

# Bayesian integration of visual and vestibular signals for heading

**John S. Butler**

**Stuart T. Smith**

Max-Planck Institute for Biological Cybernetics,  
Tübingen, Germany, &  
Albert Einstein College of Medicine, Bronx, NY, USA

Neuroscience Research Australia, Randwick, Australia



Max-Planck Institute for Biological Cybernetics,  
Tübingen, Germany,  
Toronto Rehabilitation Institute, Toronto, Canada,  
Department of Psychology, University of Toronto,  
Toronto, Canada, &  
Centre for Vision Research, York University,  
Toronto, Canada



**Jennifer L. Campos**

Max-Planck Institute for Biological Cybernetics,  
Tübingen, Germany, &  
Department of Brain and Cognitive Engineering,  
Korea University, South Korea



**Heinrich H. Bühlhoff**

Self-motion through an environment involves a composite of signals such as visual and vestibular cues. Building upon previous results showing that visual and vestibular signals combine in a statistically optimal fashion, we investigated the relative weights of visual and vestibular cues during self-motion. This experiment was comprised of three experimental conditions: vestibular alone, visual alone (with four different standard heading values), and visual–vestibular combined. In the combined cue condition, inter-sensory conflicts were introduced ( $\Delta = \pm 6^\circ$  or  $\pm 10^\circ$ ). Participants performed a 2-interval forced choice task in all conditions and were asked to judge in which of the two intervals they moved more to the right. The cue-conflict condition revealed the relative weights associated with each modality. We found that even when there was a relatively large conflict between the visual and vestibular cues, participants exhibited a statistically optimal reduction in variance. On the other hand, we found that the pattern of results in the unimodal conditions did not predict the weights in the combined cue condition. Specifically, visual–vestibular cue combination was not predicted solely by the reliability of each cue, but rather more weight was given to the vestibular cue.

**Keywords:** visual, vestibular, heading, Bayesian, spatial conflict, human

**Citation:** Butler, J. S., Smith, S. T., Campos, J. L., & Bühlhoff, H. H. (2010). Bayesian integration of visual and vestibular signals for heading. *Journal of Vision*, 10(11):23, 1–13, <http://www.journalofvision.org/content/10/11/23>, doi:10.1167/10.11.23.

## Introduction

When passively moving throughout the environment, humans are able to easily detect the direction of their movements. This is achieved through a composite of sensory and motor information including dynamic visual information (i.e., optic flow) and non-visual information (i.e., vestibular, proprioception, etc.). Optic flow consists of the pattern of retinal information that is generated during self-motion through space (Gibson, 1950). Inertial inputs are provided mainly through the vestibular organs in the inner ear, the otoliths and semicircular canals, which detect linear and rotational accelerations, respectively (Angelaki & Cullen, 2008). While there has been a

great deal of important research investigating how each of these cues can be used independently to perceive heading direction, far less is understood about the relative contributions of each when both cues are available. Because heading estimation is essential to effective navigation and is something that humans are able to perform relatively effortlessly, this paradigm provides a unique opportunity to better evaluate and quantify multisensory self-motion perception.

## Optic flow cues

Optic flow can be used to detect various different properties of self-motion, including speed and distance

(Frenz & Lappe, 2005, 2006; Harris, Jenkin, & Zikovitz, 2000; Lappe, Bremmer, & van den Berg, 1999; Redlick, Jenkin, & Harris, 2001; Sun, Campos, & Chan, 2004; Sun, Campos, Young, Chan, & Ellard, 2004; van den Berg, 1992; Wilkie & Wann, 2005). The capacity for observers to use optic flow to estimate heading has also been actively investigated. For instance, heading estimation based on optic flow has been studied using discrimination tasks (Britten & Van Wezel, 2002; Crowell & Banks, 1993; Fetsch, Turner, DeAngelis, & Angelaki, 2009a), magnitude estimation tasks using a pointer to indicate heading (Ohmi, 1996), and tasks that require an observer to select from a distribution of heading directions represented by oriented lines (Royden, Banks, & Crowell, 1992). The accuracy of heading judgments based on optic flow alone has also been shown to be contingent on factors such as heading eccentricities (i.e., the angle between the heading direction and the center of the stimulus; Crowell & Banks, 1993), the speed of simulated self-motion (Warren, Morris, & Kalish, 1988), the ability to compensate for eye movements using information from the extra-ocular muscles (Royden et al., 1992), and the coherence of the dots presented in the display (Fetsch et al., 2009a; Gu, Angelaki, & Deangelis, 2008; Gu, DeAngelis, & Angelaki, 2007).

