail), shown with the *diamond* and *bowtie* symbols, were associated with an ability to discriminate small disparity values (fine stereoacuity) provided that temporal frequency was moderate (1–5 Hz). Reproduced from Figure 2 of Patterson, R. (1990). Spatiotemporal properties of stereoacuity. Optometry and Vision Science, 67, 123–128. Copyright Wolters Kluwer Health Lippincott Williams & Wilkins. By permission

large disparities (4–8 arcdeg) whose contours are oriented orthogonal to one another. This type of 'transient stereopsis' likely occurs before binocular rivalry has a chance to become manifest. In a different line of research, it has been found (Wilcox & Hess, 1995) that Dmin (stereoacuity) and Dmax thresholds depend upon the overall size of stimuli.

With naturalistic stimuli such as real-world scenes and events, which would contain broad-band spatial-frequency/temporal-frequency content, it would be expected that a wide range of binocular disparities could be processed by the collective action of the spatial-frequency/temporal-frequency channels of the visual system. More specifically, one can tie together the results of Schor and Wood (1983), Schor et al. (1984), and Patterson (1990) to the properties of the dorsal and ventral pathways, discussed in Chap. 2.

Collectively, the results from these studies suggest that fine stereoacuity, and a small region of depth from the observer within which stereo operates, is mediated by visual channels responsive to fine spatial detail and low rates of temporal modulation (i.e., they are sluggish channels; Patterson, 2007). These channels draw connections mainly from the central retina: fine stereoacuity occurs near the horopter and fixation point (Blakemore, 1970), which translates into stimulation near the fovea. These channels would likely be related to the ventral stream discussed in Chap. 2. See Fig. 6.3. These studies also suggest that a large region of depth from the observer, together with poor stereoacuity, is mediated by visual channels responsive to relatively coarse spatial detail and moderate rates of temporal modulation (brisk channels). These channels draw connections from both the central and peripheral retina. These channels would likely be related to the dorsal-stream also discussed in Chap. 2.

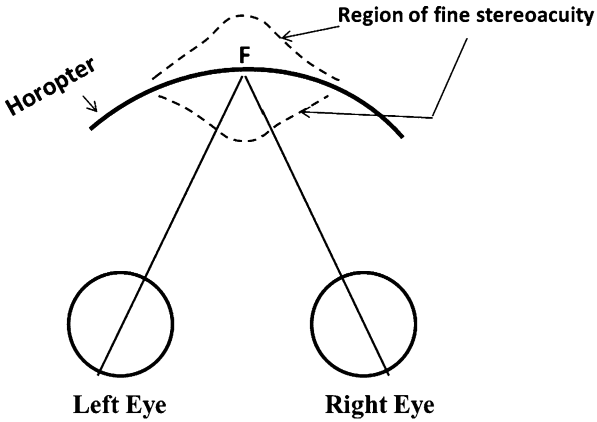


**Fig. 6.3** Top down view of the region of fine stereoacuity. Fine stereoacuity is mediated by visual channels responsive to fine spatial detail and low rates of temporal modulation. These channels draw connections mainly from the central retina, thus fine stereoacuity occurs near the horopter and fixation point. In short, the region of fine stereoacuity would be Gaussian distributed in front of and behind the fixation point and horopter (F = fixation point)

There are a number of issues raised by having two stereo pathways with different spatial/temporal properties. The spatial multiplexing method can involve decreased display spatial resolution which would differentially impact the two cortical streams. Performance on certain stereo tasks that demand high spatial-frequency processing (ventral stream)—such as stereoacuity tasks (i.e., fine depth discrimination)—could be impaired because the spatial resolution of the display could be too low. The temporal multiplexing (field sequential) method entails decreased display temporal resolution which could also differentially affect the two cortical streams. Performance on certain stereo tasks that demand high temporal acuity (dorsal stream)—such as heading control—may be impaired if the frame alternation rate was too low. Moreover, situations in which the field of view is restricted (e.g., head-worn displays) would emphasize ventral stream functioning over dorsal stream functioning and thus these situations could also impair certain stereo tasks that demand high temporal acuity and a large field of view (i.e., tasks dominated by dorsal stream functioning).

With the time-multiplexing (field sequential) technique and a large field of view, peripheral areas of the retinae, which respond to moderate and high rates of temporal modulation, will be strongly stimulated. Disruptive peripheral flicker may be perceived when viewing large field-of-view immersive displays that induce stereo with the time-multiplexing method (Patterson, 2007). The remedy for this problem is to employ a high frame rate so that the visual system temporally integrates the intermittent information seen in the periphery. The actual frame rate needed would likely depend upon the visual size or extent of the display and other display characteristics (e.g., image contrast).

Finally, with stereo displays that depict real-world scenes in perspective view, the bottom of the display may contain relatively coarse details (objects in the scene appearing closer to the observer) while the top of the display may contain fine details (objects appearing farther away). Thus, there could be a spatial-frequency gradient across the vertical extent of the display, going from lower spatial frequencies (bottom) to higher spatial frequencies (top). In this case it is likely that larger disparities could be processed at the bottom of the display and smaller disparities at the top of the display, with finer depth discrimination occurring near the top of the display.

But, as stated earlier, with depicted real-world scenes and events, there is the opportunity for the imagery to contain sufficient broad-band spatial-frequency/temporal-frequency content such that a wide range of binocular disparities could be processed by the collective action of the ventral cortical stream and the dorsal cortical stream pathways.

## 6.2 Distance Scaling of Disparity

The amount of depth that is perceived when viewing a stereo display will depend upon the observer's egocentric viewing distance. Egocentric viewing distance is defined as the distance between the observer and the point of fixation (display). For a fixed amount of lateral separation between corresponding left-eye and right-eye images presented on a stereo display, changes in the observer's egocentric viewing distance will alter both the magnitude of binocular disparity that is projected to the two eyes, and the amount of depth that the observer perceives. Increases in viewing distance will lessen the magnitude of the disparity projected to the two eyes yet perceived depth will increase. Decreases in viewing distance will increase the magnitude of disparity yet perceived depth will decrease. There is no one-to-one correspondence between a given magnitude of disparity and the amount of depth perceived by an observer; egocentric viewing distance will moderate the disparity-depth relationship. In order for stereo depth to be perceived, the visual system re-calibrates the disparity information for different egocentric viewing distances. This re-calibration is called 'distance scaling of disparity', or 'disparity scaling' (Ono & Comerford, 1977; Patterson, 2007, 2009; Patterson & Martin, 1992; Patterson, Winterbottom, & Pierce, 2006).

Disparity scaling is a concept that makes one distinguish between egocentric viewing distance and relative depth. Whereas egocentric viewing distance refers to the distance between an observer and the point of fixation, relative depth refers to the depth interval between an object and some reference point such as the horopter. Stereopsis provides information about relative depth and not egocentric distance. (The same would be true for another immersive relative depth cue called motion parallax.) However, egocentric distance information is needed in order for the perception of relative depth from stereopsis to be anchored and stable in a scene. The visual system performs a cue-integration operation and combines binocular disparity information with cues to egocentric viewing distance in order to visually compute perceived relative depth. If the egocentric viewing distance cues are unreliable or misperceived, then perceived depth becomes unreliable or misperceived.

An important distinction is how disparity changes with egocentric viewing distance in the real world versus how disparity changes with viewing distance when a stereo display is viewed. This distinction touches upon the concept of 'depth constancy' as explained below.

### *6.2.1 Disparity Change with Egocentric Viewing Distance in Real World Viewing*

Under viewing conditions in the real-world, the magnitude of binocular disparity varies approximately inversely with the square of the egocentric viewing distance (Fig. 6.4). If viewing distance to a pair of objects, separated in depth, is halved, then the disparity derived from that depth would be approximately four times its initial value. If viewing distance to the objects is instead doubled, the disparity would be approximately one-fourth its original value. With real-world viewing, and with a relatively large viewing distance and symmetrical convergence, disparity magnitude is computed as: $\mathrm{r}\left(\mathrm{in\,radians}\right)=\left(\mathrm{I}\times\mathrm{d}\right)/\mathrm{D}_{\mathrm{m}}^{2}$, where r is disparity, I is interpupillary distance (in meters), d is the depth interval (in meters), and $D_m$ is viewing distance in meters (Cormack & Fox, 1985a). In real-world viewing, there is no one-to-one

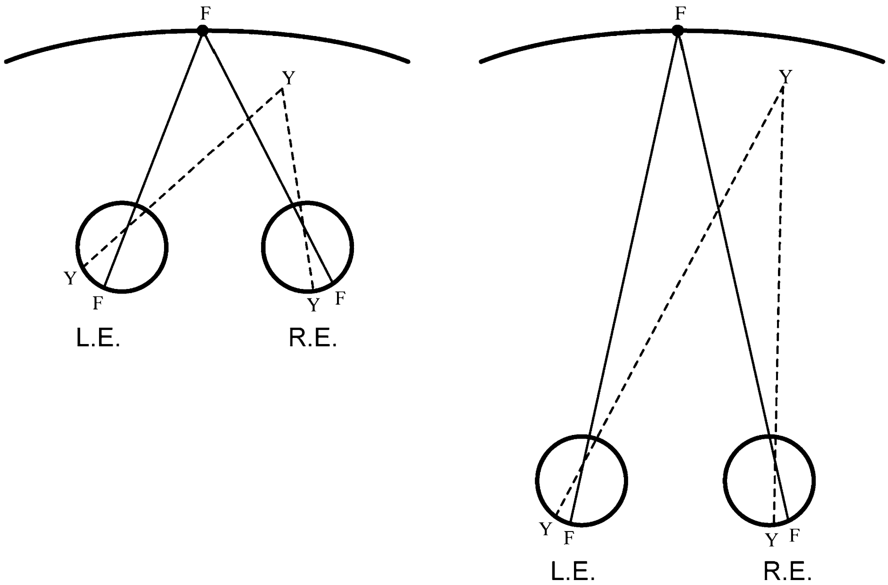


**Fig. 6.4** Drawing depicting how changes in egocentric viewing distance affect the magnitude of binocular disparity. Diagram on the left side of the figure shows a top down view of two eyes viewing a fixation point F at a close distance, whereas the diagram on the right side depicts two eyes viewing the same fixation point F at a farther distance. In both diagrams, the depth interval between F and Y is the same magnitude. Increasing the viewing distance (shown on the *right* diagram) causes disparity magnitude to decrease from its original value shown on the *left* diagram. In order for the observer to perceive stable and reliable relative depth, information about binocular disparity is likely visually combined with cues to egocentric viewing distance

correspondence between disparity magnitude and the amount of perceived depth because egocentric viewing distance is a moderating variable.

In real-world viewing, changes in the observer's egocentric viewing distance will alter disparity magnitude, but—because the visual system recalibrates the disparity for different viewing distances—the amount of depth between stationary objects appears to remain constant. The perception of stable and constant depth despite changes in egocentric viewing distance is called 'depth constancy'. In the real world, the perception of stereo depth typically remains constant with changes in egocentric viewing distance, thus stereo depth constancy appears to exist under naturalistic viewing.

### *6.2.2 Disparity Change with Egocentric Viewing Distance in Stereo Display Viewing*

When viewing stereo displays, the magnitude of binocular disparity varies approximately inversely with the first power of egocentric viewing distance. If viewing distance to a stereoscopic display depicting a given depth interval is halved, then disparity will be approximately twice its initial value, and if viewing distance is doubled, disparity will be approximately one-half its original value. When viewing stereo displays with symmetrical convergence and targets located near the midsagittal plane, disparity magnitude is computed as: r (radians) = $S/D_m$, where r is disparity, S is the separation of the half-images on the stereoscopic display in meters, and $D_m$ is egocentric viewing distance in meters (Cormack & Fox, 1985b).

In stereo display viewing, changes in the observer's egocentric viewing distance will alter both the magnitude of binocular disparity and the amount of perceived depth. For a fixed amount of lateral separation between corresponding left-eye and right-eye images on a stereo display, an observer will typically perceive depth as being proportional to viewing distance. If viewing distance is doubled, perceived depth will double, and if viewing distance is halved, perceived depth will be halved. Perceived depth of a virtual object in a stereo display is computed as: $d = (D_m \times S)/(I \pm S)$, where d is predicted depth (in meters), $D_m$ is viewing distance in meters, S is separation between half-images on the display screen (in meters), and I is interpupillary distance (in meters); when disparity is crossed, the denominator is $(I + S)$, and when disparity is uncrossed, the denominator is $(I - S)$ (Cormack & Fox, 1985b). Perceived depth in a stereoscopic display does vary in accord with this relation (Boydstun, Rogers, Tripp, & Patterson, 2009; Patterson et al., 1995; Patterson, Moe, & Hewitt, 1992; Richards, 2009; Ritter, 1977; Wallach & Zuckerman, 1963). Here, the visual system may recalibrate the disparity for different egocentric viewing distances but the recalibration derived from real-world viewing becomes misapplied when a stereo display is viewed. Perceived depth in a stereo display is not constant with changes in viewing distance. Stereo depth constancy does not appear to exist when stereo displays are viewed.

### *6.2.3 Remedy for Potential Accommodation–Vergence Conflict*

Recall from Chap. 4, one remedy for a potential accommodation–vergence conflict is to present the stereo depth of the fixated virtual object so that its displayed images fall within the depth of field of the human eye. Now, an estimate of the perceived depth of a virtual object in a stereo display is computed as $d = (D_m \times S)/(I \pm S)$, where d is predicted depth (in meters), $D_m$ is viewing distance in meters, S is separation between half-images on the display screen (in meters), and I is interpupillary distance (in meters), given above. And an estimate of the total depth of field can be calculated by using the expressions provided in Chap. 4. Recall that, first, the viewing distance in meters ($D_m$) should be recast in diopters ($D_D$): $D_D = 1/D_m$. Next, the estimated closest point of the depth of field (in meters) from the observer is calculated as: $Closest_{DOF} = 1/(D_D + 0.5D)$. And the estimated farthest point of the depth of field (in meters) from the observer would be: $Farthest_{DOF} = 1/(D_D - 0.5D)$. Combining the calculations for perceived depth with calculations for the depth of field will help determine whether the stereo depth of the fixated virtual object will allow its displayed images to fall within the depth of field. This should mitigate any potential conflict between accommodation and vergence responses. See Fig. 6.5.

