The interaction of eye movements and retinal signals during the perception of 3-D motion direction

Julie M. Harris
School of Psychology, University of St. Andrews, St. Andrews, Scotland, UK

When an object is tracked with the eyes, veridical perception of the motion of that object and other objects requires the brain to take account of and compensate for the eye movement. Here, I explore the effects of version and vergence eye movements on three-dimensional (3-D) motion perception. After demonstrating that eye movement compensation can be poor for detecting small objects moving in depth, I develop two models for how eye movement and visual signals may interact during the perception of 3-D motion direction. The first model assumes that the visual system is aiming to form an explicit representation of 3-D motion. From the results of a second experiment, on 3-D motion direction judgements, I show that this model could only hold with almost perfect 3-D motion compensation, contradicting the results from the first experiment. A second model assumes a much simpler strategy for estimating 3-D motion direction, based on recent experimental work. It predicts that compensation for vergence is not required because the Z-component of 3-D motion is not needed for direction judgements, consistent with the experimental results. This suggests that, for 3-D motion direction discrimination and angle judgements, extraretinal signals from vergence are not used.

Keywords: binocular vision, motion-in-depth, eye movements, trajectory perception, 3-D motion, version, vergence

Introduction

For motion perception to be accurate during eye movements, the brain must use extraretinal signals, to obtain information about what the eyes are doing, as well as use signals indicating what motion occurs on the retina. For example, when the eyes pursue a moving object accurately, there is no retinal motion signal. Yet, objects do appear to move during pursuit, albeit more slowly than when the eyes fixate a stationary target (the Aubert–Fleischl phenomenon, e.g., see Freeman, 2001). This and related phenomena can be understood if it is assumed that both retinal and extraretinal representations of motion are used by the brain but that there are errors in one or the other or both (Freeman, 2001; Freeman & Banks, 1998).

Many models have been developed to account for how and when the visual system compensates for the retinal motion created by congruent version eye movements. For example, the speed transducer model obtains motion with respect to the head by summing retinal and pursuit velocity and can account for illusions of speed and motion direction during pursuit eye movements (Freeman, 2001). Mislocalization of flashed signals during an eye movement can be explained in terms of time delays between eye and retinal signals (e.g., Schlag & Schlag-Rey, 2002).

Because compensation for pursuit eye movements is not perfect, pursuit of one target can affect the perceived motion of other targets. When the perceived direction of motion of a vertically moving spot is measured while observers pursue a horizontally moving spot, biases in perceived direction are found (Becklen, Wallach, & Nitzberg, 1984; Festinger, Sedgwick, & Holtzman, 1976; Souman, Hooge, & Wertheim, 2005; Swanston & Wade, 1988). Interestingly, the precision of direction perception is similar whether the eyes move or not (Krukowski, Pirog, Beutter, Brooks, & Stone, 2003).

The literature on how eye movements bias perception has focused almost exclusively on version eye movements. These are yoked, conjugate movements of both eyes, usually to follow a target moving in the frontoparallel plane (I will refer to horizontal motion in this plane as X-direction motion). When objects move three-dimensionally, with a component of their motion toward or away from the observer (termed ‘Z-direction motion here), the situation becomes more complicated. The Z-component of three-dimensional (3-D) motion results in almost equal and opposite motions in the two eyes. These disconjugate eye movements are termed vergence eye movements. Because the neural control of version and vergence are quite different (e.g., Collewijn, Erkelens, & Steinman, 1997; Semmlow, Yuan, & Alvarez, 1998), the literature described above, on the effects on perception during version eye movements, gives one little to go on to predict what happens when eyes undergo vergence movements to follow targets moving in depth.

Different scaling factors apply to convert retinal motion into the Z- or X-components of motion in the world, making matters more complicated. Retinal motion must be scaled by viewing distance, D, to obtain X-motion in world units (e.g., meters per second). However, Z-motion in world units is obtained from the difference between retinal motions, scaled by $D^2$ (e.g., Regan, 1993; see Equation 7). Correctly perceiving 3-D motions, containing both X- and Z-motion components, therefore requires scaling for viewing distance, whether the eyes move or...
not. Thus, when using perceived motion to probe how extraretinal and retinal information are combined, it should be remembered that perceived motion is subject to scaling (with different scaling factors for lateral and depth components), probably after retinal and extraretinal information have been combined.

One way for the brain to obtain an estimate of 3-D motion is to obtain independent measures of the X-component and the Z-component. Cumming & Parker (1994) suggested how the Z-component may be obtained, and Beverley & Regan (1975) demonstrated that 3-D motion trajectories can be precisely discriminated. There is evidence, from motion threshold studies, that there may be two independent mechanisms for 3-D motion detection (Sumnall & Harris, 2002), although the Z-component mechanism appears to be much less sensitive (Harris, McKee, & Watamaniuk, 1998; Sumnall & Harris, 2000; Tyler, 1971). If 3-D motion perception relied on a combination of independently obtained X-components and Z-components of motion, one might expect perception of 3-D motion direction to be biased during eye movements. It is known that the X-component is likely to be slightly biased during X-direction pursuit due to incomplete eye movement compensation. No studies have previously directly measured how the Z-component might be biased due to vergence. However, there are suggestions that perception is likely to be biased. When vergence is the only cue to the presence of Z-motion, Z-motion is not perceived over a range of speeds, distances, and stimulus sizes (Brenner, Van Den Berg, & Van Damme, 1996; Erkelens & Collewijn, 1985a; Regan, Erkelens, & Collewijn, 1986). This suggests that the extraretinal signal for motion-in-depth associated with vergence eye movements is weak if not absent, but note that there is evidence for the use of extraretinal vergence signals for other tasks. For example, changes in vergence can be used to partially scale disparity for perceived curvature (Rogers & Bradshaw, 1995; Watt, Akeley, Ernst, & Banks, 2005). Vergence is also thought to be involved in the scaling of object size: For example, afterimages can change size when vergence is changed (Mon-Williams, Tresilian, Plooy, Wann, & Broerse, 1997). However, some of the studies suggesting poor extraretinal information from vergence used stimuli containing conflicting cues. In particular, some used large stimuli that changed their disparity while their size was held constant; hence, there were cues to change in depth from disparity, alongside cues to no change in depth from constant size. Here, I used small, impoverished stimuli that reduce the salience of the no-motion size cue (Gray & Regan, 1999).