Human behavioral data are consistent with monkey neurophysiological studies, which have shown that there are neurons in the primate extrastriate visual cortex that respond preferentially to complex optic flow (Duffy & Wurtz, 1991; Lappe, Bremmer, Pekel, Thiele, & Hoffmann, 1996; Orban et al., 1992; Schaafsma & Duysens, 1996; Tanaka & Saito, 1989). In particular, neurons within the Medial Superior Temporal (MST) area have response properties appropriate for the precise encoding of visual heading (Gu et al., 2008, 2007; Gu, Watkins, Angelaki, & DeAngelis, 2006; Heuer & Britten, 2004; Perrone & Stone, 1998) and stimulation studies reveal a causal link to heading judgments (Britten & van Wezel, 1998, 2002).

## Vestibular cues

The vestibular system is important for a variety of behaviors ranging from highly reflexive responses (e.g., the vestibulo-ocular reflex) to maintaining posture and balance, to linear and rotational self-motion perception, and finally, complex navigational processing. It has been well documented that humans can, to some extent, perform displacement, velocity, and acceleration judgments using non-visual cues alone (Bertin & Berthoz, 2004; Bühlhoff, Little, & Poggio, 1989; Harris et al., 2000; Israel, Chapuis, Glasauer, Charade, & Berthoz, 1993; Israel, Grasso, Georges-Francois, Tsuzuku, & Berthoz, 1997; Jurgens & Becker, 2006; Siegle, Campos, Mohler, Loomis, & Bühlhoff, 2009; Sun, Campos, Young et al., 2004; Wallis, Chatziasstros, & Bühlhoff, 2002).

Importantly, vestibular information in the absence of visual information (e.g., passive movement in the dark)

has also been shown to be generally sufficient for detecting heading in humans and non-human primates (Butler, Smith, Beykirch, & Bühlhoff, 2006; Fetsch et al., 2009a; Gu et al., 2008, 2007; Gu, Fetsch, Adeyemo, Deangelis, & Angelaki, 2010; Ohmi, 1996; Telford, Howard, & Ohmi, 1995). However, there is some evidence that vestibular heading estimates can be more variable than visual heading estimates depending on the task. For instance, Ohmi (1996) reported that, when using a pointing task, heading judgments based on vestibular inputs alone appeared to be much more variable than conditions in which only optic flow was available (see also Telford et al., 1995). However, when using a 2-interval forced choice task, observed thresholds are similar to those observed for heading estimates derived from optic flow alone (i.e.,  $1^{\circ}$ – $4^{\circ}$ ) in both humans (Butler et al., 2006; Fetsch et al., 2009a) and non-human primates (Fetsch et al., 2009a; Gu et al., 2008).

Recently, important neurophysiological studies have also demonstrated a functional and behavioral link between the Medial Superior Temporal dorsal (MSTd) area and heading perception based solely on vestibular signals (Gu et al., 2008, 2007, 2010, 2006). Because MSTd has long been associated with optic flow processing, it is very interesting that it also contains neurons that respond to passive self-motion trajectories in the dark. This intriguing finding emphasizes the need to take a multimodal approach to evaluating egocentric heading perception.

## Combined visual and vestibular heading perception

Even though visual and vestibular information appear to be sufficient for estimating various properties of self-motion perception when isolated, there are many reasons why it would be advantageous for the brain to use a combination of cues when more than one is available. For instance, visual information may not be entirely reliable during dimly lit conditions (e.g., driving at night) or in sparse visual environments providing little optic flow (e.g., translating in heavy fog). However, visual information would be extremely important during accelerations that are below the vestibular threshold, during movements at a constant velocity, or during conditions in which vestibular information is ambiguous. For instance, because the otoliths cannot distinguish linear accelerations from backward tilt, this can lead to self-motion illusions such as the somatogravic illusion (MacNeilage, Banks, Berger, & Bühlhoff, 2007).