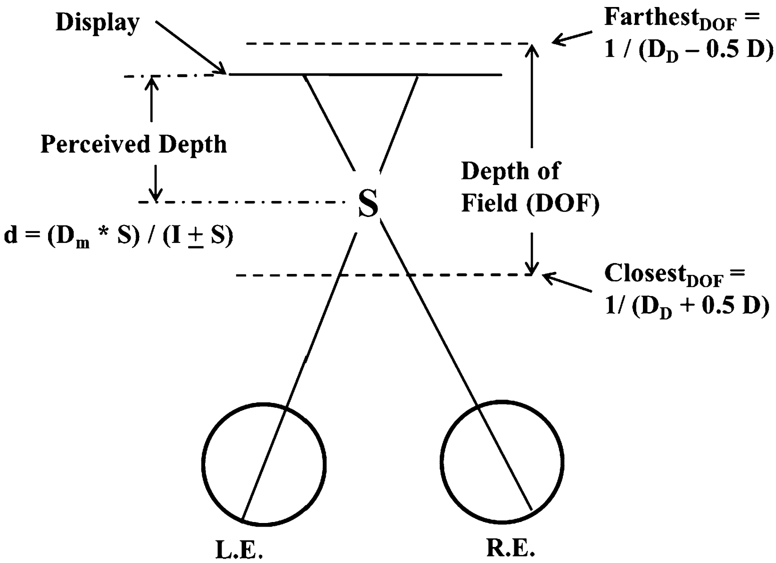


**Fig. 6.5** Depiction of top down view of person viewing S, a virtual stimulus, on a simple stereo display (L.E. = left eye; R.E. = right eye), with perceived depth of S in the crossed disparity (front depth) direction shown as example [in this case, the expression for perceived depth would be: $d = (D_m \times S)/(I + S)$; in this expression, S is separation, different from the virtual stimulus S, as discussed below]. Calculations for the perceived depth of the stereo stimulus are given on left side of the drawing and calculations for the extent of the depth of field are given on the right side of the drawing. On the left side of the drawing, d is predicted depth in meters, $D_m$ is viewing distance in meters, S is lateral separation between the right-eye and left-eye disparate images on the display in meters (not virtual stimulus S), and I is interpupillary distance in meters (Cormack & Fox, 1985b). On the right side of the drawing, Farthest$_{DOF}$ is the estimated farthest extent of the depth of field (in meters), Closest$_{DOF}$ is the estimated closest extent of the depth of field (in meters), and $D_D$ is viewing distance in diopters (D)

Thus, changes in viewing distance affect disparity magnitude differently for stereo displays versus naturalistic viewing. This distinction should have implications for mixed-reality or augmented-reality applications (Patterson, 2007). With mixed-reality and augmented-reality applications, virtual objects seen in stereo are projected into real-world scenes, with the potential for creating perceptual interactions between virtual objects and real objects. If a user has freedom to move around the environment and change egocentric viewing distance, then the perceived depth of the real objects should remain unchanged (due to depth constancy) while the stereo depth of the virtual objects will likely vary (due to a lack of depth constancy). There are other differences between real-world viewing and stereo display viewing. Virtual objects in a stereo display typically do not provide motion parallax cues as do real objects (see below). (Also, the perceived depth of closely adjacent objects can undergo perceptual interaction in the form of perceived attraction or repulsion; Westheimer, 1986; Westheimer and Levi, 1987).

For both real-world viewing and stereo display viewing, vergence likely provides information about egocentric viewing distance for the disparity scaling operation

(Foley, 1980; Owens & Leibowitz, 1976, 1980; von Hofsten, 1976), as might vertical disparity (Gillam, Chambers, & Lawergren, 1988; Gillam & Lawergren, 1983; Rogers & Bradshaw, 1993), but they would do so only for short viewing distances. Field cues, such as linear perspective or texture perspective, might provide distance information for disparity scaling at long distances (Cormak, 1984).

There are neurophysiological studies providing evidence that changes in egocentric viewing distance affect the responsiveness of disparity-sensitive cortical neurons in ways consistent with disparity scaling, such as studies using prisms to manipulate vergence angle (Genovesio & Ferraina, 2004; Trotter, Celebrini, Stricanne, Thorpe, & Imbert, 1996). For example, Dobbins, Jeo, Fiser, and Allman (1998) suggested that such distance scaling operations are likely common to all visual cortical areas in the brain. Genovesio and Ferraina (2004) suggested that such distance scaling operations are used to remap visual input from a retinal coordinate system to a body-centered coordinate system.

This concludes Chap. 6. The next chapter deals with a unique contextual factor.

# Chapter 7
# Contextual Factors, Continued

This chapter covers a unique contextual factor that affects stereo viewing: perceptual constancy. Here, the general concept of perceptual constancy is used to represent a set of three perceptual constancies: size constancy, speed constancy, and depth constancy. These constancies represent stable and veridical perceptions of size, speed, and/or depth when objects in the natural world are seen. Such stable and verdical perceptions occur via a visual cue-integration process that combines information about retinal size, retinal speed, or retinal (binocular) disparity with visual estimates of egocentric viewing distance. Perceptual constancy is a 'contextual' factor because it involves the surrounding conditions that exist (e.g., distance cues) when someone views a stereo display.

## 7.1 Perceptual Constancy

Perceptual constancy used here refers to the perception of stable and veridical characteristics of objects (e.g., their size, speed, or depth) despite changes in the retinal image produced by variation in the observer's viewing distance. This leads us to the distinction between proximal stimulation and distal stimulation.

### *7.1.1 Proximal and Distal Stimuli*

Stimuli can be subdivided into two categories, proximal and distal stimuli (Epstein, 1997). *Proximal stimuli* refer to the images that impinge upon the retina; *distal stimuli* refer to the physical stimuli that exist out in the world. The function of the visual system is to glean knowledge about distal stimuli from the information carried by proximal stimuli, a task made difficult by changing environmental conditions such as changes in egocentric viewing distance that alter the retinal image



R.E. Patterson, *Human Factors of Stereoscopic 3D Displays*,
DOI 10.1007/978-1-4471-6651-1_7

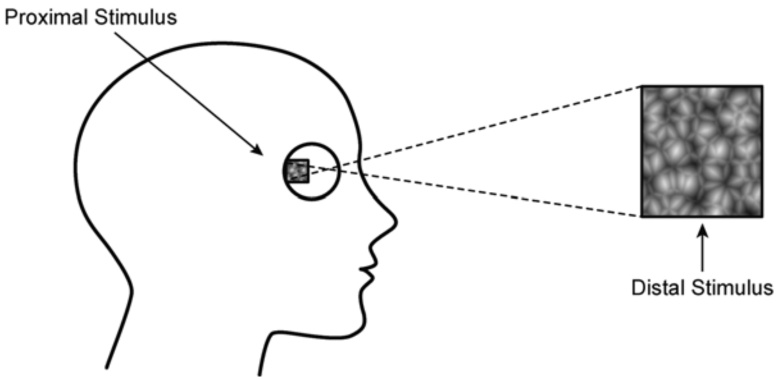


**Fig. 7.1** Diagram depicting proximal stimulus versus distal stimulus. Proximal stimulus refers to the stimulus impinging upon the retina, and the distal stimulus refers to the object out in the environment

(see Fig. 7.1). The solution is that the visual system integrates the various proximal cues of size, speed, and/or disparity with cues to egocentric viewing distance, in order to compute a stable and reliable perception of distal size (size constancy), speed (speed constancy), and/or depth (depth constancy). One way to think about this process is that the proximal cues of size, speed, and/or disparity become recalibrated for different distances.

### *7.1.2 Size Constancy*

In the perception of size, when an observer moves closer to an object, retinal image size increases, and when an observer moves farther away from an object, retinal image size decreases. Specifically, decreasing viewing distance by half will approximately quadruple the area of the retinal image; and doubling viewing distance will decrease the area of the retinal image to be approximately one quarter its original size. In order for an observer to correctly perceive that the object's distal size did not vary (size constancy), the visual system likely combines information about visual angle (proximal stimulus) and cues to egocentric viewing distance (Foley, Ribeiro-Filho, & Da Silva, 2004; see Fig. 7.2).

### *7.1.3 Speed Constancy*

An analogous phenomenon likely occurs with perceived speed. In the perception of speed, when an observer moves closer to a laterally moving object, for example, the speed of the object's retinal image increases, and when an observer moves farther

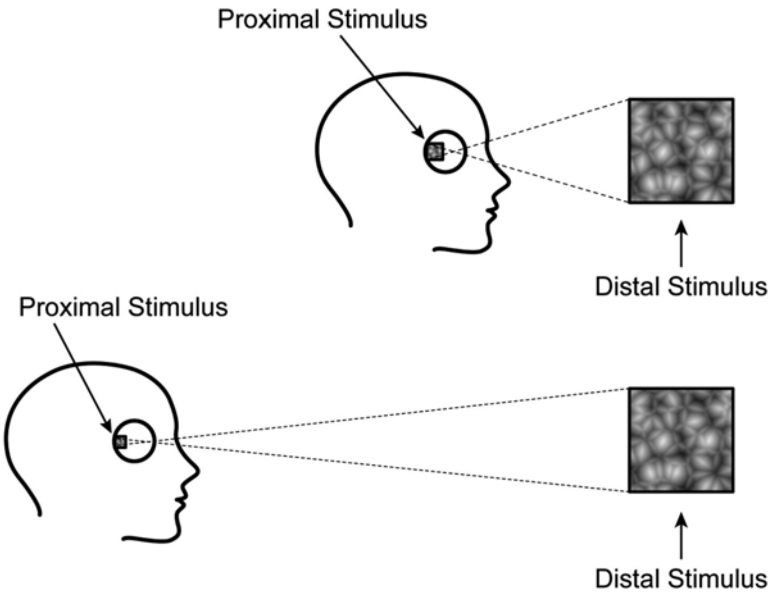


**Fig. 7.2** Drawing showing how changes in viewing distance affect the size of the proximal stimulus (retinal image). Increases in viewing distance (going from *top panel* to *bottom panel*) decrease the size of the proximal stimulus (i.e., retinal image) cast from a constant and stable distal stimulus. In order for the observer to perceive a stable and reliable size of the distal stimulus when the observer changes viewing distance, information about retinal size is likely to be visually combined with cues to egocentric viewing distance

away from a laterally moving object, the speed of the retinal image decreases. Specifically, decreasing viewing distance by half will approximately double the speed of the retinal image; and doubling viewing distance will decrease the speed of the retinal image to be approximately one half its original retinal speed. In order for the observer to correctly perceive that the distal speed of the object did not change (speed constancy), the visual system may combine information about retinal speed (proximal stimulus) and cues to egocentric viewing distance. However, the existence of speed constancy is controversial. Some authors have found evidence for speed constancy (Geri, Pierce, & Patterson, 2008; Wallach, 1939; Zohary & Sittig, 1993), but others have not (McKee & Welch, 1989).

### *7.1.4 Depth Constancy*

In the perception of stereoscopic depth, when an observer moves closer to a pair of objects located in different depth planes, the binocular disparity between the objects increases, and when an observer moves farther away from the pair of objects, binocular disparity decreases (see Fig. 6.4, Chap. 6). Recall that, in real world viewing, decreasing viewing distance by half will approximately quadruple the magnitude of the binocular disparity; and doubling viewing distance will decrease the binocular disparity to be approximately one quarter its original magnitude. In order for an observer to correctly perceive that the depth relations among the distal stimuli in the real world do not vary (depth constancy) when egocentric viewing distance changes, the visual system likely combines information about binocular disparity (proximal stimulus) and cues to egocentric viewing distance (Howard, 2002; Howard & Rogers, 1995; Ono & Comerford, 1977; Patterson, Moe, & Hewitt, 1992; Ritter, 1977). (But also recall that, in stereo display viewing, depth constancy does not occur: increases in viewing distance will make the virtual stimulus to appear in greater depth from the display, and decreases in viewing distance will make the virtual stimulus to appear with lesser depth from the display. This topic was discussed briefly in Chap. 6.)

These three constancies are representative of the kinds of distortions to the proximal (retinal) images that can occur when the observer moves about in a free environment. More realistically, the observer can change body position in ways more complicated than simply varying viewing distance. The observer, can bend his/her body, twist, and change perspective, all while viewing a given stimulus or set of stimuli, which alters the proximal images in systematic fashion. In general, in real-world viewing, it is believed that the visual system likely compensates for these kinds of changes to the proximal image by integrating various sets of visual cues in order to keep perception of the distal world stable and reliable.

### *7.1.5 Distance Scaling with the Perceptual Constancies*

The three constancies mentioned above, namely size, speed, and depth, all involved the perception of spatial extent. Size constancy entails the perception of size which involves spatial extent in two dimensions. Speed constancy entails the perception of speed which involves spatial extent and time. Depth constancy (in real-world viewing) entails the perception of stereo depth which involves spatial extent of the lateral distance between the position of corresponding images in the two eyes (i.e., binocular disparity).

For these three constancies, the visual system is said to scale or recalibrate retinal size, retinal speed, and/or binocular disparity, in accordance with cues to egocentric viewing distance. This process is called *distance scaling*. The cues to egocentric viewing distance used by the visual system for distance scaling are likely to include proprioception from ocular vergence, and field cues such as linear perspective and texture perspective. It is unlikely that proprioception from accommodative effort is used for distance scaling, except at short egocentric viewing distances, for the reasons discussed in Chap. 4.

Proprioception from ocular *vergence* is likely to be a strong distance cue. Von Hofsten (1976) found that perceived distance was determined by relative differences in convergence angle. Owens and Leibowitz (1976, 1980) suggested that vergence may be an important cue to distance perception. And Wetzel, Pierce, and Geri (1996) reported a strong relationship between vergence and perceived size, which implicates vergence as a distance cue in size constancy.

The *field cues* (e.g., linear perspective, texture perspective) may play a role in distance scaling, especially for long viewing distances (Patterson & Martin, 1992). To this author's knowledge, this idea has yet to be tested empirically. Proprioception from *accommodation* is likely not to be a strong distance cue. Although Edgar, Pope, and Craig (1993) reported that people who misaccommodate when viewing virtual displays superimposed upon real scenes also misperceive distance and size, Owens and Leibowitz (1976, 1980) found no relation between accommodation and perceived distance. Given that the depth of field of the eye becomes quite large with increases in egocentric viewing distance, which should minimize the role of accommodative effort and thus proprioception (see Chap. 4), it seems unlikely that accommodation would play a major role in distance scaling except perhaps for short viewing distances.

### *7.1.6 Perceptual Constancies When Viewing Stereo Displays*

These three constancies of size, speed, and depth, all interrelate and interact when viewing in the real world, and the tendency for that interaction is strong when a stereo display is viewed (even if depth constancy does not hold when viewing a stereo display, perceived depth will still covary with viewing distance in systematic fashion). In particular, conditions that make egocentric viewing distance be misperceived or misregistered by the visual system should affect perceived size, perceived speed, and perceived depth all together. If a shorter viewing distance is registered, then virtual size should appear smaller, virtual speed should appear slower, and virtual depth should appear less, relative to the registration of a longer viewing distance. If a longer viewing distance is registered, then virtual size should appear larger, virtual speed should appear faster, and virtual depth should appear greater, relative to a the registration of shorter viewing distance. Displays that cause egocentric viewing distance to be misregistered by the visual system (e.g., vergence drifting toward a resting level due to a low-luminance display) may cause the size, speed, or depth of displayed imagery to be changed, owing to distance scaling based on an incorrectly registered distance (Patterson & Martin, 1992; Patterson, Winterbottom, & Pierce, 2006). Changes in the apparent size, apparent speed, and/or apparent depth of a virtual stimulus may influence perceived distance in reciprocal fashion. These various stimulus attributes of size, speed, depth, and distance all covary together when viewing in the real-world, and the visual system is adept at interrelating them when stereo displays are viewed.

Recall that the remedy for a potential accommodation–vergence conflict entails presenting the stereo depth of the fixated virtual object such that its displayed images fall within the depth of field (Fig. 6.5, Chap. 6). In some circumstances, one may need to decrease the stereo depth of the virtual object so that its displayed images do fall within the depth of field, which may, in turn, require that the useable range of stereo depth be shortened. Decreasing the relative depths of the virtual objects, in turn, may make their size appear smaller, speed appear slower, or distance appear less. In other words, attempting to reduce a potential accommodation–vergence conflict by reducing the perceived depth of a virtual stimulus (so that its displayed images fall within the depth of field) may cause unwanted variation in perceived size, speed, or distance.