In the first experiment, I estimated the amount of compensation for vergence eye movements that occurs when detecting whether a small target moved toward or away from the observer. This was done so that I could predict how observers might behave in the second experiment, where I measured the effects of vergence and version eye movements during judgements of 3-D motion direction. Anticipating the results, it was found that compensation was poor, or very poor, predicting that the Z-component of perceived motion should be dominated by the relative retinal signal. With very little compensation for vergence, I predicted, by developing a fairly standard model of eye movement compensation (Model 1), that the judgement of 3-D motion direction should be substantially biased during vergence eye movements.

However, it has recently been suggested that the brain may not make an explicit 3-D representation of motion when estimating 3-D motion direction. Instead, the visual system may use the visual direction between the fixation location and the end point of the object motion, along with some estimate of distance to the end point, to obtain an often incorrect but consistent estimate of direction (Harris & Drga, 2005). I also describe a model (Model 2) that follows from this work and that predicts no bias during vergence, even with poor eye movement compensation. Finally, the second experiment was designed to distinguish between the two models and was consistent with Model 2.

Part I: Comparing version and vergence gains

In this section, I measured the extent to which eye movements are compensated for during vergence and compared this with what is known from the literature about compensation for both vergence and version eye movements.

Experiment 1: Compensation for vergence

I start with a simple model that follows from previous literature (e.g., see Freeman, 2001), in that when the eyes move to follow a target, the estimated object motion, O, is obtained from a combination of the estimated eye motion, E, and the estimated retinal motion, R, so that O = E + R. The estimated eye motion and retinal motions are related to their actual counterparts by a gain factor, so that E = eE and R = rR. The gain is the slope of the relationship between input (eye motion or retinal motion) and output (estimated eye motion or estimated retinal motion) and represents the extent to which the signal is used by the visual system. For example, a gain of 1 means that the signal is used veridically, and a gain of 0 would mean that it was not used at all. Later, I will develop this model in more detail, considering both X- and Z-components of motion (Model 1). For now, let:

\[ O' = (r_oR_z + e_zE_z), \]
where the $z$ subscript denotes that I am considering motions and gains for motion in the $Z$-direction only.

I compared two experimental conditions. In the first, I measured thresholds for discriminating whether the direction of motion was toward or away from the observer with only a single moving target present. I assume that the observer requires a fixed extent of perceived motion to detect the perceived motion at the defined threshold, $O_{th}$. When there is no reference present, the eyes follow the target. Relative retinal motion, $R$, is close to 0, and the $Z$-component required to obtain threshold performance is measured as $Z_0$. From Equation 1, I obtain:

$$O_{th} = eZ_0.$$  

(2)

In the second condition, a stationary reference target was also present. If the eyes continue to follow the moving target, the $Z$-component of the motion delivered as an extraretinal signal is $Z_r$. There is now also a relative motion component, in the opposite direction, of $-Z_r$. Again, following Equation 1, I obtain:

$$O_{th} = Z_r(e - r).$$  

(3)

Equations 2 and 3 can be rearranged to obtain the gain ratio, $e/r = g_z$, given by:

$$g_z = \left( \frac{Z_r}{Z_r - Z_0} \right).$$  

(4)

Therefore, if I measure the extent of motion required for threshold performance with a reference present ($Z_r$), where both retinal and eye movement signals are available, and without a reference ($Z_0$), when only eye motion signals can be used, I can estimate the eye movement compensation gain ratio.

**Methods**

**Stimuli**

Stimuli were presented on a CRT monitor. Active stereo-goggles (Stereographics Inc.), synchronized to the screen refresh rate of 120 Hz, allowed left and right eyes to view left and right stereo half-images via alternate CRT monitor frames. I generated motion of a single target point (subtending a constant size on the screen of 8.3 arcmin, average luminous intensity $3.85 \times 10^{-5}$ cd, equivalent to a 6.6 cd m$^{-2}$ patch of pixels with constant color-gun values), which could be moving directly toward or away from the observer. Screen background was dark, and the experiment took place in a darkened room. Observers were not able to see the screen border or anything else in the room. Note that speed of motion was constant in the world; thus, image speed varied depending on where the point was located in 3-D.

**Procedure**

Using a 2AFC task, observers were shown motion over a range of depths (26 cm behind to 26 cm in front of the target, equivalent to 42.6 min binocular disparity behind the target and 72.5 min disparity in front) and were asked whether the target moved toward them or away from them. On some trials, there was a stationary reference stimulus present 0.32 deg above the moving target (of the same size and luminance as the moving target); on others, only the moving target was present. Stimulus presentation time was 1 s, and viewing distance was 1 m. Trials were randomly interleaved from these two conditions. I used 3 naive observers, who were experienced in depth and/or motion-in-depth experiments.

**Results**

“Percent toward” responses were plotted as a function of the extent of target motion. Data are shown in Figure 1, along with the best fit psychometric functions, fitted using the four-parameter logistic equation.

Observers were sensitive to the motion when a reference was present (filled blue symbols) but found the task much harder when there was no stationary reference (open red symbols). Threshold was defined as half the difference between motion extents corresponding to 75% and 25% on the psychometric function. With the reference present, thresholds, expressed as arcmin, were 4 min for Observer A, 1.8 min for Observer B, and 21 min for Observer C. For Observer A, the threshold without a reference was 18 arcmin. For the other two observers, data were very poorly fit by a sigmoid curve ($r^2$ fits <0.3) because performance was close to chance over the whole range. Thresholds were calculated as 53 and 936 arcmin, respectively. While these estimated values should be viewed with caution, due to the poor psychometric function fit, they do demonstrate that these two observers could barely distinguish between motion toward or motion away.

I used Equation 4 to estimate gain ratios for each observer, which were found to be 0.29 for Observer A, 0.03 for Observer B, and 0.02 for Observer C.

**Discussion**

This experiment revealed that, even for the small targets used in the experiment, the gain ratio was low or sometimes very low. This confirms that $Z$-direction motion is indeed difficult to detect when there is no
reference present, suggesting that eye movements are very poorly compensated for, if at all.

The results in this study are comparable with those already reported in the literature. Erkelens & Collewijn (1985a) used large displays (30 × 30 deg) moving in counterphase in the left and right eyes. No motion-in-depth was perceived. Although there was cue conflict induced because there were no size changes, this did not stop a vivid perception of motion-in-depth being appreciated by their observers when a stationary reference was added to the display. This does strongly suggest that relative motion provides a very powerful cue to motion-in-depth but that information about the vergence eye movement does not (this can only happen when \( g_z \approx 0 \)).