In general, there have been relatively few attempts to quantify the relative weighting of visual and vestibular cues for self-motion perception. Harris et al. (2000) investigated the relative contributions of visual and vestibular cues when estimating egocentric displacement and observed a higher weighting of vestibular cues. Bertin and Berthoz (2004) also reported a reduction in errors

when participants were asked to reproduce a trajectory experienced with an initial vestibular stimulation combined with a prolonged visual stimulus. Telford et al. (1995) showed slight, non-significant improvements in heading estimates when vestibular information was combined with visual inputs.

New evidence from recent studies now supports the contention that visual and vestibular information combine in a statistically optimal fashion (Fetsch et al., 2009a; Gu et al., 2008). Specifically, it has been shown that both psychophysical measures and neural responses demonstrate a reduction in variance when visual and vestibular cues are combined, compared to the patterns of responding observed in the unisensory conditions. While there is also some evidence to suggest that optimal integration in naïve humans appears to occur only when the optic flow stimuli are presented in stereo rather than binocularly (Butler, Campos, Bühlhoff, & Smith, submitted), this effect has not been observed in well-trained, non-human primates (Fetsch et al., 2009a). In general, these results conform to the predictions specified by Bayesian decision theory that have been widely supported by other types of cue-integration paradigms (e.g., visual-auditory (Alais & Burr, 2004) and visual-haptic (Bühlhoff & Yuille, 1991; Ernst & Banks, 2002; Ernst, Banks, & Bühlhoff, 2000; Yuille & Bühlhoff, 1996)).

## Current study

While the above mentioned reductions in variance observed during combined cue conditions support the idea that optimal integration occurs for visual–vestibular interactions, these results only provide *predictions* regarding relative cue-weighting rather than the actual weighted values. In order to experimentally quantify these cue weights, it is important to not only compare combined cue conditions to unisensory conditions but also to evaluate conditions under which the two cues are presented simultaneously, yet provide discrepant information about the same movement parameter (i.e., movement direction in this case; see also Fetsch et al., 2009a). In the current experiment, we achieved this by creating a spatial conflict ( $\Delta = \pm 6^\circ$  or  $\pm 10^\circ$ ) between the heading direction specified by optic flow and that specified by passive linear motion (i.e., vestibular signals). Participants performed a 2-interval forced choice task in which they were asked to report which of the two movements were directed “more to the right.” The egocentric heading information was provided either through optic flow alone, vestibular information alone, or both combined (either congruent or incongruent).

Here we show that all subjects combined visual and vestibular signals in a statistically optimal fashion such that a reduction in variance was observed during combined cue conditions relative to each of the unisensory conditions. However, in this case, a simple linear cue combination

process did not predict the visual and vestibular weights, but rather, a consistent bias toward the vestibular cue was observed.

## Methods

### Participants

Eight (5 male) observers experienced in psychophysical testing participated in the experiment for payment. All were right-handed, had normal or corrected-to-normal vision and all but one (author, JB) were naïve to the purpose of the experiment. The average age was 25 years (range 21–32). Participants gave their informed consent before taking part in the experiment, which was performed in accordance with the ethical standards specified in the 1964 Declaration of Helsinki.

### Apparatus

The experiment was conducted in the Motion Laboratory at the Max Planck Institute for Biological Cybernetics, which consists of a Maxcue 600, six-degree-of-freedom Stewart platform manufactured by Motion-Base PLC (see Figure 1). Participants wore noise-cancellation headphones with two-way communication capability, while white noise was played to mask the noise of the platform. Subwoofers installed underneath the seat and foot plate were used to produce somatosensory vibrations to mask the vibrations from the platform motors. Participants responded using a simple four-button response box. All visual motion information was displayed on a projection screen, with a field of view of  $86^\circ \times 65^\circ$  and a resolution of  $1400 \times 1050$  pixels with a refresh rate of 60 frames per second. Participants viewed the projection screen through an aperture, which reduced the field of view to  $50^\circ \times 50^\circ$ . This ensured that the edges of the screen could not be seen, thereby increasing immersion and avoiding conflicting information provided by the stability of the frame and the visual motion information being projected on the screen. The image was presented in stereo generated using red–cyan anaglyphs. All experiments were coded using a graphical real-time interactive programming language (Virttools, France).