Such perceptual effects involving the induction of perceptual constancy operations may not matter with stereo displays depicting alphanumeric symbols, because such symbols, for example, can take on different apparent sizes (depending upon the situation) as long as they are readable. But such perceptual effects may matter greatly with stereo displays depicting real-world scenes, because the size, speed, and depth of the elements in such scenes could be misperceived.

## 7.2 Summary of Contextual Factors

Fine stereoacuity requires fine spatial detail and steady stimulation, and it is likely mediated by the ventral cortical stream. Fine stereoacuity occurs near the horopter and fixation point (i.e., stimulation near the fovea). A large depth range requires coarse spatial detail and moderate luminance temporal modulation, and is likely mediated by the dorsal cortical stream. For both stereo display viewing and real-world viewing, reliable and valid cues to egocentric viewing distance are necessary for the perception of stable and reliable stereo depth. And if the stereo display does not track head position for updating the displayed imagery, the depth perceived is predicted to vary differently from real-world depth when users move around their environment. As an observer moves, the depth position of virtual stimuli may vary while the depth position of physical objects remains constant. This situation may complicate the use of mixed-reality or augmented-reality head-worn displays. More generally, misregistration of distance cues may cause the size, speed, and/or depth of displayed stereo imagery to be misperceived due to the influence of the perceptual constancy processes.

Now that we have covered the contextual factors, we now turn to a discussion of the high-level factors that are relevant to the viewing of stereo displays.

# Chapter 8
# High-Level Factors

This chapter covers three high-level factors that are related to stereo display viewing: high-level cue conflict, intuitive reasoning, and direct manipulation interfaces. These factors are considered 'high level' because they involve cognitive functioning.

Before discussing these high-level factors, we must first address the issue of why I link stereo display viewing with cognition. I do so because information processing is not limited to the visual perceptual system when stereo displays are viewed. Such displays engage higher levels of cognition as well. These higher cognitive levels, in turn, exert profound influence on visual processing. One theoretical context in which to view the connection of stereo viewing to cognition is to posit that the visual system evolved to extract and make meaning from stimulation in the world. Perceiving stimuli presented on a visual display should involve the process of meaning making/meaning extraction (Patterson, 2012).

This meaning making tendency on the part of the brain (which obviously includes the visual system) can be shown in a demonstration following the lead of Michotte (1962; see also Heider & Simmel, 1944). Michotte demonstrated that the relative timing and spatial arrangement of moving elements on a display can induce individuals to mentally project meaning into the elements. In a typical demonstration of this kind for example (Fig. 8.1), a blue disk moves rightward, from the left-hand side of a display, into the center where a stationary red disk is positioned. When the blue disk appears to make contact with the red disk, the blue disk ceases its movement and at the same time the red disk quickly moves rightward to the right-hand side of the display. It is typical for observers to say that the blue disk "caused" the red disk to move. On the other hand, if the red disk begins its movement before the blue disk reaches the center, so that at some point in the sequence both disks are moving rightward at the same time, observers typically say that the blue disk is "chasing" the red disk. Of course the actual stimuli are nothing more than two colored disks presented on a visual display, yet manipulating the relative timing, distance, and size of the disks can induce different attributions of causation and the projection of meaning into the stimuli.



R.E. Patterson, *Human Factors of Stereoscopic 3D Displays*,
DOI 10.1007/978-1-4471-6651-1_8

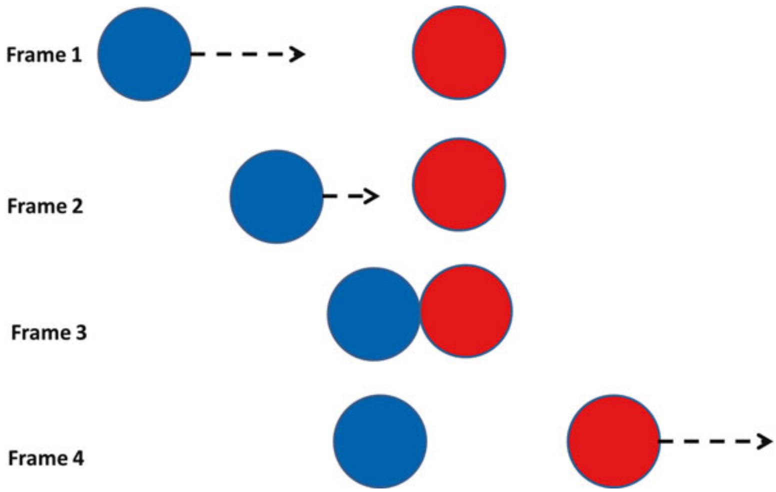


**Fig. 8.1** A depiction of a typical demonstration showing how the relative timing and spatial arrangement among moving elements can induce individuals to mentally project meaning into simple events. In the demonstration depicted here, a *blue disk* moves rightward in Frames 1 and 2 and into the center of the display where a stationary *red disk* is positioned. At exactly the time when the *blue disk* appears to make contact with the *red disk*, in Frame 3, the *blue disk* ceases its movement (and comes to rest in the center of the display) and the *red disk* suddenly begins to move rightward to the right-hand side of the display (Frame 4). It is typical for observers to say that the *blue disk* "caused" the *red disk* to move. On the other hand, if the *red disk* begins its movement before the *blue disk* reaches the center, so that both disks are moving rightward at some point in the sequence, observers typically say that the *blue disk* is "chasing" the *red disk*

Other studies have supported the idea that the visual system has a strong tendency to make/extract meaning from stimulation by showing that visual search and recognition are enhanced by the presence of a meaningful context. Biederman (1972), and Biederman, Glass, and Stacy (1973), showed that the speed or accuracy of identifying a cued object was greater when the object was presented within a coherent real-world scene than when the object was presented in a jumbled scene. Biederman (1981) found that violations of semantic relations of an object in a scene were as readily detected as physical violations, and that extensive semantic processing of a scene was readily achieved from a single fixation. Oliva and Torralba (2006) also reported that the semantic information in a scene could be extracted from a 200-ms exposure, equivalent to a single fixation. See Weisstein and Harris (1974), Weisstein, Williams, and Harris (1982), Enns and Rensink (1990), Grill-Spector and Kanwisher (2005), and Thorpe, Fize, and Marlot (1996) for related research. This idea of meaning extraction is consistent with the well-known tendency of humans to remember the meaning and interpretation of pictures but forget much of their physical details even after only a brief time period (Gernsbacher, 1985; Mandler & Ritchey, 1977).

Thus, users of stereo displays would be expected to possess a natural tendency to make/extract meaning from visual stimulation. This suggests that certain kinds of stereo displays—immersive stereo displays composed of naturalistic scenes and cues—likely engage our cognitive systems in a particular way. Specifically, such stereo displays may engage our *intuitive reasoning system*. This, in turn, may lead to viewing discomfort, an idea developed in the next section. To this author's knowledge, some of these ideas have yet to be conclusively supported.

## 8.1 High-Level Cue Conflict

Immersive stereo displays attempt to re-create real-world scenes by presenting various cues to depth and distance. These cues may include binocular disparity, motion parallax, linear perspective, and texture perspective. In creating this kind of display, it is important that the various cues convey the same magnitude of distance and depth, and thus the cues should be 'in registration' for ease of viewing. If not, then discomfort is likely to occur due to cue conflict.

Patterson and Silzars (2009) discussed the concept of high-level cue conflict and considered a hypothetical example of an individual who is viewing a stereo display that is depicting a football game in which players run the length of a football field. In this example, the binocular disparity information would convey depth intervals of only inches or feet in magnitude because disparities corresponding to larger magnitudes of depth would not be fusible. Yet, in this example, the linear and texture perspective in the scene would convey depths of tens of yards, consistent with a football field. Patterson and Silzars suggested that, in this example, the cue conflict between disparity (inches or feet of depth) and perspective (tens of yards) would create viewing discomfort. More generally, Patterson and Silzars argued that cue conflict is particularly severe with misregistered motion parallax and binocular parallax because these are the two strongest immersive depth cues under most viewing conditions.

### *8.1.1 Motion Parallax*

In real-world viewing, even small body movements create very noticeable parallax shifts, called motion parallax, which is a key source of relative depth information. Motion parallax refers to the depth perception induced by the apparent movement of objects relative to the observer that is produced by body movements—such as body sway or locomotion. Over the years, a number of authors have noted the significance of depth information from motion parallax (Gibson, 1950, 1966, 1979; Simpson, 1993; von Helmholtz, 1866). Interestingly, Nawrot (2003) and Nawrot and Joyce (2006) have suggested that pursuit eye movements are necessary for the veridical perception of depth from motion parallax.

For an example of motion parallax, consider Fig. 8.2 (see also Blake & Sekuler, 2005). Here, an observer (oval marked "O") is fixed to a moving vehicle (note that the same analysis applies if the observer performs head movements or self-locomotes

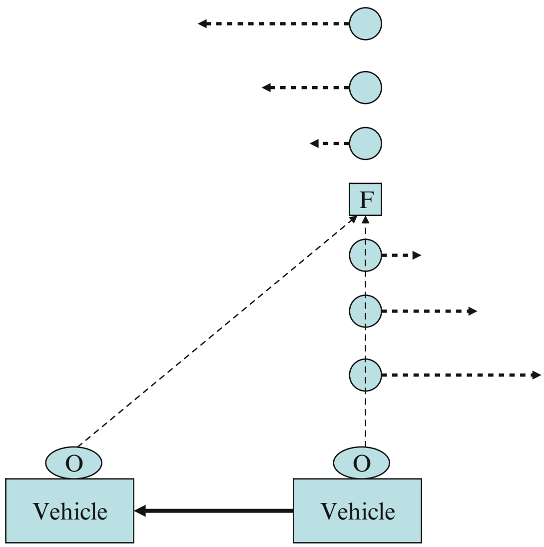


**Fig. 8.2** Top down view of the type of situation that yields motion parallax information to a moving observer. In this example, an observer (*oval* marked "O") is fixed to a moving vehicle which moves laterally while the observer, whose line of sight is perpendicular to the direction of vehicle movement, fixates object F. Three arbitrary objects (*circles*) are positioned at different depths in front of fixation, and three objects (*circles*) are positioned at different depths behind fixation. The *dashed arrows* attached to the various objects indicate the velocity of the apparent movement of the objects relative to the observer as the observer locomotes. Objects positioned at increasingly closer depths to the observer and in front of fixation appear to move faster in a direction opposite to the observer's motion, while objects positioned at increasingly farther depths behind fixation appear to move faster in the same direction as the observer. The apparent relative movement of stationary objects fixed to a rigid scene provides relative depth information to a locomoting observer. Reproduced from Figure 3 of Patterson and Silzars (2009), Immersive stereo displays, intuitive reasoning, and cognitive engineering. Journal of the Society for Information Display, 17, 443–448. Copyright Society for Information Display. By permission

without a vehicle). The observer, whose line of sight is perpendicular to the direction of vehicle movement, fixates object F. Arbitrarily, three objects (circles) are shown positioned at different depths in front of fixation, and three objects (again circles) are positioned at different depths behind fixation. The dashed arrows attached to the various objects indicate the velocity of the apparent movement of the objects *relative to the observer* as the observer locomotes. It can be seen that objects positioned at increasingly closer depths to the observer and in front of fixation appear to move faster in a direction opposite to the observer's motion, while objects positioned at increasingly farther depths behind fixation appear to move faster in the same direction as the observer (i.e., the latter objects' apparent movement is due to the rotational eye movement that occurs when the observer maintains fixation on F while translating laterally).

When the observer moves from one position to the next position, there are portions of the background scene that are initially hidden from view by the objects but which become revealed as the observer reaches his or her new position; this is termed '*dynamic disclosure*' (Patterson & Silzars, 2009). Moreover, there are also portions of the background scene that are initially visible to the observer but which become hidden from view by the objects as the observer reaches the new position; this is termed '*dynamic occlusion*'. Consistent with the concepts of dynamic disclosure and occlusion, the visual system interprets the relative movement of the objects as being due to observer's self-motion—*the objects themselves appear fixed to a rigid scene* as the observer moves past them (and the relative motion of the objects, as noted above, provides relative depth information).

For example, imagine that an observer is a passenger inside a moving automobile traveling along a highway. Also imagine that the passenger turns her head to the right to look out the side window at a scene containing, in increasing distance away from the observer, a picket fence, a horse, a farm house, and a mountain range located at a far distance behind the farm house. Assume that the observer fixates the farm house.

Because the picket fence is closest to the observer, it will appear to undergo the fastest opposing relative motion; the horse, being slightly farther away from the observer, will appear to undergo slightly slower opposing relative motion; and because the house is fixated, it will appear to undergo no relative motion. The mountain range, being at a far distance behind the farm house, will appear to undergo congruent relative motion and appear to move in the direction of the observer's self-motion. Moreover, patches of the ground that are initially hidden from view by the fence, horse, and house will become revealed as the observer travels down the highway (dynamic disclosure), while at the same time portions of the ground that are initially visible to the observer will become hidden from view by the fence, horse, and house (dynamic occlusion).

The visual system will interpret the apparent relative movement of the fence, horse, house, and mountain range—motion parallax—as being due to the observer's self-motion. These objects will appear fixed to the scene as the observer moves past them, and their apparent relative motion will provide relative depth information to the locomoting observer.

### *8.1.2 Motion Parallax and Heading Control*

Motion parallax is thought to support the ability of individuals to control heading when locomoting (e.g., Patterson et al., 2006). One important cue for determining heading is the *optic flow* produced when an observer moves through a textured environment (Gibson, 1950). When an observer moves forward and maintains fixation, the pattern of relative motion in the retinal images contains a point from which the relative motion vectors radiate outward. This point is called the *focus of expansion or FOE* (see Fig. 8.3) and corresponds to the location in the environment toward which the observer is heading at that moment. This pattern of relative motion may be used by individuals to steer vehicles, direct locomotion, and control postural

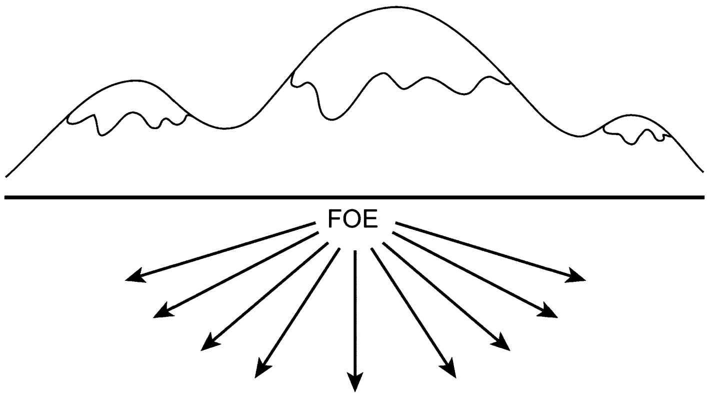


**Fig. 8.3** Drawing depicting optic flow and the focus of expansion. The *arrows* depict the direction of apparent relative motion of stationary elements in a scene when an observer translates in the forward direction (i.e., viewer translating forward toward the mountains in the distance). The point of origin of the relative motion vectors is called the focus of expansion (FOE) which gives the instantaneous heading

balance (Bardy, Warren, & Kay, 1996, 1999; Warren, 1998). This pattern of relative motion selectively activates specialized regions in extrastriate visual cortex (see, e.g., Vanduffel et al., 2002).