Also of interest is a second paper by Erkelens & Collewijn (1985b), where they were able to measure the absolute retinal disparity of their targets during a vergence movement. Because the eyes do not exactly keep up with a target moving in depth, quite substantial absolute disparities were generated (1–2 deg), but no motion-in-depth was perceived. This suggests a key difference between perception of motion-in-depth and lateral motion, namely, that neither eye movement signals nor absolute disparity changes seem to contribute much to perception of the motion-in-depth. In other words, for a large stimulus, the perceived motion from vergence is almost nonexistent and the perceived motion from an isolated retinal signal is very poor. Again, in that study, there was a cue conflict situation, with size cues signaling no motion. This could explain, at least partially, why no motion-in-depth was seen.

Regan et al. (1986) showed that, for six observers viewing small targets, there was a range of absolute motion-in-depth for which no motion-in-depth was perceived. However, motion-in-depth perception was not completely abolished: Thresholds for detection were raised two to seven times above those when a stationary reference was present. It is important to note that for most observers, this threshold was very close to the largest fusible disparity. In other words, you really have to make the motion-in-depth large for it to be seen. Eye tracking showed that there were large absolute disparities available on the retina that observers were rather insensitive to. This leads one to expect that the visual system will be much more sensitive to the relative motion in the stimuli than the absolute motion.

More recent work also supports the idea that extraretinal signals are not used in 3-D motion perception. Brenner et al. (1996) used randomly shaped 3-D objects moving in depth, which were around 6 deg in size. Changing vergence by around 3 deg (simulated motion-in-depth of 21.6 cm s \(^{-1}\)) over 1 s did not result in any perceived motion.

Part II: Eye movements during 3-D motion direction judgements

Now that I have estimated the gain for Z-motion detection, I go on to consider how eye movements, both version and vergence, would be expected to affect the perception of 3-D motion direction. In this section, I will develop two possible models of how eye movement compensation, or the lack of it, would be expected to bias the perception of 3-D motion direction. I then perform an experiment to distinguish between the two models. I start by describing the combinations of target motion and eye movement to be explored. Target motions were
along a 3-D trajectory, towards the observer at a range of angles between 19.8 deg to the left of straight ahead and 19.8 deg to the right of straight ahead. There were three eye movement conditions:

1. Fixation point stationary directly in front of the observer and presented in the plane of the presentation screen.
2. Fixation point presented in the plane of the screen, moving to the right, with the same X-component speed and final position as the largest X-motion component in the target set.
3. Fixation point moving in depth, with the same Z-component of speed, either in the opposite direction to the target, so that eye motion was away from the observer (Condition 3a), or in the same direction, so that motion was toward the observer (Condition 3b). For Condition 3a, the final Z-position was the same for target and fixation point. For Condition 3b, the final position of the fixation point was as far behind as the target was in front.

Model 1: Full 3-D motion extraction

Here, I assume that the visual system attempts to independently extract the X- and Z-components of 3-D motion, from a combination of retinal motion and eye movement compensation signals. I further assume that the Z-component and the X-component can be treated as separable components, probably mediated by separate mechanisms.

The retinal motion \( R \), with components \( R_x \) and \( R_z \), is defined as the difference between the retinal motion caused by the actual motion of the object, \( O \), with components \( O_x \) and \( O_z \), and the actual eye motion, \( E \), with components \( E_x \) (version eye movement component) and \( E_z \) (vergence eye movement component):

\[
R = \begin{pmatrix} R_x \\ R_z \end{pmatrix} = \begin{pmatrix} O_x - E_x \\ O_z - E_z \end{pmatrix}.
\]

(5)

Note that for the purposes of this study, I am agnostic about whether these vectors represent distances moved or velocity. The transformation of retinal motion from left and right eyes (\( R_{\text{right}}, R_{\text{left}} \)) to X- and Z-relative components (e.g., see Summell & Harris, 2000) is:

\[
R = \begin{pmatrix} R_x \\ R_z \end{pmatrix} \approx \begin{pmatrix} (R_{\text{right}} + R_{\text{left}})/2 \\ R_{\text{right}} - R_{\text{left}} \end{pmatrix}.
\]

(6)

Notice that the retinal motion X-component is the average of the left and right eye motions, whereas the Z-component is obtained from the difference between the two eyes’ retinal motion signals. I now develop Equation 1 to consider both X- and Z-components of motion. The estimated object motion is now a vector, \( O' \), and as before is obtained from a combination of the estimated eye motion, \( E' \), and the estimated relative motion, \( R' \), so that \( O' = E' + R' \). Note that to obtain an estimate of object motion, \( O' \), defined in world coordinates, the contributions of relative motion, \( R \), and eye motion, \( E \), which are both defined in retinal coordinates, must be scaled for viewing distance, \( D \), and interocular separation, \( I \), and can be expressed in terms of X- and Z-components as:

\[
O' = \begin{pmatrix} O_x' \\ O_z' \end{pmatrix} \approx \begin{pmatrix} r_x R_x + e_x E_x \\ r_z R_z + e_z E_z \end{pmatrix} \begin{pmatrix} D \\ D^2/I \end{pmatrix},
\]

(7)

where \( r_x \) and \( r_z \) are the retinal motion gains (i.e., the proportionate error) and \( e_x \) and \( e_z \) are the gains of the extraretinal motion signal. For simplicity, I have assumed that scaling is performed after the relative and eye contributions have been combined.

Equation 7 includes terms that represent some compensation for the eye movement. Full compensation would be represented by a gain of 1, \( e_x = 1 \), and no compensation would be represented by a gain of 0, \( e_x = 0 \). It should be noted that in studies on the effects of eye movements on perception, it is not possible to disentangle the contributions of gain from the eye motion signal and relative motion gain (e.g., Freeman & Banks, 1998). Thus, it is more appropriate to express the gain, \( g_x \), as a ratio, \( g_x = e_x/r_x \).

From the literature on lateral pursuit during target motion, I expect only partial compensation for a version eye movement. For example, when pursuing a fixation target horizontally, the direction of a moving vertical target is perceived to have some horizontal component, demonstrating a gain, \( g_x \), of less than 1 (Festinger et al., 1976; Souman et al., 2005; Swanston & Wade, 1988).