### Stimuli

The visual stimulus consisted of a limited lifetime starfield. Each star was a Gaussian blob with a limited lifetime in the range of 0.5–1.0 s. The maximum number of Gaussian blobs on the screen at one time was 200 and the minimum was 150. The participants were seated 100 cm from the screen. All blobs were the same size, but as a

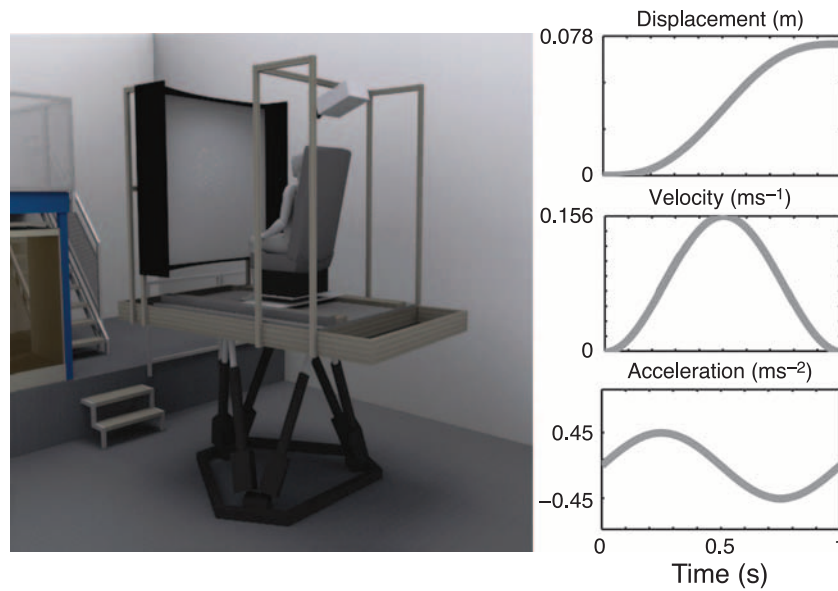


Figure 1. (Left) Apparatus and (right) motion profile.

result of their virtual depth, they subtended angles of  $0.1^\circ$  to  $0.2^\circ$ , which corresponded to a virtual depth of 1.4 m to 0.6 m. The starfield was presented in stereo on a gray background to facilitate the fusing of the red and cyan stereo images and reduce the cross-talk between the red and green images. The visual and vestibular motion profile (m) was

$$s(t) = 0.049 \frac{(2\pi t - \sin(2\pi t))}{4\pi^2}, \quad 0 \leq t \leq 1s, \quad (1)$$

where  $t$  is time. The motion profile had a maximum displacement of 0.078 m, a raised cosine velocity profile with a maximum velocity of  $0.156 \text{ m s}^{-1}$ , and a sine-wave acceleration profile with a maximum acceleration of  $0.45 \text{ m s}^{-2}$  (see Figure 1). To maximize the effects of combining the two cues, the motion parameters and stimuli were intentionally chosen to ensure that the thresholds in the two unisensory conditions were comparable. This was done in order to reveal the maximum effects of combining the two cues. These motion parameters were also well above detection threshold (Benson, Spencer, & Stott, 1986; van den Berg, 1992).

## Procedure

Participants performed a 2-interval forced choice task (2IFC) in which they were asked to judge “in which of the two intervals did you move more to the right” (see Figure 2). Each trial consisted of two linear heading motions, the standard and the comparison (counterbalanced across trials).

In the vestibular condition, the standard angle was always fixed at  $\Theta = 0^\circ$  (the naso-occipital direction), while

the eight comparison angles ranged from  $-10^\circ$  to  $10^\circ$  on an equidistant log scale centered around  $0^\circ$ . In the visual alone condition, there were four standard headings,  $\Theta = -10^\circ, -6^\circ, 6^\circ, 10^\circ$ , with eight comparison angles centered