When an observer travels on a straight path without eye or head movements, the translational information contained in the retinal images coming from the optic flow is sufficient for the observer to recover heading direction (Warren, 1998). When an observer moves on a curved path, or on a straight path while making eye or head movements, the flow pattern on the retina (retinal flow) will be a combination of a translational component and a rotational component that interferes with the recovery of heading from optic flow. In this latter case, an observer must decompose the retinal flow into its rotational and translational components (e.g., Li & Warren, 2000, 2002; Royden, Crowell, & Banks, 1994; Warren, 1998). The visual system can use the motion parallax information for performing this decomposition to recover heading direction (Longuet-Higgins & Prazdny, 1980; Rieger & Lawton, 1985). Patterson et al. (2006) found that increasing the density and vertical extension of object contours in a simulated real-world environment improved heading control, presumably due to the increased motion parallax information conveyed by the added contour information. Interestingly, heading control may be mediated by activity of a visual pathway (i.e., the dorsal cortical stream) different from the pathway (ventral cortical stream) that supports perceptual judgments of heading (e.g., Aglioti, DeSouza, & Goodale, 1995; Bridgeman, Gemmer, Forsman, & Huemer, 2000; Goodale, Milner, Jakobson, & Carey, 1991; Milner & Goodale, 1995).

The significance of motion parallax both for the control of heading as well as for conveying immersive depth (akin to binocular parallax) underscores the important role that motion parallax plays in visual functioning.

### 8.1.3 Motion Parallax/Binocular Parallax Conflict

When an individual locomotes, the apparent differential movement of objects in different depth planes *relative to the individual* (i.e., motion parallax) is part of what our brain uses to place us in a stable environment. And it is this relative movement that is absent from many, if not most, types of stereoscopic displays, including 3D movies. For example, when motion parallax cues are absent from a stereo display, the apparent movement of a given virtual object located in depth in front of the display will seem to follow the direction of the observer's self-motion, rather than be in a direction opposite to the observer's self-motion (the latter of which would be expected from motion parallax). That is, there is no apparent motion parallax-related movements of objects when viewed from different positions at different moments in time (see Fig. 8.4). There is no dynamic disclosure or occlusion of background elements. The visual system interprets the location of the virtual object as occupying a fixed position along the observer's line of sight—when the observer moves in a given lateral direction, so does the virtual 3D object.

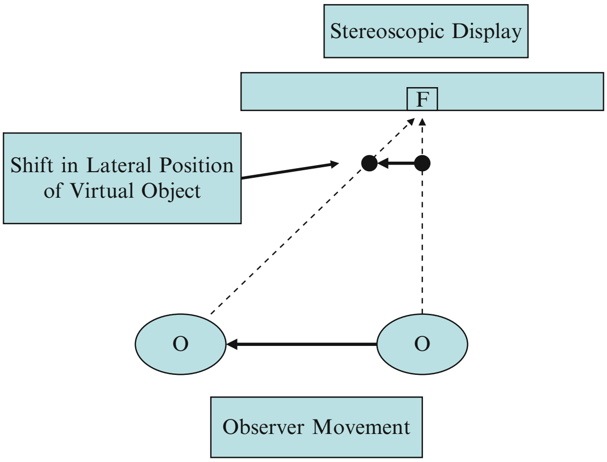


**Fig. 8.4** The shift in lateral position of a virtual object whose depth is simulated in a typical stereoscopic display; top down view. The diagram depicts an observer (*oval* marked "O") who performs a leftward lateral movement while maintaining fixation on a virtual object which appears in depth in front of the stereoscopic display. ('F' depicts the point of the observer's line of sight when projected back to the display.) Because the visual system interprets the virtual object as being positioned along the observer's line of sight as she moves laterally, the virtual object will be perceived as moving laterally in the direction of the observer's self-movement. This shift in the lateral position of the virtual object is not motion parallax. Reproduced from Figure 4 of Patterson and Silzars (2009), Immersive stereo displays, intuitive reasoning, and cognitive engineering. Journal of the Society for Information Display, 17, 443–448. Copyright Society for Information Display. By permission

Thus, when an individual makes a lateral head or body movement while watching a stereo movie, such as when changing his or her sitting posture, there is no corresponding motion parallax information represented in the 3D movie as there would be when viewing objects out in the real world. (Note that, even at the recommended 20-m distance when viewing 3D cinema, a lateral shift in head position of 1 cm would produce a parallax angle of approximately 1.7 arcmin, which would likely be above the observer's detection threshold.) Patterson and Silzars (2009) suggested that this absence of motion parallax in 3D movies, or other stereo displays, can create high-level cue conflict between motion parallax (represented as a zero value to the visual system) and binocular parallax (represented as a non-zero value). The remedy for this high-level cue conflict is either to: (1) track the observer's head position and update the content of the stereo display accordingly so as to present motion parallax information congruent with the head movement and with the binocular parallax present in the display; or (2) restrain the head and body from making movements, thus in theory there would be no need to track the observer's head position and update the content of the stereo display—the motion parallax information should be zero because there would be no head movement.

With respect to #1 above, tracking the observer's head position and updating the content of the stereo display so that motion parallax is congruent with binocular parallax, Merritt, Cole, and Ikehara (1991) have performed such a study (see also Merritt, 1987, 1989, 2011). Merritt et al. (1991) investigated the effects of binocular parallax (stereo) and/or motion parallax on a rapid-sequential-positioning remote-manipulation task. These authors found that the presence of binocular parallax produced better performance than no binocular parallax, either with or without the presence of motion parallax; motion parallax added a small performance benefit over binocular parallax. The results of this study suggested that binocular parallax and motion parallax do not combine in an additive fashion in affecting remote-manipulation task performance, at least under the conditions tested. Nonetheless, because observer discomfort was not assessed in the Merritt et al. study, it is still possible that high-level cue conflict between the presence of binocular parallax and the absence of motion parallax in certain conditions in the Merritt et al. study may have created discomfort, as suggested by Patterson and Silzars.

Patterson and Silzars (2009) suggested that the discomfort experienced by many observers when the typical stereo display is viewed, or when a 3D movie is seen, is likely due to the presence of high-level cue conflict as discussed above, rather than due to accommodation-vergence conflict, except for short egocentric viewing distances. Patterson and Silzars argued that high-level cue conflict, between the presence of binocular parallax and the absence of motion parallax, was one primary reason for the discomfort typically experienced when many stereo displays (and 3D movies) are viewed, rather than a ubiquitous conflict between oculomotor (accommodation-vergence) responses. One potential reason why high-level cue conflict between the two parallax cues may lead to discomfort is discussed next.

## 8.2 Intuitive Reasoning

It is possible that this high-level cue conflict between the presence of binocular parallax and the absence of motion parallax may present problems for cognitive functioning. To develop this idea fully, we need to discuss a two-process theory of reasoning and decision making articulated in the contemporary cognitive psychology literature (see Patterson, Pierce, Bell, Andrews, & Winterbottom, 2009; Patterson & Silzars, 2009).

A number of authors (e.g., Evans, 2003, 2008; Evans & Stanovich, 2013; Hammond, 2007; Hogarth, 2001, 2002; Kahneman & Frederick, 2002; Sloman, 1996; Stanovich & West, 2000, 2002) have proposed that human reasoning and decision making is composed of a blend of two types of complementary processes or systems: an *analytical process* and an *intuitive process* (see Fig. 8.5).

Analytical reasoning and decision making refers to the rendering of conscious decisions that entail deliberation and the contrasting of options and an assessment of their likelihood and possible consequences. The analytical process is slow, deliberative, reflective, effortful, and it entails the operation of (declarative) working memory, whose capacity is limited (for discussion of working memory, see Baddeley & Hitch, 1974; Miyake & Shah, 1999). Intuitive reasoning and decision making refers to the rendering of largely unconscious decisions based on situational pattern

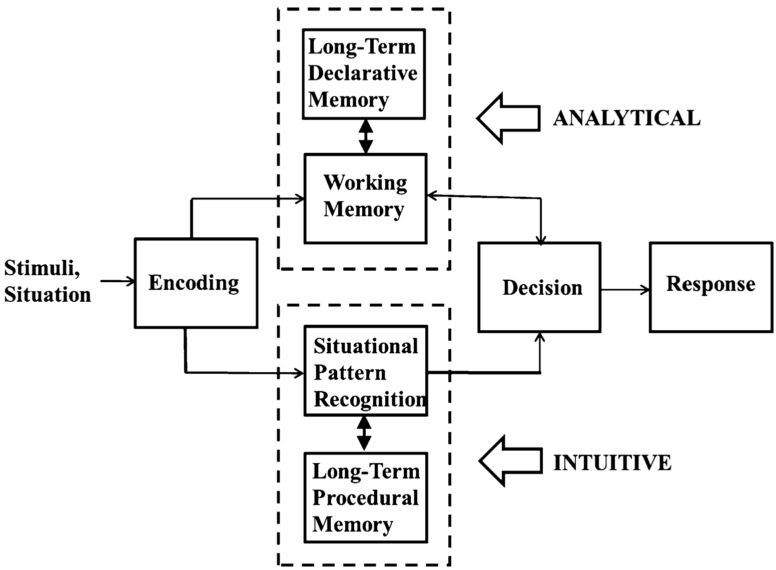


**Fig. 8.5** Diagram of dual-process reasoning and decision-making model. Both the analytic reasoning and decision-making system (including working memory and long-term declarative memory, shown in *upper dashed rectangle*) and the intuitive reasoning and decision-making system (including situational pattern recognition and long-term procedural memory, shown in *lower dashed rectangle*) are depicted

**Table 8.1** Environmental and task characteristics for engaging two different types of reasoning and decision making as discussed by Hammond et al. (1997)

| Analytical | Intuitive |
|---|---|
| Few cues | Many cues |
| Objective measurement | Perceptual measurement |
| Discrete or unknown values | Continuous values |
| Low redundancy | High redundancy |
| Sequential display | Simultaneous display |

The characteristics of intuitive reasoning and decision making match the characteristics of immersive stereoscopic displays

recognition without contrasting or deliberating options. The intuitive system is fast, automatic, implicit, relatively effortless, related to high-level perception, and does not involve working memory (it likely involves procedural memory). In most everyday situations, these two systems are thought to be involved in some blended way during reasoning and decision making, although different situational contexts may selectively engage one or the other process.

Patterson and Silzars (2009) suggested that the stimulus properties that engage intuitive reasoning correspond well with the type of information and cues provided by immersive stereo displays (i.e., stereo displays that depict simulated real-world scenes, such as 3D movies). This is because the intuitive process is thought to yield unconscious inferences and predictions based on previously experienced situational patterns and associations (i.e., statistical regularities) in the world; the intuitive process produces judgments and decisions based on recognition and similarity. Accordingly, intuitive reasoning should be strongly activated by immersive (i.e., simulated real-world) tasks and displays that entail the perception of simultaneous, redundant situational cues (Hammond, 2007; Hammond, Hamm, Grassia, & Pearson, 1997). Table 8.1 presents several of the relevant stimulus properties for inducing intuitive reasoning and decision making as discussed by Hammond et al. (1997).

One reason why high-level cue conflict between binocular parallax and motion parallax may lead to discomfort with the viewing of stereo displays, such as with 3D movies, is that the intuitive reasoning system is attempting, and failing, to make reasoned sense out of the incoming conflicted perceptual information (Patterson & Silzars, 2009). This conflict may be most serious when it exists between motion parallax and binocular parallax because these two parallax cues are the most perceptually immersive. But conflict between other cues, such as linear perspective versus binocular parallax, may also be important.

Patterson and Silzars (2009) proposed that information displays should be subdivided into two categories: analytical reasoning-inducing displays (which would entail alphanumeric symbology), and intuitive reasoning-inducing displays (which would involve immersive displays, including stereo displays). High-level cue conflict presumably would be particularly important with the intuitive reasoning-inducing displays (e.g., immersive stereo displays).

Another form of cue conflict was reported by Stuart et al. (2009), who describe an evaluation of a helmet-worn display (HWD) by rotary wing pilots. This HWD projected imagery onto the visor of the helmet, which magnified the binocular disparity, thus distorting cues to depth and distance. Stuart et al. reported that that the pilots, with practice, learned to discount the conflicting/unreliable sources of depth information. The authors speculated that the pilots learned to discount the distorted binocular cues in favour of the veridical monocular cues, such as familiar size, motion parallax and linear perspective. Thus, it may be that individuals who have practice viewing a given stereo display may learn to discount any cue conflict between the presence of binocular parallax and the absence of motion parallax, similar to the rotary wing pilots in the Stuart et al. study. This result should be replicated in further studies under controlled conditions. Nonetheless, such learning, if it occurs, may take a significant investment in time and thus it is not clear how practical such a solution would be for many situations in which conventional stereo displays (e.g., 3D movies) are viewed.

## 8.3 Direct Manipulation Interfaces

Direct-manipulation interfaces refers to interfaces in which the sensing of the position of the user's hand, arm, handheld stylus, or head (i.e., head tracking) will enable the user to manipulate or interact with the virtual objects or scenes presented on the display (Hutchens, Hollan, & Norman, 1985). Thus, in the present context, direct-manipulation interfaces refer to *interactive stereo displays*. Specific to purposes of the present discussion, an interactive stereo display that included head tracking could offer a solution to the problem of high-level cue conflict: one could track the user's head position and then use that information to update the virtual objects or scenes presented on the stereo display in a way that would be consistent with the amount of motion parallax that would normally be experienced, as Merritt et al. (1991) have done. As the user moves his/her head, the virtual objects or scene would be presented from a slightly different perspective, the exact amount of which would be determined by the head movement. In other words, the tracking of the user's head movement would be employed to create an appropriate amount of motion parallax information (Merritt et al., 1991).

It is likely that the presence of motion parallax, together with binocular parallax, in an interactive stereo display would minimize or prevent the occurrence of the high-level cue conflict that was deemed serious by Patterson and Silzars (2009). The reduction or elimination of high-level cue conflict, in turn, could increase the viewing comfort and the quality of the sense of depth and immersion when an interactive stereo display is viewed. To this author's knowledge, these suggestions have yet to be empirically tested.

# Chapter 9
# High-Level Factors, Continued

This chapter covers three high-level factors that are related to stereo display viewing: hand/arm tracking and proprioception, interactive stereo displays and spatial reasoning, and spatial mental models and working memory. These factors are considered 'high level' because, like the factors discussed in Chap. 8, they involve cognitive functioning.

## 9.1 Hand/Arm Tracking and Proprioception

At the end of the last chapter, we discussed how direct-manipulation interfaces involve interfaces in which the sensing of the position of the user's hand, arm, or head will enable the user to manipulate or interact with the virtual objects or scenes presented on the display (Hutchens, Hollan, & Norman, 1985). We also discussed how an interactive stereo display, which is a direct-manipulation interface combined with stereo viewing, may offer a solution to the problem of high-level cue conflict by tracking the user's head position and then using that information to update the virtual objects or scenes presented on the display, consistent with the amount of motion parallax that would normally be experienced (Merritt, Cole, & Ikehara, 1991).