The much smaller literature on perception of absolute changes in vergence predicts close to zero compensation for an eye movement in depth, and this is what I found in Experiment 1. The low gains found suggest that eye movement compensation is very poor.

To summarize, I can be fairly confident that for version, there is partial compensation for eye movements, and hence, \( e_x < r_x \); thus, \( g_x < 1 \). For vergence, it appears that if there is a useful extraretinal signal, it is small, and hence, \( e_z \ll r_z \); thus, \( g_z \ll 1 \).

I now consider what Model 1 would predict when observers perceive object motion during eye movements, given the gain constraints described above. Figure 2a shows solid arrows representing relative motions between the fixation point and target (\( R \)) for three sample trajectory angles, with a stationary fixation point (Condition 1, \( O = R \) if the eyes are stationary).

As viewed from above, the origin was in the plane of the display screen and each motion trajectory contained a component in depth toward the observer (down the page). The solid line in Figure 3a illustrates notional data for
Condition 1, plotting perceived angle as a function of physical angle. From earlier studies (Harris & Dean, 2003; Harris & Drga, 2005; Welchman, Tuck, & Harris, 2004), one can see that observers make large, systematic errors in their responses when asked to estimate the angle of motion. Indicated angles are typically much larger than those displayed and there are large individual differences between observers; hence, the slope $E/\theta$ would be greater than 1.

In Condition 2, the fixation point moved with the same X-component of motion as the trajectory with largest X-component (note that the fixation point motion moved with a constant X-component regardless of the 3-D trajectory presented). This value of fixation X-motion was chosen to make clear predictions about what would be perceived if observers used $R$ but not $E$ (if the eyes move with motion $E_x$ but the extraretinal eye movement signal is not used, $e_x = 0$; hence, $g_x = 0$). The arrows in Figure 2b represent the relative motion ($R$) between fixation point and target when the fixation point moved to the right (Condition 2). Use of only relative retinal motion ($g_z = 0$) to perceive the trajectory would shift the response curve rightward along the X-axis of the notional data plots (dotted red line in graph, Figure 3a), and I would expect the shift to be equivalent to the motion of the fixation point. If the eye movement was partially compensated for ($1 > g_z > 0$), I would expect a smaller but nonzero shift, somewhere between the two lines in Figure 3a. For full compensation ($g_z = 1$), there would be no shift: The data sets would overlap. Remember that, from the existing literature, I expect 50% or better compensation here (e.g., Freeman, 2001).

A different pattern of data was expected when the fixation point moved in depth. I concentrate on predictions made on the basis that the gain is very low (Experiment 1), and here, assume it to be 0 ($g_z = 0$). Later, when the experimental data is analyzed, I will consider more carefully what would happen for other gain values. Here, I used two possible configurations of fixation point movement in $Z$. Note that all target trajectories used a constant $Z$-component of target motion. Figure 2c illustrates the relative motion ($R$) between fixation and target when the fixation point moved away from the observer (Condition 3a). Because target and fixation motions are in opposite directions in depth, if observers used only relative retinal motion ($g_z = 0$), they should report narrower angles (dotted blue line in Figure 3b) than for Condition 1 (solid black line): Specifically, the slope of the function should be half that of the function for Condition 1, plotting perceived angle as a function of physical angle. From earlier studies (Harris & Dean, 2003; Harris & Drga, 2005; Welchman, Tuck, & Harris, 2004), one can see that observers make large, systematic errors in their responses when asked to estimate the angle of motion. Indicated angles are typically much larger than those displayed and there are large individual differences between observers; hence, the slope $\theta'/\theta$ would be greater than 1.

In Condition 2, the fixation point moved with the same X-component of motion as the trajectory with largest X-component (note that the fixation point motion moved with a constant X-component regardless of the 3-D trajectory presented). This value of fixation X-motion was chosen to make clear predictions about what would be perceived if observers used $R$ but not $E$ (if the eyes move with motion $E_x$ but the extraretinal eye movement signal is not used, $e_x = 0$; hence, $g_x = 0$). The arrows in Figure 2b represent the relative motion ($R$) between fixation point and target when the fixation point moved to the right (Condition 2). Use of only relative retinal motion ($g_z = 0$) to perceive the trajectory would shift the response curve rightward along the X-axis of the notional data plots (dotted red line in graph, Figure 3a), and I would expect the shift to be equivalent to the motion of the fixation point. If the eye movement was partially compensated for ($1 > g_z > 0$), I would expect a smaller but nonzero shift, somewhere between the two lines in Figure 3a. For full compensation ($g_z = 1$), there would be no shift: The data sets would overlap. Remember that, from the existing literature, I expect 50% or better compensation here (e.g., Freeman, 2001).

A different pattern of data was expected when the fixation point moved in depth. I concentrate on predictions made on the basis that the gain is very low (Experiment 1), and here, assume it to be 0 ($g_z = 0$). Later, when the experimental data is analyzed, I will consider more carefully what would happen for other gain values. Here, I used two possible configurations of fixation point movement in $Z$. Note that all target trajectories used a constant $Z$-component of target motion. Figure 2c illustrates the relative motion ($R$) between fixation and target when the fixation point moved away from the observer (Condition 3a). Because target and fixation motions are in opposite directions in depth, if observers used only relative retinal motion ($g_z = 0$), they should report narrower angles (dotted blue line in Figure 3b) than for Condition 1 (solid black line): Specifically, the slope of the function should be half that of the function for Condition 1, plotting perceived angle as a function of physical angle. From earlier studies (Harris & Dean, 2003; Harris & Drga, 2005; Welchman, Tuck, & Harris, 2004), one can see that observers make large, systematic errors in their responses when asked to estimate the angle of motion. Indicated angles are typically much larger than those displayed and there are large individual differences between observers; hence, the slope $\theta'/\theta$ would be greater than 1.
Condition 1. If there were partial compensation \((1 > g_z > 0)\), the slope would be somewhere in between.

For Condition 3b, when the fixation point moved toward the observer, it had the same \(Z\)-component of motion as the target itself; the relative motion of all trajectories would therefore be lateral only (but of different magnitude, see multiple arrowheads in Figure 2d). If observers used relative motion only \((g_z = 0)\), the perceived angle should be either +90 deg (when the direction is to the right) or -90 deg (when the direction is to the left), shown by the light blue solid line in Figure 3b; in other words, a sigmoid curve fitted to the data should have infinite slope. If there were partial compensation \((1 > g_z > 0)\), the slope would be between infinity and that for Condition 1. Remember that Experiment 1 suggested that there was almost no useful signal available from vergence eye movements; hence, I expect very little eye movement compensation in this condition.

**Model 2: 3-D motion as a change in visual direction**

Recent work on the perception of 3-D motion direction suggests that, for some tasks, the visual system does not attempt to obtain the \(X\)- and \(Z\)-components of 3-D motion to calculate the trajectory direction. Instead, the change in visual direction between the start point and end point of the motion can be used, along with an estimate of distance to the configuration. For example, a recent study found that large systematic errors in perceived trajectory angle occurred, consistent with the visual system assuming a constant (although often incorrect) \(Z\)-component of motion (Harris & Drga, 2005). This study and related studies use an absolute setting task, rather than forced-choice psychophysics. A concern with such absolute techniques is that observers may develop strategies for performing the task, which are not linked directly to the angle that is perceived. There is evidence to suggest that such concerns are not warranted in this case.

In previous work, we have demonstrated consistency of performance across a variety of tasks and procedures. For example, in a previous study (Harris & Dean, 2003), we compared performance for indicating 3-D motion direction using several different tasks. When we compared a pointer task with a drawing task, the results were not significantly different and were also similar to the angles produced in a verbal and an interception task. In a different study (Welchman et al., 2004), we compared performance using a pointer task for computer-generated stimuli with performance for real target motion along a 2-D motion track. For the former, we used a pointer of radius 25 cm, with pivot point located at the same physical distance as the target start point. For the latter, we used a pointer whose pivot point was located 29 cm from the observer, with a pointer radius of 14 cm. Despite different pointer lengths and pivot-point locations, reported angles were very similar. Further, Welchman et al. (2004) described a control experiment in which pivot-point location was either at the same position as the start point of the motion or was on the table in front of the observers. Again, results were very similar in the two conditions. We are therefore confident that previously published results are not simply artifacts of the chosen task.

More recently, preliminary work has suggested that, rather than using a constant estimate of the distance, the visual system may underestimate the \(Z\)-component (it is known that the \(Z\)-component of speed of 3-D motion is systematically underestimated when compared with the \(Z\)-component; see Brenner et al., 1996; Welchman, Lam, & Buellthoff, 2006, suggest that the \(Z\)-component’s relatively low reliability compared with the \(X\)-component can account for this). Lages (2005, 2006) presented a Bayesian model of 3-D motion trajectory perception where this underestimation was formulated as a prior probability of seeing small disparities (and hence small distances moved in the \(Z\)-direction). Here, I based my 3-D trajectory model on the same assumptions as Harris & Drga (2005), in that observers use a constant although incorrect estimate of the \(Z\)-component of motion.

Consider what information from eye movements would be involved in perception if only the change in visual direction and a constant (incorrect) estimate of \(Z\) were used to obtain motion direction. The change in visual direction as a target moves between two locations is equivalent to the unscaled \(X\)-component of motion. I can therefore define the relative change in visual direction as \(R_x\), as for Model 1. The perceived motion is the sum of retinal \(R_e\) and version \(E_v\) contributions to the motion, scaled for viewing distance, as for Model 1. The key difference is that neither eye nor retinal contributions to the \(Z\)-component of motion are used. Instead, a constant, \(k\), is assumed here:

\[
O' \approx \left( \frac{r_x}{k} + e_x e_v \right) \left( \frac{D}{D^2/l} \right),
\]

where \(k\) is the binocular disparity corresponding to a fixed estimate of \(Z\)-distance moved through, \(Z_{est}\), where \(k \approx l Z_{est} D^2\). Remember that this distance is assumed by the visual system and will be incorrect for all but one actual distance. Again, \(r_x\) and \(e_x\) are the gains on the relative change in lateral position and on version eye movement signal, respectively, and \(g_z = e_x / r_x\). In this study, motion-in-depth was always toward the observer; hence, I expect \(k\) to be constant. For motion away from the observer, I would expect \(k\) to be the opposite sign. Data qualitatively consistent with this hypothesis have been presented in Lages (2005).

Now, consider the possible situation where no eye movement compensation signal is available \((g_z = 0)\) and
when observers use the change in visual direction to estimate 3-D motion trajectory. When the fixation point moves laterally, any eye movement compensation (or noncompensation) will have the same effect on both models (compare Equations 7 and 8). Therefore, I expect the same biases in performance. The solid line in Figure 3c illustrates notional data for Condition 1, plotting perceived angle, $\theta'$, as a function of physical angle, $\theta$. Once again, use of purely relative changes in visual direction ($g_x = 0$) would result in responses that would be shifted rightward along the X-axis (dotted line in graph, Figure 3c). Again, I expect partial compensation ($0 < g_x < 1$) when the eyes move; hence, the shift of the curve in Figure 3c may be smaller than if only relative motion were used.

A pattern of data quite different to that for Model 1 was expected when the fixation point moved in depth. Equation 8 shows that neither changes in relative depth moved through, $R_z$, or eye movement compensation signals, $E_z$, will affect the perceived trajectory. Figure 3d illustrates this point, showing a plot of indicated trajectory angle as a function of physical trajectory angle for Condition 1, with no fixation motion, and Conditions 3a and 3b, where the fixation point moved away and toward the nose, respectively. Notice that the three curves have essentially identical slopes.

Model 2 therefore predicts very different effects from Model 1 when the eyes move to follow motion-in-depth, unless there is a perfect relative gain of $g_x = 1$, when Model 1 performance would be the same as Model 2 (see paragraphs above and Experiment 1 for discussion of why this is very unlikely). Perception independent of eye movements occurs under Model 2 because changes in the Z-component of motion (whether caused by object or vergence eye movement) have very little effect on visual direction. Thus, under Model 2, compensation for vergence eye movements is unnecessary. Below, I will describe an experiment that tests which of the two models best fits human performance when observers are asked to indicate 3-D motion trajectory directions while the fixation point moves.

**Experiment 2: Trajectories during eye movements**

I measured the perceived trajectory angle of 3-D object motion when the eyes were fixated on a stationary reference point, when the fixation point moved laterally, and when the fixation point moved in depth.