An interactive stereo display may also promote an increased sense of spatial relations among the virtual objects in a scene owing to *proprioception*. Proprioception refers to the unconscious sensing of the position, movement, and spatial orientation arising from one's own body without the use of other senses (e.g. vision, audition). Proprioception and the kinesthetic sense are two terms often seen as interrelated, yet there is considerable disagreement regarding their definitions. Here, I will use the terms as defined by Kandel, Schwartz, and Jessel (2000): There are two submodalities of proprioception: the limb-position sense, which involves the sensing of the stationary position of limbs, and kinesthesia, which entails the sensing of the movements of the limbs (Kandel et al., 2000).



R.E. Patterson, *Human Factors of Stereoscopic 3D Displays*,
DOI 10.1007/978-1-4471-6651-1_9

Kandel et al. discuss how proprioception and its submodality of kinesthesia are important for controlling the movements of the limbs, manipulating objects, and maintaining posture. This is achieved by the activity of proprioceptors in the muscles that monitor their length, tension, and pressure, and the presence of noxious stimuli. The muscle spindles, one kind of proprioceptor, convey information about the length of the muscle and the velocity of its stretch in order to register information about changes in muscle angle and position. The golgi tendon organ, another proprioceptor, registers information about muscle tension. There are also proprioceptors registering information from joints and ligaments. Figure 9.1 shows the four lobes of the brain and the cerebellum. The cerebellum is responsible for processing the unconscious aspects of proprioception/kinesthesia.

Proprioception and its submodality kinesthesia may provide information about the spatial relations among virtual objects depicted with an interactive stereo display through the cumulative signals about body position and movement as one interacts with the virtual objects over time. Keehner, Khooshabeh, and Hegarty (2008) suggested, along a related vein, that one advantage of an interactive stereo display is that information about spatial relations can come from the motor commands made to control the display. In support of this idea, Keehner et al. cite Philbeck, Klatzky, Behrmann, Loomis, and Goodridge (2001), who have shown that motor command signals during active exploration may provide cues about navigation through space. The sensing of body position and movement through proprioception/kinesthesia, as well as the commands for executing motor movements themselves, may provide information about spatial relations when an interactive stereo display is used.

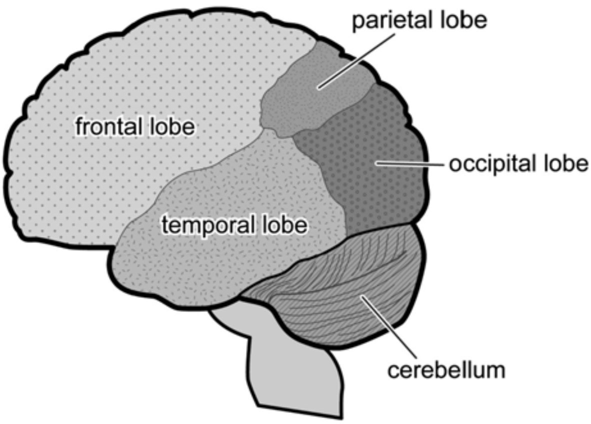


**Fig. 9.1** Diagram of the brain (left side visible) showing the four lobes and the cerebellum. The cerebellum is responsible for processing the unconscious aspects of proprioception and kinesthesia, that is the unconscious sensing of the position, movement, and spatial orientation arising from one's own body. Proprioception is important for sensing and controlling the movements of the limbs, manipulating objects, and maintaining posture. Proprioception may also provide information about spatial relations when interactive stereo displays are used

This information about spatial relations via proprioception/kinesthesia, in turn, may be neurally combined with the binocular parallax and motion parallax present when an interactive stereo display is viewed. The use of an interactive stereo display may enable proprioception/kinesthesia to enhance the sense of spatial relations among the virtual objects by reinforcing and bolstering the presence of (congruent) depth cues coming from binocular parallax and motion parallax.

Another advantage of an interactive stereo display comes from Hutchens et al. (1985), who suggested that a reduction in cognitive effort can occur when using a direct-manipulation interface because the capability of such an interface to accomplish a given task can be well matched to the user's goals. Finally, Keehner et al. (2008) suggested that consistent advantages should be obtained when using a direct-manipulation interface provided that relatively simple stimuli are used (e.g., familiar objects) and the tasks involve simple visual recognition or inspection. This topic of employing an interactive stereo display for various tasks deserves to attract much research in the future.

Lloyd and Nigus (2012) reported, in a study investigating aerial refueling techniques, that ratings of visual comfort were higher when head tracking was used with a stereoscopic display, but such ratings were lower when head tracking was used with a non-stereoscopic display. This study provided empirical evidence in support of the hypothesis that head-tracking is beneficial for stereoscopic displays in terms of visual comfort. In general, the use of direct-manipulation interfaces (e.g., for generating motion parallax) should improve the usability of stereo displays (Merritt et al., 1991).

## 9.2 Interactive Stereo Displays and Spatial Reasoning

Spatial reasoning refers to the ability to reason and make judgments about objects and their spatial relations (Gardner, 1993/1983; Renz & Nebel, 2007). There are two aspects of spatial reasoning (Anderson, 2010): (1) *visual*, which entails judgments about the visual details of an object (or objects) like its color or size (Moyer, 1973; Thompson & Kosslyn, 2000); and (2) *spatial*, which involves judgments about the position or location of an object, or about the spatial relations among parts of an object, or of multiple objects (Brooks, 1968; Roland & Friberg, 1985). Stereo displays that include direct-manipulation interfaces should aid both aspects of spatial reasoning.

### *9.2.1 Visual Aspect*

There are tasks that involve the visual aspect of spatial reasoning for which interactive stereo displays may be beneficial. One could use an interactive stereo display to manipulate a virtual object in order to scan its features, or one could rotate or manipulate a set of displayed objects in order to compare their features.

Memory for visual detail fades very rapidly, even within the first tens of seconds (Gernsbacher, 1985; Mandler & Ritchey, 1977). To counter this memory fading, a virtual object could be manipulated so as to bring into view its relevant parts and the user would not have to rely on visual memory—the user can just directly view the virtual object. Employing an interactive stereo display for viewing certain visual details could be used for offloading the burden placed on memory for those details.

### 9.2.2 *Spatial Aspect*

There are tasks that involve the spatial aspect of spatial reasoning for which an interactive stereo display may be beneficial. An interactive stereo display could be used to manipulate a virtual object through three dimensions in order to clearly see its parts, or to manipulate a set of virtual objects through three dimensions in order to see their spatial layout. This, in turn, could enhance the speed and accuracy of judgments made about the object(s) because it takes cognitive effort to manipulate objects mentally via visual imagery. Shepard and Metzler (1971; see Fig. 9.2) reported that the time needed to determine that two perspective drawings portray the same (or different) three-dimensional object increased with the angular difference between the two portrayed orientations. Studies have shown that individuals would rather rotate shapes on a display rather than rotate them mentally (Kirsh & Maglio, 1994).

Using an interactive stereo display to manipulate virtual objects could eliminate the need to manipulate the objects mentally. Ragan, Kopper, Schuchardt, and Bowman (2013) investigated the ability of individuals to distinguish structural gaps and intersections between components of 3D models that simulated an underground cave system. The authors found that individuals made significantly fewer errors with either an increased field of regard or with the addition of head-tracked rendering. The results also indicated that the individuals performed significantly faster when the system provided the combination of stereo and head-tracked rendering. Luursema, Verwey, Kommers, and Annema (2008) showed that computer-implemented stereopsis, together with dynamic exploration, provided a significant benefit for the learning of human anatomy.

Another type of task for which interactive stereo displays may be beneficial is the alignment of the relationships among different kinds of maps (geospatial alignment). To discuss this idea further, several distinctions must first be made among (1) route map, (2) survey map, (3) an egocentric representation of space, and (4) an allocentric representation of space (Anderson, 2010; Burgess, 2006; Klatzky, 1998). A route map refers to a path linking up locations but it contains no spatial information (Anderson, 2010). An example of a route map would be a drawing showing a depiction of your driveway, an arrow pointing left followed by a stop sign, and then an arrow pointing rightward showing an entrance to a highway. It would be a drawing showing a path that linked up different locations (driveway, stop sign, highway) but it would be devoid of any real spatial information—the spatial relations among

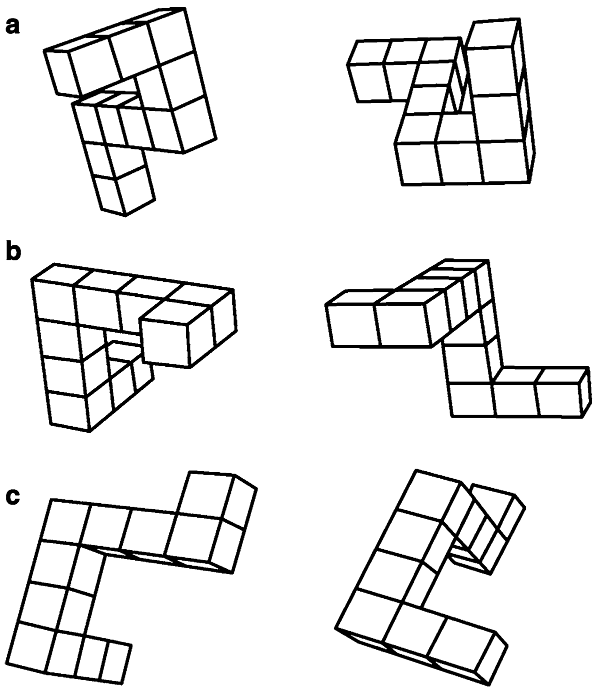


**Fig. 9.2** Mental rotation. Shepard and Metzler (1971) showed that the time needed to determine that two perspective drawings portrayed the same (or different) three-dimensional object increased with the angular difference between the two portrayed orientations. Are the objects the same in Panel (**a**)? Panel (**b**)? Panel (**c**)? Yes, yes, and no. Reproduced from Figure 1 of Shepard, R.N. & Metzler, J. (1971). Mental rotation of three-dimensional objects. Science, 171, 701–703. Copyright The American Association for the Advancement of Science. By permission

the locations would not be preserved (see Fig. 9.3, left panel). A survey map, on the other hand, is a representation of locations that preserves their spatial relations (Anderson, 2010). An example of a survey map would be a map of the city in which you live (Fig. 9.3, right panel).

An egocentric representation of space (Burgess, 2006; Klatzky, 1998) refers to space as viewed from the observer's perspective. An example of an egocentric representation would be the viewing of objects in your front yard relative to your own location. An allocentric representation of space is space independent of the observer's perspective or position; it involves representing the location of an object relative to the location of other objects (Anderson, 2010). An example of an allocentric representation would be the intrinsic spatial relations among the objects in your front yard independent of your location.

Turning back to geospatial alignment, one could use an interactive stereo display to align an egocentric representation of a set of objects or scene with an allocentric representation, such as a survey map. Using an interactive stereo display to align egocentric and allocentric maps could improve the speed and accuracy of judgments in certain applied contexts. People find it difficult to mentally integrate different

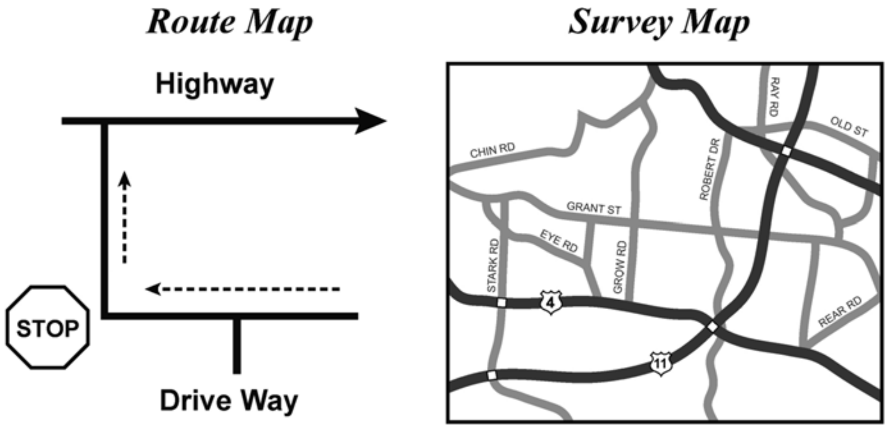


**Fig. 9.3** Route map on the *left*, which is a path linking up locations but it contains no spatial relations. Survey map on the *right*, which is an arrangement of locations that preserves spatial relations

representations of space. The degree to which an allocentric map is rotated from an egocentric viewpoint increases the difficulty of navigation (Gugerty, deBoom, Jenkins, & Morley 2000). The time taken to recognize photographs from different viewpoints around an environment increases linearly with the difference in angle between the egocentric viewpoint and the other viewpoints (Diwadkar & McNamara, 1997). Thus, individuals will typically mentally rotate a map when it is incongruent with their egocentric viewpoint, which takes time and is error prone (Anderson, 2010). An example for which geospatial alignment would be useful would be an individual using an interactive stereo display to manipulate a survey map that is being created or embellished with egocentric information from another source. More generally, an interactive stereo display can be used to create and align multiple sets of objects for a variety of different purposes.

A final type of task for which interactive stereo displays could be beneficial is the creation of spatial layouts among set of objects derived from the translation of words (Anderson, 2010; Taylor & Tversky, 1992). Using an interactive stereo display in this way may increase the speed and accuracy of the layout construction.

In short, interactive stereo displays allow the user to manipulate the spatial attributes (e.g., orientation, angle of view) of displayed objects and scenes presented on the display. This, in turn, could enhance the user's spatial reasoning abilities in a wide variety of tasks.

## 9.3 Spatial Mental Models and Working Memory

Spatial reasoning likely involves the generation of mental models. A *mental model* refers to an organized knowledge structure that involves imagined possibilities and projection (Johnson-Laird, 2010; Klein & Hoffman, 2008). Theoretical approaches that attempt to explain human reasoning include the ideas that reasoning entails: a form of mental logic from which inferences are drawn (O'Brien, 2009); rules of inference relating to social exchange (Cosmides, Tooby, Fiddick, & Bryant, 2005); mental computation of probabilities (Oaksford & Chater, 2001); or mental models of imagined possibilities from which inferences are developed (Johnson-Laird & Byrne, 2002). There is empirical support for the mental model theory of human reasoning (Byrne & Johnson-Laird, 2009; Johnson-Laird, 2010).

The generation of a spatial mental model is a mentally-demanding process that likely depends upon *working memory*. Brunye and Taylor (2008) found that visuospatial and central-executive working-memory processes were involved in the creation of spatial mental models of route descriptions (geographical relations from an egocentric perspective) or survey descriptions (geographical relations from an allocentric perspective). It is possible that interactive stereo displays can enhance the visuospatial and central-executive working-memory processes by making spatial relationships in three dimensions more richly defined, which could lessen the demand on working memory.

### *9.3.1 Working Memory*

Many authors (Baddeley, 2003; Conway, Jarrold, Kane, Miyake, & Towse, 2008; Cowan, 1995, 1999; Kane, Conway, Hambrick, & Engle, 2008; Miyake & Shah, 1999) believe that high-level cognitive abilities (e.g., planning, reasoning, problem-solving) are dependent upon a conscious process called *working memory*. Working memory plays a role in the cognitive operations of maintaining and regulating task-relevant information in an active state (Gazzaniga, Ivry, & Mangun, 2002). The core of working memory is the *central executive* (or the supervisory attention system; Norman & Shallice, 1986). The central executive (Fig. 9.4) activates memory representations of goal states, switches attention to coordinate simultaneously-performed multiple tasks (Baddeley & Logie, 1999; Miyake & Shah, 1999), produces appropriate responses while inhibiting inappropriate responses (Kane et al., 2008), and maintains memory activation through rehearsal. One component of the central executive is *endogenous attention*, which is the control of processing through voluntary effort. (*Exogenous attention* is the capture of processing by a highly salient stimulus; Miyake & Shah, 1999.)