**Methods**

**Stimuli**

For stimulus generation and presentation, I used the same apparatus as for Experiment 1. At the 1-m viewing distance, I simulated motion of a single target point (subtending a constant size on the screen of 8.3 arcmin, average luminous intensity $3.85 \times 10^{-5}$ cd, equivalent to a $6.6$ cd m$^{-2}$ patch of pixels with the same color-gun values) moving in a straight line along different 3-D trajectories for 1 s. The screen background was dark; the experiment took place in a darkened room, and observers were not able to see the screen border or anything else in the room. Speed of motion was constant in the world; thus, image speed varied depending on where the point was located in 3-D. The motion extended in depth for 0.132 m (approximately 31 arcmin binocular disparity) from the screen, toward the observer. Trajectory angle was varied between 0 deg (directly toward the nose) and 19.8 deg (note that I used trajectory angles up to 19.8 deg; the eccentricity of each point, its visual angle, is always smaller than 3.13 deg; hence the use of the small angle approximation in Equations 7 and 8 is justified) by using different lateral motion components (lateral position of end points were 0, 0.012, 0.024, 0.037, and 0.0475 m from the start point directly in front of the nose; positive X-direction is defined as rightward as seen by the observer). The lateral motion component could be either to the left or to the right of the nose.

The stimulus comprised only the moving target and a fixation point (subtending 8.3 arcmin, located 0.32 deg above the moving target). The pair of points was presented in darkness, and no other visual reference was present. There were three different fixation conditions:

1. The fixation point was stationary directly in front of the observer and presented in the plane of the presentation screen, 1 m from the observer.
2. The fixation point was presented in the plane of the screen and moved laterally to the right, with the same speed and final position as the largest lateral motion component above (speed: 0.0475 m s$^{-1}$, 2.72 deg s$^{-1}$; extent: 0.0475 m, 2.72 deg).
3. The fixation point moved in depth, with the same speed as the depth component of the motion (0.132 m/s; extent: 0.132 m). Motion could be away from the observer (Condition 3a, disparity extent: 24.1 arcmin) or toward the observer (Condition 3b, disparity extent: 31.4 arcmin; the same distance in depth corresponds to slightly different disparities when in front or behind the start point).

Observers were informed that the fixation point might move and that, if it did, they should follow it with their eyes. They were also told that the target could follow any trajectory along any angle from +90 deg to −90 deg.

Due to the use of active stereo-goggles, I was unable to monitor either version or vergence eye movements in this experiment. The equations developed above assume accurate following eye movements. The gains, $g_x$, and $g_z$, refer to the extent to which retinal and eye movement signals are used, rather than the extent to which the eyes
can accurately follow the target. However, the lateral pursuit stimulus in Condition 2 was optimal for high version gain. High pursuit gain is typically found for pursuit stimuli moving at around 2 deg s$^{-1}$ (e.g., see Freeman, 2001; Turano & Heidenreich, 1999). Further, such high gain is usually achieved around 100 ms after stimulus onset (Pola & Wyatt, 1991). I would also expect good vergence following for the stimuli in Condition 3. After a delay of around 160 ms (Rashbass & Westheimer, 1961), high gain vergence is seen for speeds of binocular disparity change of up to 12 deg s$^{-1}$.

Procedures

Six observers took part, five were completely naïve and were paid for their participation. Observers were asked to indicate the motion direction of the target by moving a pointer to indicate the perceived angle of motion with respect to the midline. The pointing apparatus consisted of a square wooden block (30 x 30 cm), on which a narrow wooden pointer that could be rotated around the center of the block (radius 14 cm) was mounted. The pointer was positioned horizontally on a desk in front of the observer. Observers were required to move the pointer to indicate the motion direction.

Before each trial began, the room lights were extinguished and observers were asked to fixate the stationary fixation point. The experimenter pressed a button to start the trial; the target stimulus appeared, and either the target moved for 1 s along a 3-D trajectory or both the target and fixation point moved for 1 s. A light was then turned on to allow observers to see the pointer while they moved it to the desired angle. Observers were encouraged to report the motion direction as quickly as possible, and most responded within 0.5 s. The indicated angle, $\theta$, was recorded by the experimenter (by reading from a scale marked on the pointer mount), the lights were extinguished, and the pointer was reset to zero before the next trial began.

For each experimental condition, observers performed two blocks of trials (in each block, only one fixation condition applied). In each block, they were asked to make two settings for each of the nine different presented trajectory angles, $\theta$.

Results

I first measured trajectory angles perceived for a stationary fixation point and plotted perceived angle as a function of physically presented trajectory angle (Condition 1). Observers did not experience motion of the fixation point in this condition.

For Condition 2, the fixation point moved in the X-direction. Some observers noticed that the fixation point moved laterally in this condition. I expected the perceived angle data to be shifted along the X-axis, under either of the models described above. It is possible to estimate the size of the shift by fitting a sigmoid curve using each observer’s data for Condition 2 and measuring the X-axis intercept. I obtained X-axis intercepts for each individual separately, as well as an average X-axis intercept, using the pooled data from six observers.

For Condition 3, the fixation point moved either away from the observers (Condition 3a) or toward them (Condition 3b). No observers reported noticing that the fixation point had moved. I was interested in the relative slopes of perceived versus physical trajectory for each condition pair. Under Model 1, assuming no compensation for eye movements ($g_x = 0$), I expected a shallow slope for Condition 3a, when the eyes moved away (half that for Condition 1) and a very steep one for Condition 3b, when the eyes moved toward the observer. Under Model 2, I expected the slopes for Conditions 3a and 3b to be very similar to that for Condition 1. Best-fit sigmoid functions were calculated and their slope recorded. This was done separately for each observer, as well as for the average data. I was then able to calculate the ratio of slopes for each condition pair. Under Model 1 (with $g_x = 0$), ratios of 0 were expected for Conditions 1/3b and 3a/3b and ratios of 2 for Conditions 1/3a. Under Model 2, all ratios should be 1.

Figure 4a shows average data and best fit curves, from six observers, for Condition 1, where the fixation point was stationary (open circles), and Condition 2, where the fixation point moved to the right (filled red triangles). The best fit curve had an X-axis intercept at 5.99 deg. This can be transformed to give a change of 1.38 cm in the apparent maximum distance moved in the X-direction. The actual intercept, had observers used only relative motion ($g_x = 0$), would have been 4.75 cm or at 19.8 deg on the x-axis (shown by the arrow in Figure 4a). X-component intercepts, calculated separately for each observer, are shown in Figure 4b. Shifts were very different across observers, but notice that none was as large as 4.75 cm. This suggests that none of the observers completely relied on the relative motion between fixation point and target. All to some extent appeared to be partially compensate for their pursuit eye movements ($0 < g_x < 1$), and some had excellent compensation (very small intercepts; hence, $g_x$ was just less than 1 for several observers, see right-hand axis in Figure 4b). By comparing Conditions 1 and 2, one can tell that there was partial compensation for a lateral eye movement, but it cannot distinguish between the two models described above.

Next, I compared data from Conditions 1 and 3. Figure 5a shows average data, from the same six observers, for Condition 1, where the fixation point was stationary (open circles), Condition 3a, where the fixation point moved away (open blue squares), and Condition 3b, where the fixation point moved toward the observer (filled light blue triangles). The data are very similar for all conditions. The slope of the best fit sigmoid for Condition 3a (dark blue...
line) is slightly shallower than that for the other two conditions.

Ratios of slopes of each observer for each of the three condition pairs are shown in Figure 5b. Remember that Model 1, if only relative motion were used and there was no compensation for eye movements (\( g_z = 0 \)), predicted that the ratio of slopes for Conditions 3a/3b (top graph) would be 0, that the ratio of slopes for Conditions 1/3a (middle graph) would be 2, and that the ratio of slopes for Conditions 1/3b (lower graph) would be 0. Model 2 predicted that the ratio would be 1 for all three pairs. All observers showed ratios close to 1, as predicted by Model 2, rather than falling close to the predictions for Model 1 (if \( g_z = 0 \)).

The data suggest a mechanism following Model 2, rather than Model 1. However, recall that the predictions for Model 1 were developed on the assumption that there was no compensation for eye movement signals (that \( g_z = 0 \)). Using Equation 7, it can be shown that, for the X- and Z-components used in this experiment, the gain, \( g_z \), is related to the above ratios by the following equation:

\[
g_z = 2 - \frac{\text{slope } 1}{\text{slope } 3a}
\]

Figure 5. (a) Perceived angle as a function of physical trajectory angle, averaged across six observers. Open circles show performance for Condition 1 (stationary fixation point). Filled red triangles show performance for Condition 2 (fixation point moved laterally). The solid lines are the best fit sigmoid curves fitted to the data using the four-parameter logistic equation. The vertical arrow shows where the expected X-axis intercept would be in Condition 2, if \( g_x = 0 \). (b) I also fitted sigmoids to individual observer data and found the actual angle at which perceived angle was zero: the X-axis shift. Best fit shifts are shown in the bar graph, expressed in centimeters of lateral motion, for each observer (equivalent gain is shown on the right axis). A shift of 4.25 cm would correspond to complete use of relative motion (\( g_x = 0 \)).

Figure 4. (a) Perceived angle as a function of physical trajectory angle, averaged across six observers. Open circles show performance for Condition 1 (stationary fixation point). Filled red triangles show performance for Condition 2 (fixation point moved laterally). The solid lines are the best fit sigmoid curves fitted to the data using the four-parameter logistic equation. The vertical arrow shows where the expected X-axis intercept would be in Condition 2, if \( g_x = 0 \). (b) I also fitted sigmoids to individual observer data and found the actual angle at which perceived angle was zero: the X-axis shift. Best fit shifts are shown in the bar graph, expressed in centimeters of lateral motion, for each observer (equivalent gain is shown on the right axis). A shift of 4.25 cm would correspond to complete use of relative motion (\( g_x = 0 \)).

\[
g_z = 2 - \frac{\text{slope } 1}{\text{slope } 3a}
\]
Equations 9 and 10 can be used to estimate the gain that would be needed to explain observer performance in this experiment, under Model 1. Remember that, in Experiment 1, gains were very low for detection of the Z-component of motion when the eyes moved to follow a target. Figure 6 shows calculated gains using Equations 5 and 6 for Conditions 3a and 3b.

Gains ranged from 0.68 to 1.24, and there were large, nonsystematic differences between the two conditions. Gains above 1 are nonsensical and must therefore reflect experimental error. What is consistent across the data is that all the calculated gains are large, averaging 0.96. In principle, it is therefore possible that Model 1 is used but only if there is almost complete compensation for vergence eye movements ($g_z \approx 1$).

**Discussion**

The results of Experiment 2 would be very surprising if they really reflected close to full compensation for vergence ($g_z = 1$). For performance to be very similar when the eyes moved either with or against the target in depth, the gains on the signals from eye and retina would have to be almost identical. From previous literature (see details where models were developed above), one can see that the vergence signal appears to be a very poor source of information about object motion when an object moving in depth is followed using a vergence eye movement. More importantly, the gains calculated here are not consistent with those from Experiment 1, where I measured thresholds for Z-motion detection in the presence and absence of a stationary reference. Gains there were found to be 0.28, 0.03, and 0.02 for three observers. I therefore reject Model 1 on the basis that if the gain on the vergence signal to detect Z-motion direction (control experiment) is almost zero, it should also be almost zero for the Z-component of motion in the 3-D direction task. Instead, I suggest that Model 2 may better describe the visual mechanisms used to perform the 3-D motion direction task, using visual direction and a constant estimate of distance, and hence not taking account of vergence eye movements at all.

**Part III: General discussion**

The aim of this paper was to explore how both version and vergence eye movements affect the perception of 3-D motion trajectory. In the first section, I measured the relative contributions of retinal and extraretinal information to the detection of Z-direction motion. The $e_z/r_z$ gain ratio was very low, demonstrating a relatively poor contribution of vergence signal to the motion perceived. In the second section, I considered some of the theoretical issues involved in suggesting models of how eye movements may interact with perception when either the object or eyes move in depth. Two models were developed for how eye and visual signals might interact when observers are asked to determine the direction of an object moving in 3-D. In the experiment, I found biases in the perceived trajectories when the fixation point moved laterally but not when the fixation point moved in depth, consistent with a model based on the use of visual direction information.
Below, I will discuss each of these results in the context of recent studies of motion and position processing during version and vergence eye movements.