Working memory also includes the phonological loop, which processes speech and linguistic information (Baddeley, 1986; Baddeley & Hitch, 1974; Logie, 1995). This component can be further subdivided into a phonological temporary store and an active rehearsal mechanism (Baddeley & Logie, 1999). Another component is the visuospatial sketchpad, which processes visual and spatial information. This component can be further subdivided into a visual cache (termed *visual short-term memory* in Fig. 9.4) and an active spatial-based rehearsal mechanism called the *inner scribe* (Baddeley & Logie, 1999; Logie, 1995). It is this visual short-term memory process, together with the central-executive, that may be enhanced by the use of interactive stereo displays.

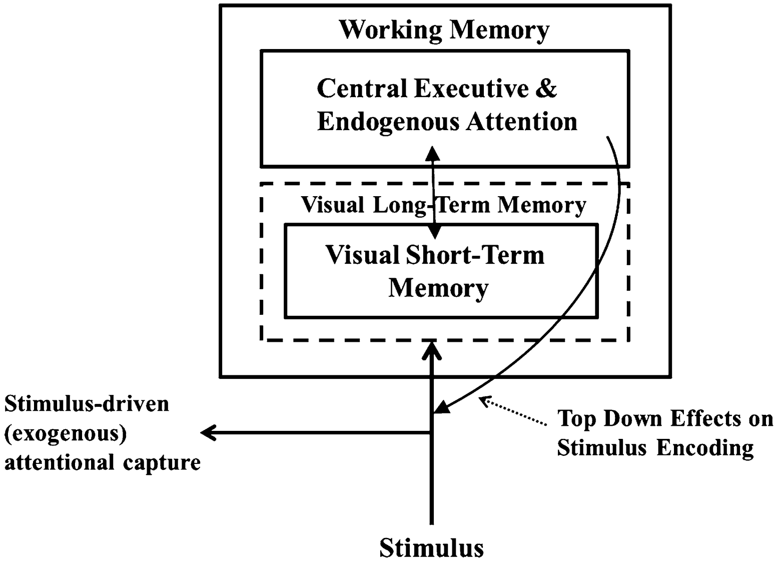


**Fig. 9.4** Diagram of the parts of working memory. Information initially flows from the bottom of the diagram up to the top. The central executive activates modality-specific representations from long-term memory, maintains and updates goals, plans actions, and inhibits competing responses. The central executive also allocates endogenous attention for performing multiple tasks. Exogenous attention is triggered by stimulation and can control access to working memory. One part of working memory is visual short-term memory, which is part of the visuospatial sketchpad (not shown). It is the processing by this visual short-term memory, together with the central-executive, that may be enhanced by the use of interactive stereo displays

It is possible that, by making spatial relationships more fully defined and perceived, interactive stereo displays could enhance the cognitive generation of mental models and lessen demands on working memory. That interactive stereo displays can facilitate cognitive processing is a hypothesis to be tested in future research.

Finally, recall the discussion of the dual-process model of reasoning and decision making outlined previously in Chap. 8 (see Fig. 8.5). In that chapter, it was stated that high-level cue conflict may be minimized or eliminated in part by the use of interactive stereo displays that provided congruent motion parallax and binocular parallax via head tracking and imagery updating. The congruent parallax information may permit the *intuitive* reasoning system, which is one of the two systems of the dual-process model, to make reasoned sense out of the incoming perceptual information. In the section just reviewed above in this chapter, it was stated that interactive stereo displays could lessen the demands placed on working memory in certain contexts, which is part of the *analytical* system of the dual-process model. Thus, interactive stereo displays seem to be important for both the intuitive and analytical cognitive systems involved in human reasoning and decision making.

## 9.4 Summary of High-Level Factors

High-level cue conflict refers to a potential perceptual conflict between the presence of binocular parallax and the absence of corresponding motion parallax in a stereo display (Merritt et al., 1991). High-level cue conflict may lead to discomfort in the viewing of immersive (simulated real-world) stereo displays because the intuitive reasoning system fails to make reasoned sense out of the conflicted perceptual information. The presence of motion parallax, together with congruent binocular parallax, in an interactive stereo display would minimize high-level cue conflict, benefit the intuitive system, and increase viewing comfort and the quality of the sense of depth and immersion by making spatial relationships more fully defined and perceived, as Merritt et al. (1991) have done. Interactive stereo displays may aid spatial reasoning because they may generate a sense of spatial relations derived from proprioception/kinesthesia. Spatial reasoning may involve the formation of mental models, which depends upon working memory, a part of the analytical reasoning system. Interactive stereo displays may enhance the user's mental models, and lessen demands placed on working memory. Interactive stereo displays thus seem to be important for both the intuitive and analytical cognitive systems involved in human reasoning and decision making.

We now turn to a set of recommendations for stereoscopic display design, which is the last chapter of this book.

# Part III
# Recommendations for Stereoscopic Display Design

# Chapter 10
# Recommendations for Stereoscopic Display Design

We have covered a lot of material in this book, so here is a summary:

1. *Interocular cross talk*. Limit interocular cross talk to a value less than 2 %. This is an issue of the physics of display design and its solution lies in the ability to keep separate the information delivered to the two eyes.
2. *Interocular differences in luminance and contrast*. Keep both interocular luminance differences and interocular contrast differences less than 25 %, ideally less than 5 % each.
3. *Accommodation–vergence mismatch*. View stereo displays from a distance of 1 m or greater if possible. Present the stereo depth of the fixated virtual object such that its displayed images fall within the depth of field of the human eye (by combining calculations for perceived depth with calculations for depth of field, equal to a 0.5 diopter tolerance for average display luminance).
4. *Brightness*. Brighter display luminance 'stops-down' pupil size, which in turn expands the depth of field, thus expanding the tolerance zone for comfortable 3D viewing (where accommodation/vergence mismatch would not interfere).
5. *Stereoanomaly*. In certain situations, there may be a need to screen for stereoanomaly and stereoblindness. To enhance the chances for veridical depth perception, imagery should be displayed under non-degraded conditions and binocular disparity bolstered with other congruent depth and distance cues.
6. *Spatio-temporal frequency effects*. Human sensitivity across a wide range of disparity magnitudes in crossed and uncrossed directions can be predicted from knowledge about the spatio-temporal luminance modulation of the displayed imagery. For displayed imagery with fine details, the total effective disparity range can be 80 arcmin, centered on fixation (horopter), and stereoacuity can be as low as 20 arcsec. For imagery with coarse details, the total effective disparity range can be 8 arcdeg (centered on fixation), and stereoacuity can be 5 arcmin (which can improve to 20 arcsec with transient stimulation). To prevent the appearance of disruptive peripheral flicker, the field of view should be limited with time-multiplexed displays or a high frame rate should be used.



R.E. Patterson, *Human Factors of Stereoscopic 3D Displays*,
DOI 10.1007/978-1-4471-6651-1_10

7. *Distance scaling of disparity*. Viewing distance is important for determining the amount of depth that is perceived in stereo displays. To perceive stereo depth, the visual system re-calibrates binocular disparity in accordance with changes in viewing distance. Changes in viewing distance affect the amount of perceived depth with stereo displays differently than perceived depth in the real-world, which may complicate the use of mixed-reality or augmented-reality displays.
8. *High-level cue conflict*. Depth and distances cues should be congruent in stereo display viewing, including binocular parallax, motion parallax, and perspective cues. Stereo displays can be divided into *analytical reasoning-inducing stereo displays* (alphanumeric symbology presented with disparity), and *intuitive reasoning-inducing stereo displays* (immersive geophysical stereo imagery). High-level cue conflict likely occurs primarily with the latter type of display, stereo displays that depict a simulation of the physical world without congruent motion parallax (e.g., 3D movies).

# References


Aglioti, S., DeSouza, J. F. X., & Goodale, M. A. (1995). Size-contrast illusions deceive the eye but not the hand. *Current Biology, 5*, 679–685.

Ames, A., Jr. (1925). The illusion of depth from single pictures. *Journal of the Optical Society of America, 10*, 137–148.

Anderson, J. R. (2010). *Cognitive psychology and its implications*. New York: Worth.

Andre, J. T., & Owens, D. A. (1999). Predicting optimal accommodative performance from measures of the dark focus of accommodation. *Human Factors, 41*, 139–145.

Arditi, A. (1986). Binocular vision. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and performance: Vol. 1. Sensory processes and perception* (pp. 23.1–23.40). New York: Wiley.

Backus, B. T., Fleet, D. J., Parker, A. J., & Heeger, D. J. (2001). Human cortical activity correlates with stereoscopic depth perception. *Journal of Neurophysiology, 86*, 2054–2068.

Baddeley, A. D. (1986). *Working memory*. New York: Oxford University Press.

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews. Neuroscience, 4*, 829–839.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York: Academic Press.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multi-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28–61). New York: Cambridge University Press.

Banks, M. S., Read, J. C. A., Allison, R. S., & Watt, S. (2012). Stereoscopy and the human visual system. *SMPTE Motion Imaging Journal, 121*, 24–43.

Bardy, B. G., Warren, W. H., Jr., & Kay, B. A. (1996). Motion parallax is used to control postural sway during walking. *Experimental Brain Research, 111*, 271–282.

Bardy, B. G., Warren, W. H., Jr., & Kay, B. A. (1999). The role of central and peripheral vision in postural control during walking. *Perception & Psychophysics, 61*, 1356–1368.

Biederman, I. (1972). Perceiving real-world scenes. *Science, 177*, 77–80.

Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 213–253). Hillsdale, NJ: Erlbaum.

Biederman, I., Glass, A. L., & Stacy, E. W. (1973). Searching for objects in real-world scenes. *Journal of Experimental Psychology, 97*, 22–27.

Blake, R. (1989). A neural theory of binocular rivalry. *Psychological Review, 96*, 145–167.

Blake, R. (2001). A primer on binocular rivalry, including current controversies. *Brain and Mind, 2*, 5–38.




R.E. Patterson, *Human Factors of Stereoscopic 3D Displays*,
DOI 10.1007/978-1-4471-6651-1

Blake, R., & Sekuler, R. (2005). *Perception*. New York: McGraw-Hill.
Blakemore, C. (1970). The range and scope of binocular depth discrimination in man. *Journal of Physiology, 211*, 599–622.
Boydstun, A., Rogers, J., Tripp, L., & Patterson, R. (2009). Stereo depth perception survives significant interocular luminance differences. *Journal of the Society for Information Display, 17*, 467–471.
Breese, B. B. (1899). On inhibition. *Psychological Monographs, 3*, 1–65.
Bridgeman, B., Gemmer, A., Forsman, T., & Huemer, V. (2000). Processing spatial information in the sensorimotor branch of the visual system. *Vision Research, 40*, 3539–3552.
Brooks, L. R. (1968). Spatial and verbal components of the act of recall. *Canadian Journal of Psychology, 22*, 349–368.
Brunye, T. T., & Taylor, H. A. (2008). Working memory in developing and applying mental models from spatial descriptions. *Journal of Memory and Language, 58*, 701–729.
Burgess, N. (2006). Spatial memory: How egocentric and allocentric combine. *Trends in Cognitive Sciences, 10*, 551–557.
Burkhalter, A., & Van Essen, D. C. (1986). Processing of color, form and disparity information in visual areas VP and V2 of ventral extrastriate cortex in the macaque monkey. *The Journal of Neuroscience, 6*, 2327–2351.
Burt, P., & Julesz, B. (1980). Modifications of the classical notion of Panum's fusional area. *Perception, 9*, 671–682.
Byrne, R. M. J., & Johnson-Laird, P. N. (2009). 'If' and the problems of conditional reasoning. *Trends in Cognitive Sciences, 13*, 282–287.
Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology, 197*, 551–566.
Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics, 27*, 183–228.
Conway, A. R. A., Jarrold, C., Kane, M. J., Miyake, A., & Towse, J. N. (2008). Variation in working memory: An introduction. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 3–20). New York: Oxford University Press.
Cormack, R. H., & Fox, R. (1985a). The computation of retinal disparity. *Perception & Psychophysics, 37*, 176–178.
Cormack, R. H., & Fox, R. (1985b). The computation of disparity and depth in stereograms. *Perception & Psychophysics, 38*, 375–377.
Cormak, R. H. (1984). Stereoscopic depth perception at far viewing distances. *Perception & Psychophysics, 35*, 423–428.
Cosmides, L., Tooby, J., Fiddick, L., & Bryant, G. A. (2005). Detecting cheaters. *Trends in Cognitive Sciences, 9*, 505–506.
Cowan, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.
Cowan, N. (1999). An embedded-process model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). New York: Cambridge University Press.
Cumming, B. G., & DeAngelis, G. C. (2001). The physiology of stereopsis. *Annual Review of Neuroscience, 24*, 203–238.
Cumming, B. G., & Parker, A. J. (1999). Binocular neurons in V1 of awake monkeys are selective for absolute, not relative, disparity. *Journal of Neuroscience, 19*, 5602–5618.
De Lange, H. (1952). Experiments on flicker and some calculations on an electrical analogue of the foveal systems. *Physica, 18*, 935–950.
De Lange, H. (1954). Relationship between critical flicker-frequency and a set of low-frequency characteristics of the eye. *Journal of the Optical Society of America, 44*, 380–389.
De Valois, R. L., & De Valois, K. K. (1988). *Spatial vision*. New York: Oxford University Press.
Diwadkar, V. A., & McNamara, T. P. (1997). Viewpoint dependence in scene recognition. *Psychological Science, 8*, 302–307.

Dobbins, A. C., Jeo, R. M., Fiser, J., & Allman, J. M. (1998). Distance modulation of neural activity in the visual cortex. *Science, 281*, 552–555.

Edgar, G. K., Pope, J. C., & Craig, I. (1993). Visual accommodation problems with head-up and helmet mounted displays. In C. T. Bartlett & M. D. Cowan (Eds.), *Proceedings of SPIE: Display systems: High-resolution and large screen displays and helmet, head-up, and head-down displays* (Vol. 1988, pp. 100–107). Bellingham, WA: SPIE.

Edwards, M., Pope, D. R., & Schor, C. M. (1999). Orientation tuning of the transient-stereopsis system. *Vision Research, 39*, 2717–2727.

Engber, D. (2009, April 2). The problem with 3D. *Slate Magazine*.

Enns, J. T., & Rensink, R. A. (1990). Influence of scene-based properties on visual search. *Science, 247*, 721–723.

Epstein, W. (1997). *Stability and constancy in visual perception: Mechanisms and processes*. New York: Wiley.

Evans, J. S. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Science, 17*, 454–459.

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment and social cognition. *Annual Review of Psychology, 59*, 255–278.

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8*, 223–241.

Falk, V., Mintz, D., Grunenfelder, J., Fann, J. I., & Burdon, T. A. (2001). Influence of three-dimensional vision on surgical telemanipulator performance. *Surgical Endoscopy, 15*, 1282–1288.