**Lateral movement of the fixation point**

The key issue under consideration is whether the brain behaves as if using Model 1 or Model 2. The condition in which the fixation point moved laterally does not distinguish between these models. It does, however, allow one to measure the relative extent to which the version eye movement signals are combined with retinal signals. Note that, in these experiments (as for others in the field), it was not possible to obtain absolute estimate of retinal gain, $r_x$, or eye movement gain, $e_x$. Rather, errors in the perceived direction give one an indication of the ratio $g_x = e_x/r_x$, which tells us about the relative amounts of retinal and eye information used (e.g., see Freeman & Banks, 1998).

I found that all observers showed biases, consistent with the presence of an imperfect gain on the eye signal, as a ratio of the retinal signal ($g_x < 1$). Some observers showed just a small bias, but for others, bias was high and, hence, gains were low. This result concurs with the literature on how version eye movements affect lateral motion perception. Most studies find some compensation for pursuit eye movements (e.g., Freeman, 2001; Freeman & Banks, 1998; Swanston & Wade, 1988). Several studies have shown that observers sometimes appear to use only retinal information and, therefore, do not take account of eye movements. Festinger et al. (1976) asked observers to judge the direction of a target point moving physically upward during horizontal pursuit of a different point. Observers appeared to use the relative retinal information between target and fixation point, perceiving the motion as if along an oblique trajectory, and hence did not compensate at all for the eye movements. Rock & Linnett (1993) asked observers to perform a shape discrimination task, based on lines flashed at 250- to 400-ms intervals. With no reference, eye position information was taken into account; yet, when a reference was present, shape judgements were made relative to the reference, even if it was itself moving and hence causing misperceptions of shape. Another study on shape perception also found that shape was largely determined by relative information (Li, Brenner, Cornelissen, & Kim, 2002). Brenner & Cornelissen (2000) showed that, when setting the relative distance between a pair of sequentially flashed targets during pursuit, observers set the distance as if using only retinal information, despite taking account of pursuit eye movements when setting the absolute position of one of the targets. Although it is difficult to generalize across very different studies, it seems that eye position can be taken into account when evaluating egocentric locations. When reference information is available in the retinal signal, that is, when there is relative motion between points in the retinal image, that information appears to be used in isolation, even if it is misleading.

**Depth movement of the fixation point**

The results for Condition 3, when the fixation point moved in depth, were clear and fit with Model 2, which assumed that the visual system estimates a change in visual direction of the end point of the motion. I argue that I obtained this result because the visual system does not attempt to obtain a full 3-D representation of the motion trajectory.

In the introduction to Model 1, I argued, based on previous literature and the results of Experiment 1, that I would expect compensation for vergence eye movements to be close to 0 ($g_z = 0$). If instead I leave a nonzero $E_z$ term in Model 1, the model could fit the data in Figure 5, but only if almost complete compensation for vergence eye movements occurred ($g_z \approx 1$). Although this possibility cannot be ruled out here, I think that this is highly unlikely, given what is known about the visual system’s sensitivity to absolute motion-in-depth and because of the results of Experiment 1. First, gain differences between different classes of neural signal are pervasive in the visual system; it would be odd that they were completely balanced for this one particular scenario. Second, motion-in-depth is barely perceived when a small target moves in depth with no reference present, suggesting that the visual system is not very sensitive to the extraretinal vergence signal (see Experiment 1; Brenner et al., 1996; Regan et al., 1986). Hence, I would expect relative information to almost always be used. If it was, and Model 1 were correct, the results for Conditions 3a and 3b should have been very different from one another, in contrast to what was actually found (Figure 5). If observers are unable to tell whether an object they are fixating is moving toward or away from them in depth, then it seems unlikely that any binocular extraretinal eye movement signal is being fed back into the systems used to perceive motion-in-depth direction.

**What if the eyes did not move?**

I did not measure eye movements in this study. If the eyes had not moved at all, would the interpretation of the results in this study be affected? If $E_x$ or $E_z$ were zero, the results would look like those predicted in Figure 3 for $g_x$ and $g_z$ being zero. They did not. Further, it is unlikely that the eyes did not move because the only visible points in the scene were the moving fixation point and the moving target. There was no stationary reference in Condition 2 or Condition 3, which would make it very difficult to hold the eyes still. For Condition 2, if the eyes had not undergone a version movement to follow the fixation point, only relative information would be used,
and I would expect the results to show that $g_x = 0$. This did not occur. For Condition 3, if the eyes were stationary, only relative motion information would be available. Model 1 would predict dramatically different results for Conditions 3a and 3b because the relative motion was very different in these two conditions (Figure 3b). On the other hand, Model 2 relies only on the X-component of relative motion, which was the same in Conditions 3a and 3b. Hence, even with little or no eye movements, Model 1 would not predict the results found in this study.

Generalizing these results

I have used a restricted range of stimulus parameters here, mainly because I was concerned that all target stimuli should remain within the fusible range. One clearly needs the depth to be as large as possible so that the brain has the best chance of exploiting the binocular disparity information it provides. The problem is that of using depth components large enough to have a maximal effect but small enough to not become diplopic. Motion trajectories extended for 31.4 arcmin disparity in depth toward the observer. While most observers did not experience diplopia of the target if asked to fixate the stationary reference, a few did. I therefore chose this disparity as the best compromise.

The speed of lateral pursuit was chosen to equal that of the largest horizontal component of target motion to make the model predictions clear and to cover the same target motion ranges as in previous studies (Harris & Dean, 2003; Harris & Draga, 2005). The data here do not address what would happen for larger speeds or depth extents outside the fused range. I hope to characterize the interactions between speed, vergence eye movement gain, and perception in a larger study, where precise eye movement measurements can be made.

Conclusions

For pursuit of a fused target that moves in depth, the visual system appears not to use vergence eye movement signals in the perception of the $Z$-component of 3-D motion. Despite this, perception of 3-D motion direction is not biased when the eyes move to follow another target; I suggest that this is because a full 3-D representation of motion trajectory is not required.

Acknowledgments

I would like to thank Philip Dean for collecting a large part of the data for this project. The project was funded by the EPSRC (UK). Thanks to Vit Draga, Harold Nefs, Andy Welchman, the anonymous reviewers, and the Journal of Vision editor for their comments on earlier drafts of the manuscript.

Commercial relationships: none.
Corresponding author: Julie M. Harris.
Email: Julie.Harris@st-andrews.ac.uk.
Address: School of Psychology, University of St. Andrews, St. Mary’s College, South Street, St. Andrews, KY16 9JP, Scotland, UK.

References