Farivar, R. (2009). Dorsal-ventral integration in object recognition. *Brain Research Reviews, 61*, 144–153.

Foley, J. M. (1980). Binocular distance perception. *Psychological Review, 87*, 411–434.

Foley, J. M., Ribeiro-Filho, N. P., & Da Silva, J. A. (2004). Visual perception of extent and the geometry of visual space. *Vision Research, 44*, 147–156.

Foley, J. M., & Richards, W. (1972). Effect of voluntary eye movement and convergence on the binocular appreciation of depth. *Perception & Psychophysics, 11*, 423–427.

Freeman, R. D., & Ohzawa, I. (1990). On the neurophysiological organization of binocular vision. *Vision Research, 30*, 1661–1676.

Gardner, H. (1993). *Frames of mind: The theory of MI*. New York: Basic Books. (Original work published 1983)

Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (2002). *Cognitive neuroscience: The biology of the mind*. New York: Norton.

Genovesio, A., & Ferraina, S. (2004). Integration of retinal disparity and fixation-distance related signals: Toward an egocentric coding of distance in the posterior parietal cortex of primates. *Journal of Neurophysiology, 91*, 2670–2684.

Georgieva, S., Peeters, R., Kolster, H., Todd, J. T., & Orban, G. A. (2009). The processing of three dimensional shape from disparity in the human brain. *Journal of Neuroscience, 29*(3), 727–742.

Geri, G., Pierce, B. P., & Patterson, R. (2008). Oculomotor contribution to the change in perceived speed with viewing distance. *Journal of the Optical Society of America A, 25*, 2851–2857.

Gernsbacher, M. A. (1985). Surface information loss in comprehension. *Cognitive Psychology, 17*, 324–363.

Getty, D. J., & Green, P. J. (2007). Clinical applications for stereoscopic 3-D displays. *Journal of the Society for Information Display, 15*, 377–384.

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

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

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

Gillam, B., Chambers, D., & Lawergren, B. (1988). The role of vertical disparity in the scaling of stereoscopic depth perception: An empirical and theoretical study. *Perception & Psychophysics, 44*, 473–483.

Gillam, B., & Lawergren, B. (1983). The induced effect, vertical disparity, and stereoscopic theory. *Perception & Psychophysics, 34*, 121–130.

Goodale, M. A., Milner, A. D., Jakobson, L. S., & Carey, D. P. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature, 349*, 154–156.

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science, 16*, 152–160.

Grossman, E. D., & Blake, R. (2001). Brain activity evoked by inverted and imagined biological motion. *Vision Research, 41*, 1475–1482.

Grossman, E. D., Donnelly, M., Price, P., Morgan, V., Pickens, D., Neighbor, G., et al. (2000). Brain areas involved in perception of biological motion. *Journal of Cognitive Neuroscience, 12*, 711–720.

Gugerty, L., deBoom, D., Jenkins, J. C., & Morley, R. (2000). Keeping north in mind: How navigators reason about cardinal directions. In *Proceedings of the Human Factors and Ergonomics Society 2000 Congress* (pp. 1148–1151). Santa Monica, CA: Human Factors and Ergonomics Society.

Hammond, K. R. (2007). *Beyond rationality: The search for wisdom in a troubled time*. New York: Oxford University Press.

Hammond, K. R., Hamm, R. M., Grassia, J., & Pearson, T. (1997). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. In W. M. Goldstein & R. M. Hogarth (Eds.), *Research on judgment and decision making: Currents, connections and controversies* (pp. 144–180). New York: Cambridge University Press.

Hanazato, A., Okui, M., & Yuyama, I. (2000). Subjective evaluation of cross talk disturbance in stereoscopic displays. In *Proceedings of the Society for Information Display 20th International Display Research Conference* (pp. 288–291).

Hasegawa, S., Hasegawa, A., Omori, M., Ishio, H., Takada, T., & Miyao, M. (2011). Stereoscopic vision induced by parallax images on HMD and its influence on visual functions (pp. 297–305). In *Virtual and Mixed Reality—New Trends:* Vol. 6773. *Lecture notes in computer science*.

Hegde, J., & Felleman, D. J. (2007). Reappraising the functional implications of the primate visual anatomical hierarchy. *The Neuroscientist, 13*, 416–421.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behaviour. *American Journal of Psychology, 57*, 243–259.

Hess, R. F., Liu, C. H., & Wang, Y.-Z. (2003). Differential binocular input and local stereopsis. *Vision Research, 43*, 2303–2313.

Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergence—Accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision, 8*, 1–30.

Hoffman, D. M., Karasev, V. I., & Banks, M. S. (2011). Temporal presentation protocols: Flicker visibility, perceived motion, and perceived depth. *Journal of the Society for Information Display, 19*, 255–281.

Hogarth, R. (2001). *Educating intuition*. Chicago: University of Chicago Press.

Hogarth, R. (2002). *Deciding analytically or trusting your intuition? The advantages and disadvantages of analytic and intuitive thought* (UPF Economics Business Working Paper No. 654).

Hollins, M., & Leung, E. (1978). The influence of color on binocular rivalry. In J. C. Armington, J. Krausfopf, & B. R. Wotten (Eds.), *Visual psychophysics and physiology* (pp. 181–190). New York: Academic Press.

Howard, I. P. (2002). *Seeing in depth: Vol. 1. Basic mechanisms*. New York: Porteous.

Howard, I. P., & Rogers, B. J. (1995). *Binocular vision and stereopsis*. Oxford, England: Oxford University Press.

Hutchens, E. L., Hollan, J. D., & Norman, D. A. (1985). Direct manipulation interfaces. *Human–Computer Interaction, 1*, 311–338.

Johnson, C. A. (1976). Effects of luminance and stimulus distance on accommodation and visual resolution. *Journal of the Optical Society of America, 66*, 138–142.

Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences, 107*, 18243–18250.

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, inference, and pragmatics. *Psychological Review, 109*, 646–678.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). New York: Cambridge University Press.

Kandel, E. R., Schwartz, J., & Jessel, T. (2000). *Principles of neuroscience*. New York: McGraw Hill.

Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2008). Variation in working memory capacity as variation in executive attention and control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 21–48). New York: Oxford University Press.

Keehner, M., Khooshabeh, P., & Hegarty, M. (2008). Interactive visualizations and individual differences among users. In F. Dong, G. Ghinea, & S. Y. Chen (Eds.), *User centered design for medical visualization* (pp. 1–24). Hershey, PA: Idea Group.

Kelly, D. H. (1971). Theory of flicker and transient responses. II. Counterphase gratings. *Journal of the Optical Society of America A, 61*, 632–640.

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science, 18*, 513–549.

Klatzky, R. L. (1998). Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In *Spatial Cognition: Vol. 1404. Lecture notes in computer science* (pp. 1–17). Berlin: Springer.

Klein, G., & Hoffman, R. R. (2008). Macrocognition, mental models, and cognitive task analysis methodology. In J. M. Schraagen, L. G. Militello, T. Ormerod, & R. Lipshitz (Eds.), *Naturalistic decision making and macrocognition*. Hampshire, England: Ashgate.

Kooi, F. (2011). A display with two depth layers: Attentional segregation and declutter. In C. Roda (Ed.), *Human attention in digital environments* (pp. 245–258). Cambridge, England: Cambridge University Press.

Kooi, F. L., & Toet, A. (2004). Visual comfort of binocular and 3D displays. *Displays, 25*, 99–108.

Lack, L. (1969). The effect of practice on binocular rivalry control. *Perception & Psychophysics, 6*, 397–400.

Lambooij, M., Fortuin, M., IJsselsteijn, W. A., & Heynderickx, I. (2009). Measuring visual discomfort associated with 3D displays. In *Proceedings of SPIE 7237, stereoscopic displays and applications, 72370K*.

Lambooij, M., IJsselsteijn, W., Fortuin, M., & Heynderickx, I. (2009). Visual discomfort and visual fatigue of stereoscopic displays: A review. *Journal of Imaging Science and Technology, 53*, 1–14.

Leibowitz, H. W., & Owens, D. A. (1975a). Anomalous myopias and the intermediate dark focus of accommodation. *Science, 189*, 646–648.

Leibowitz, H. W., & Owens, D. A. (1975b). Night myopia and the intermediate dark focus of accommodation. *Journal of the Optical Society of America, 65*, 1121–1128.

Leibowitz, H. W., & Owens, D. A. (1978). New evidence for the intermediate position of relaxed accommodation. *Documenta Ophthalmologica, 46*, 133–147.

LeVay, S., & Voight, T. (1988). Ocular dominance and disparity coding in cat visual cortex. *Visual Neuroscience, 1*, 395–414.

Levelt, W. J. M. (1965). *On binocular rivalry*. Soesterberg, The Netherlands: Institute for Perception, RVO-TNO.

Li, L., & Warren, W. H., Jr. (2000). Perception of heading during rotation: Sufficiency of dense motion parallax and reference objects. *Vision Research, 40*, 3873–3894.

Li, L., & Warren, W. H., Jr. (2002). Retinal flow is sufficient for steering during observer rotation. *Psychological Science, 13*, 485–491.

Likova, L. T., & Tyler, C. W. (2007). Stereomotion processing in the human occipital cortex. *NeuroImage, 38*, 293–305.

Lit, A. (1949). The magnitude of the Pulfrich stereo-phenomenon as a function of binocular differences of intensity at various levels of illumination. *American Journal of Psychology, 62*, 159–181.

Livingstone, M. S., & Hubel, D. H. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science, 240*, 740–749.

Lloyd, C. J., & Nigus, S. G. (2012). Effects of stereopsis, collimation, and head tracking on air refueling boom operator performance. In *Proceedings of the Image Society Annual Conference,* Scottsdale, AZ.

Loewenfeld, I. E. (1993). *The pupil: Anatomy, physiology and clinical applications*. Ames, IA: Iowa State University Press.

Logie, R. H. (1995). *Visuo-spatial working memory*. Hove, England: Erlbaum.

Longuet-Higgins, H. C., & Prazdny, K. (1980). The interpretation of a moving retinal image. *Proceedings of the Royal Society of London: Series B, 208*, 385–397.

Lueder, E. (2012). *3D displays (Wiley series in display technology)*. Chichester, England: Wiley.

Luursema, J.-M., Verwey, W. B., Kommers, P. A. M., & Annema, J.-H. (2008). The role of stereopsis in virtual anatomical learning. *Interacting with Computers, 20*, 455–460.

Mandler, J. M., & Ritchey, G. H. (1977). Long-term memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 386–396.

McIntire, J. P., Havig, P. R., & Geiselman, E. E. (2012). What is 3D good for? A review of human performance on stereoscopic 3D displays. In P. L. Marasco, P. R. Havig, D. D. Desjardins, K. R. Sarma (Eds.), *Head- and helmet-mounted displays XVII and display technologies and applications for defense, security, and avionics VI. Proceedings of SPIE, 8383*, *83830X-1-13*.

McIntire, J. P., Havig, P. R., & Geiselman, E. E. (2014). Stereoscopic 3D displays and human performance: A comprehensive review. *Displays, 35*, 18–26.

McKee, S. P., & Welch, L. (1989). Is there a constancy for velocity? *Vision Research, 29*, 553–561.

Merritt, J. (1987). Visual motor realism in 3D teleoperator display systems. True three-dimensional imaging techniques and display technologies (pp. 88–93). In *Proceedings of the Society of Photo-Optical Instrumentation Engineers*, *761*, Bellingham, WA.

Merritt, J. (1989). Virtual window viewing geometry. In P. S. Schenker (Ed.), Sensor fusion: Spatial reasoning and scene interpretation. *Proceedings of the Society of Photo-Optical Instrumentation Engineers, 1003*, Bellingham, WA.

Merritt, J. O. (2011). Performance gains when using 3D displays vs. 2D displays. *SID Symposium Digest of Technical Papers, 42*, 987–990.

Merritt, J. O., Cole, R. E., & Ikehara, C. (1991). A rapid-sequential-positioning task for evaluating motion parallax and stereoscopic 3D cues in teleoperator displays. In *Proceedings of the IEEE conference on systems, man, and cybernetics*, October 13–16, University of Virginia, Charlottesville.

Michotte, A. (1962). *The perception of causality*. Andover, MA: Methuen.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. Oxford, England: Oxford University Press.

Miyake, A., & Shah, P. (1999). Emerging consensus, unresolved issues, future directions. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 442–482). New York: Cambridge University Press.

Miyao, M., Hasegawa, A., Hasegawa, S., Shiomi, T., Hori, H., & Takada, H. (2012). 3D vision gaze with simultaneous measurements of accommodation and convergence among young and middle-aged subjects. In V. G. Duffy (Ed.), *Advances in human aspects of healthcare* (pp. 509–514). New York: CRC Press.

Moyer, R. S. (1973). Comparing objects in memory: Evidence suggesting an internal psychophysics. *Perception & Psychophysics, 13*, 180–184.

Muhlbach, L., Bocker, M., & Prussog, A. (1995). Telepresence in videocommunications: A study of stereoscopy and individual eye contact. *Human Factors, 37*, 290–305.

Nawrot, M. (2003). Eye movements provide the extra-retinal signal required for the perception of depth from motion parallax. *Vision Research, 43*, 1553–1562.

Nawrot, M., & Joyce, L. (2006). The pursuit theory of motion parallax. *Vision Research, 46*, 4709–4725.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and selfregulation* (Vol. 4, pp. 1–18). New York: Plenum Press.
O'Brien, D. (2009). Human reasoning requires mental logic. *Behavioral and Brain Sciences, 32*, 96–97.
Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences, 5*, 349–357.
Ogle, K. N. (1964). *Binocular vision*. New York: Hafner.
Ogle, K. N., & Schwartz, J. T. (1959). Depth of focus of the human eye. *Journal of the Optical Society of America, 49*, 273–280.
Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research: Visual Perception, 155*, 23–36.
Ono, H., & Comerford, T. (1977). Stereoscopic depth constancy. In W. Epstein (Ed.), *Stability and constancy in visual perception: Mechanisms and processes* (pp. 91–128). New York: Wiley.
Owens, D. A. (1979). The Mandelbaum effect: Evidence for an accommodative bias toward intermediate viewing distances. *Journal of the Optical Society of America, 69*, 646–652.
Owens, D. A., & Leibowitz, H. W. (1976). Oculomotor adjustments in darkness and the specific distance tendency. *Perception & Psychophysics, 20*, 2–9.
Owens, D. A., & Leibowitz, H. W. (1980). Accommodation, convergence, and distance perception in low illumination. *American Journal of Optometry and Physiological Optics, 57*, 540–550.
Patten, M. L., & Murphy, A. P. (2012). Relative disparity processing in the dorsal visual pathway. *The Journal of Neuroscience, 32*, 5353–5355.
Patterson, R. (1990). Spatio-temporal properties of stereoacuity. *Optometry and Vision Science, 67*, 123–125.
Patterson, R. (2007). Human factors of 3D displays. *Journal of the Society for Information Display, 15*, 861–871.
Patterson, R. (2009). Human factors of stereo displays: An update. *Journal of the Society for Information Display, 17*, 987–996.
Patterson, R. (2012). Cognitive engineering, cognitive augmentation, and information display. *Journal of the Society for Information Display, 20*, 208–213.
Patterson, R. (2013). Interactive stereoscopic displays. *Journal of the Society for Information Display, 21*, 142–150.
Patterson, R., Cayko, R., Short, L., Flanagan, R., Moe, L., Taylor, E., et al. (1995). Temporal integration differences between crossed and uncrossed stereoscopic mechanisms. *Perception & Psychophysics, 57*, 891–897.
Patterson, R., & Fox, R. (1984). The effect of testing method on stereoanomaly. *Vision Research, 24*, 403–410.
Patterson, R., Geri, G. A., Dyre, B. P., Pierce, B. J., Akhtar, S. C., Covas, C. M., et al. (2006). Active heading control in simulated flight based on vertically extended contours. *Perception & Psychophysics, 68*, 593–600.
Patterson, R., & Martin, W. L. (1992). Human stereopsis. *Human Factors, 34*, 669–692.
Patterson, R., Moe, L., & Hewitt, T. (1992). Factors that affect depth perception in stereoscopic displays. *Human Factors, 34*, 655–667.
Patterson, R., Pierce, B. J., Bell, H., Andrews, D., & Winterbottom, M. (2009). Training robust decision making in immersive environments. *Journal of Cognitive Engineering and Decision Making, 3*, 331–361.
Patterson, R., Ricker, C., McGary, J., & Rose, D. (1992). Properties of cyclopean motion perception. *Vision Research, 32*, 149–156.
Patterson, R., & Silzars, A. (2009). Immersive stereo displays, intuitive reasoning, and cognitive engineering. *Journal of the Society for Information Display, 17*, 443–448.
Patterson, R., Winterbottom, M., Pierce, B., & Fox, R. (2007). Binocular rivalry and head-mounted displays. *Human Factors, 49*, 1083–1096.
Patterson, R., Winterbottom, M., & Pierce, B. (2006). Perceptual issues in the use of head-mounted visual displays. *Human Factors, 48*, 555–564.

Percival, A. S. (1892). The relation of convergence to accommodation and its practical bearing. *Ophthalmological Review, 11*, 313–328.
Peuskens, H., Sunaert, S., Dupont, P., Van Hecke, P., & Orban, G. A. (2001). Human brain regions involved in heading estimation. *Journal of Neuroscience, 21*, 2451–2461.
Philbeck, J. W., Klatzky, R. L., Behrmann, M., Loomis, J. M., & Goodridge, J. (2001). Active control of locomotion facilitates nonvisual navigation. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 141–153.
Poggio, G. F. (1995). Mechanisms of stereopsis in monkey visual cortex. *Cerebral Cortex, 5*, 193–204.
Poggio, G. F., Motter, B. C., Squatrito, S., & Trotter, Y. (1985). Responses of neurons in visual cortex (V1 and V2) of the alert macaque to dynamic random dot stereograms. *Vision Research, 25*, 397–406.
Pulfrich, C. (1922). Die Stereoskopie im Dienste der isochromen und heterochromen Photometrie. *Die Naturwissenschaften, 10*, 553–761.
Ragan, E. D., Kopper, R., Schuchardt, P., & Bowman, D. A. (2013). Studying the effects of stereo, head tracking, and field of regard on a small-scale spatial judgment task. *IEEE Transactions on Visualization and Computer Graphics, 19*(5), 886–896.
Rash, C. E., Mozo, B. T., McEntire, B. J., & Licina, J. R. (1996). *RAH-66 Comanche health hazard and performance issues for the helmet integrated display and sighting system* (Tech. Rep. 97-1). Fort Rucker, AL: U.S. Army Aeromedical Research Laboratory.
Reeves, P. (1920). The response of the average pupil to various intensities of light. *Journal of the Optical Society of America, 42*, 35–43.
Renz, J., & Nebel, B. (2007). Qualitative spatial reasoning using constraint calculi. In M. Aiello, I. Pratt-Hartmann, & J. van Benthem (Eds.), *Handbook of spatial logics* (pp. 1–61). New York: Springer (Ch. 1).
Richards, W. (1970). Stereopsis and stereoblindness. *Experimental Brain Research, 10*, 380–388.
Richards, W. (1971). Stereoanomalous depth perception. *Journal of the Optical Society of America, 61*, 410–417.
Richards, W. (2009). Configuration stereopsis: A new look at the depth-disparity relation. *Spatial Vision, 22*, 91–103.
Rieger, J. H., & Lawton, D. T. (1985). Processing differential image motion. *Journal of the Optical Society of America A, 2*, 354–360.
Ritter, M. (1977). Effect of disparity and viewing distance on perceived depth. *Perception & Psychophysics, 22*, 400–407.
Rogers, B. J., & Bradshaw, M. F. (1993). Vertical disparities, differential perspective and binocular stereopsis. *Nature, 361*, 253–255.
Roland, P. E., & Friberg, L. (1985). Localization of cortical areas activated by thinking. *Journal of Neurophysiology, 53*, 1219–1243.
Rowlands, J. J., & Killian, J. R. (1937). Seeing solid: The third dimension at work and play. *The Technology Review, 39*, 191–210.
Royden, C. S., Crowell, J. A., & Banks, M. S. (1994). Estimating heading during eye movements. *Vision Research, 34*, 3197–3214.
Schall, J., Nawrot, M., Blake, R., & Yu, K. (1993). Visual guided attention is neutralized when informative cues are visible but unperceived. *Vision Research, 33*, 2057–2064.
Schiller, P. H., Logothetis, N. K., & Charles, E. R. (1990). Role of the color opponent and broad-band channels in vision. *Visual Neuroscience, 5*, 321–346.
Schlosberg, H. (1941). Stereoscopic depth from single pictures. *The American Journal of Psychology, 54*, 601–605.
Schor, C. M., & Wood, I. (1983). Disparity range for local stereopsis as a function of luminance spatial frequency. *Vision Research, 23*, 1649–1654.
Schor, C. M., Wood, I., & Ogawa, J. (1984). Spatial tuning of static and dynamic local stereopsis. *Vision Research, 24*, 573–578.
Schwartz, A. H. (1971). Stereoscopic perception with single pictures. *Optical Spectra, 9*, 25–27.
Servos, P., Goodale, M. A., & Jakobson, L. S. (1992). The role of binocular vision in prehension: A kinematic analysis. *Vision Research, 32*, 1513–1521.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701–703.

Shibata, T., Kim, J., Hoffman, D. M., & Banks, M. S. (2011). The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision, 11*, 1–29.

Shiomi, T., Uemoto, K., Kojima, T., Sano, S., Ishio, H., Takada, H., et al. (2013). Simultaneous measurement of lens accommodation and convergence in natural and artificial 3D vision. *Journal of the Society for Information Display, 21*, 120–128.

Shipley, T., & Rawlings, S. C. (1970). The nonius horopter: I. History and theory. *Vision Research, 10*, 1225–1262.

Siegel, M., Tobinaga, Y., & Akiya, T. (1999). Kinder Gentler Stereo. In *Proceedings of SPIE: Stereoscopic displays and virtual reality systems VI* (Vol. 3639 A-03, pp. 1–10).

Simpson, W. A. (1993). Optic flow and depth perception. *Spatial Vision, 7*, 35–75.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3–22.

Sollenberger, R. L., & Milgram, P. (1993). Effects of stereoscopic and rotational displays in a three-dimensional path-tracing task. *Human Factors, 35*, 483–499.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences, 23*, 645–665.

Stanovich, K. E., & West, R. F. (2002). Individual differences in reasoning: Implications for the rationality debate. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases* (pp. 421–440). Cambridge, England: Cambridge University Press.

Stuart, G. W., Jennings, S. A., Kalich, M. E., Rash, C. E., Harding, T. H., & Craig, G. L. (2009). Flight performance using a hyperstereo helmet-mounted display: Adaptation to hyperstereopsis. In *Proceedings of the Society for Photo-Electronic Instrumentation Engineering 7326, Head- and Helmet-Mounted Displays XIV: Design and applications, 732606* (pp. 72400A–72400A-15). Bellingham, WA: International Society for Optics and Photonics.

T'so, D. Y., & Roe, A. W. (1995). Functional compartments in visual cortex: Segregation and interaction. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 325–337). Cambridge, MA: MIT Press.

Taylor, H. A., & Tversky, B. (1992). Spatial mental models derived from survey and route descriptions. *Journal of Memory and Language, 31*, 261–292.

Thompson, W. L., & Kosslyn, S. M. (2000). Neural systems activated during visual mental imagery: A review and meta-analyses. In A. W. Toga & J. C. Mazziotta (Eds.), *Brain mapping II: The systems* (pp. 535–560). San Diego, CA: Academic Press.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*, 520–522.

Toates, F. M. (1974). Vergence eye movements. *Documenta Ophthalmologica, 37*, 153–214.

Toats, F. M. (1972). Accommodation function of the human eye. *Physiological Reviews, 52*, 828–863.

Trotter, Y., Celebrini, S., Stricanne, B., Thorpe, S., & Imbert, M. (1996). Neural processing of stereopsis as a function of viewing distance in primate visual cortical area V1. *Journal of Neurophysiology, 76*, 2872–2885.

Tsao, D. Y., Vanduffel, W., Sasaki, Y., Fize, D., Knutsen, T. A., Mandeville, J. B., et al. (2003). Stereopsis activates V3A and caudal intraparietal areas in macaques and humans. *Neuron, 39*, 555–568.

Tsou, B. H., & Shenker, M. (2000). Visual factors associated with headmounted displays. In M. Bass (Ed.), *Handbook of optics: Vol. III. Visual optics: Issues pertinent to the optical designer* (2nd ed., pp. 18.1–18.15). New York: McGraw-Hill.

Tyler, C. W. (1973). Stereoscopic vision: Cortical limitations and a disparity scaling effect. *Science, 181*, 276–278.

Tyler, C. W. (1983). Sensory processing of binocular disparity. In C. M. Schor & K. J. Ciuffreda (Eds.), *Vergence eye movements: Basic and clinical aspects* (pp. 199–295). Boston: Butterworths.

Tyler, C. W., & Scott, A. B. (1979). Binocular vision. In R. E. Records (Ed.), *Physiology of the human eye and visual system* (pp. 643–671). New York: Harper & Row.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, England: MIT Press.

Van Beurden, M. H. P. H., van Hoey, G., Hatzakis, H., & IJsselsteijn, W. A. (2009). Stereoscopic displays in medical domains: A review of perception and performance effects. In *Human vision and electronic imaging XIV, Proceedings of the SPIE* (pp. 72400A–72400A-15). Bellingham, WA: International Society for Optics and Photonics.

van den Enden, A., & Spekreijse, H. (1989). Binocular depth reversals despite familiarity cues. *Science, 244*, 959–961.

van Ee, R., & Richards, W. (2002). A planar and a volumetric test for stereoanomaly. *Perception, 31*, 51–64.

Van Essen, D. C., & DeYoe, E. A. (1995). Concurrent processing in the primate visual cortex. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 383–400). Cambridge, MA: MIT Press.

Vanduffel, W., Fize, D., Peuskens, H., Denys, K., Sunaert, S., Todd, J. T., et al. (2002). Extracting 3D from motion: Differences in human and monkey intraparietal cortex. *Science, 298*, 413–415.

Velger, M. (1998). *Helmet-mounted displays and sights*. Boston: Artech House.

von Helmholtz, H. (1866). Concerning the perceptions in general. In *Treatise on physiological optics* (Vol. III, 3rd ed.) (J. P. C. Southall, Trans., 1925, Optical Society of America, Section 26; reprinted. New York: Dover, 1962).

von Hofsten, C. (1976). The role of convergence in visual space perception. *Vision Research, 16*, 193–198.

Votanopoulos, K., Brunicardi, F. C., Thornby, J., & Bellows, C. F. (2008). Impact of three-dimensional vision in laparoscopic training. *World Journal of Surgery, 32*, 110–118.

Wallach, H. (1939). On constancy of visual speed. *Psychological Review, 46*, 541–552.

Wallach, H., & Zuckerman, C. (1963). The constancy of stereoscopic depth. *American Journal of Psychology, 76*, 404–410.

Wang, B., & Ciuffreda, K. J. (2006). Depth of focus of the human eye: Theory and clinical applications. *Survey of Ophthalmology, 51*, 75–85.

Wang, L., Teunissen, K., Tu, Y., Chen, L., Zhang, P., Zhang, T., et al. (2011). Crosstalk evaluation in stereoscopic displays. *Journal of Display Technology, 7*, 208–214.

Wang, L., Tu, Y., Chen, L., Zhang, P., Zhang, T., Teunissen, K., et al. (2012). Effect of display technology on the crosstalk perception in stereoscopic video content. *IEEE Transactions on Circuits and Systems for Video Technology, 22*, 1257–1265.

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

Ware, C., & Mitchell, P. (2008). Visualizing graphs in three dimensions. *ACM Transactions on Applied Perception, 5*, 2.1–2.15.

Warren, W. H., Jr. (1998). The state of flow. In T. Watanabe (Ed.), *Highlevel motion processing: Computational, neurobiological, and psychological perspectives* (pp. 323–341). Cambridge, MA: MIT Press.

Weisstein, N., & Harris, C. S. (1974). Visual detection of line segments: An object-superiority effect. *Science, 186*, 752–755.

Weisstein, N., Williams, M. C., & Harris, C. S. (1982). Depth, connectedness, and structural relevance in the object-superiority effect: Line segments are harder to see in flatter patterns. *Perception, 11*(1), 5–17.

Westheimer, G. (1986). Spatial interaction in the domain of disparity signals in human stereoscopic vision. *Journal of Physiology, 370*, 619–629.

Westheimer, G., & Levi, D. M. (1987). Depth attraction and repulsion of disparate foveal stimuli. *Vision Research, 27*, 1361–1368.

Wetzel, P. A., Pierce, B. J., & Geri, G. A. (1996). Viewing distance and the perceived size of simulator imagery. *Journal of the Society for Information Display, 4*, 247–253.

Wheatstone, C. (1838). Contributions to the physiology of vision: 1. On some remarkable and hitherto unobserved phenomena of binocular vision. *Philosophical Transactions of the Royal Society London, 128*, 371–374.

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2013). *Engineering psychology and human performance* (4th ed., pp. 119–120). New York: Pearson.

Wickens, C. D., Todd, S., & Seidler, K. (1989). *Three-dimensional displays: Perception, implementation, applications* (Tech. Rep. CSERIAC SOAR-89-01). Wright-Patterson Air Force Base, OH: Armstrong Aerospace Medical Research Laboratory.

Wilcox, L. M., & Hess, R. F. (1995). Dmax for stereopsis depends on size, not spatial frequency content. *Vision Research, 35*, 1061–1069.

Woods, A. J. (2010). *Understanding crosstalk in stereoscopic displays*. Keynote presentation at three-dimensional systems and applications conference, Tokyo, Japan, May 19–21.

Yabuta, N. H., Sawatari, A., & Callaway, E. M. (2001). Two functional channels from primary visual cortex to dorsal visual cortical areas. *Science, 292*, 297–300.

Yeh, Y., & Silverstein, L. (1990). Limits of fusion and depth judgments in stereoscopic color displays. *Human Factors, 32*, 45–54.

Zohary, E., & Sittig, A. (1993). Mechanisms of velocity constancy. *Vision Research, 33*, 2467–2478.

# Index





R.E. Patterson, *Human Factors of Stereoscopic 3D Displays*,
DOI 10.1007/978-1-4471-6651-1