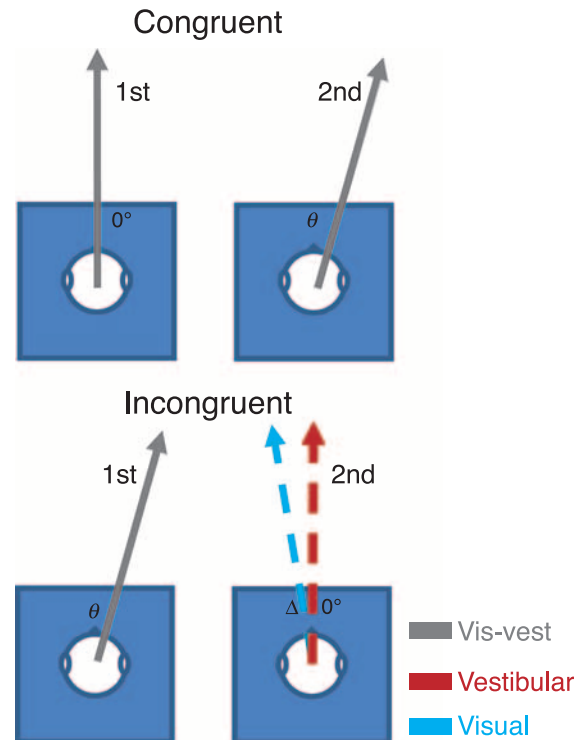


Figure 2. Experimental procedure for the congruent and incongruent trials: Participants were translated along two separate heading directions and judged in which of the two movements they moved more to the right. In the example,  $\theta$  is the comparison heading to the standard,  $0^\circ$ , which in the incongruent trials had a conflict,  $\Delta$ , between the visual and the vestibular cues.



around each of the standards. In the bimodal condition, for the standard heading, the visual and vestibular cues were either congruent or incongruent. In the incongruent conditions, the vestibular component was straight ahead ( $\Theta = 0^\circ$ ) and the visual was offset by  $\Delta$ . We used four different conflict levels corresponding to the four visual alone standards,  $\Delta = \pm 6^\circ$  and  $\Delta = \pm 10^\circ$ . The eight comparison angles were chosen using the method of constant stimuli from the range  $\theta = (-10^\circ + \Theta + \Delta/2, 10^\circ + \Theta + \Delta/2)$ .

Each trial was initiated with an auditory beep played over the headphones instructing the participant that they could begin the trial with a button press. After pressing the start button, there was a 750-ms pause before the onset of the motion. Between intervals, there was a 1000-ms pause, followed by a second auditory signal indicating the commencement of the second interval. After the second interval, the participants responded via the button box.

For the visual alone and visual–vestibular conditions, the limited lifetime Gaussian starfield appeared and remained static for 750 ms before the onset of the motion. In the vestibular alone condition, there was a 750-ms pause before the onset of the motion to ensure that the duration of all the conditions was the same.

In the vestibular alone and visual–vestibular conditions, after responding, the participants were passively moved back to the start position at a constant, sub-threshold velocity of  $0.025 \text{ m s}^{-1}$ , for about 5 s in order to commence the next trial.

## Conditions

All participants performed the task unimodally with only visual optic flow alone (Vis) and only vestibular information alone (Vest), and bimodally with combined visual–vestibular information (Vis–Vest).

## Pre-study (congruent)

Prior to the main experiment, five of the eight participants were run in two sessions in which they performed six experimental blocks, two Vis, two Vest, and two Vis–Vest blocks. In the visual condition, the standard heading was  $0^\circ$ , and in the combined condition, there was no conflict between the visual and vestibular information. These pre-study trials were used to familiarize the participants with the stimuli and the novel setup and to measure baseline values.

## Main study (incongruent)

In the main experiment, all participants completed three Vest blocks, twelve Vis blocks (comprised of three blocks for each of the four standard headings), and twelve

Vis–Vest blocks (comprised of three blocks for each of the four conflict levels). Each block had 72 trials, consisting of 8 standard and comparison headings repeated 9 times, which lasted approximately 20 min. In total, participants completed all 27 blocks of trials over ten 1.5-h sessions. In the Vis–Vest blocks, the conflict was counterbalanced such that  $\pm\Delta$  was presented in a pseudorandom fashion within a block.

The order of the blocks were pseudorandomized; each session consisting of three blocks. In the post-experiment debriefing, only one naïve participant stated that they were aware of a conflict in the Vis–Vest condition.

## Data analysis

Proportion of rightward judgments made by participants was plotted as a function of stimulus heading angle and cumulative Gaussian psychometric functions were fitted to the data using the *psignifit* toolbox (Wichmann & Hill, 2001a, 2001b; Figure 3).

From the fit, the point of subjective equality (PSE) and the just noticeable difference (JND)

$$\text{JND} = \sqrt{2}\sigma, \quad (2)$$

were calculated. The JND value is inversely proportional to cue reliability, and thus, the higher the JND, the less reliable the cue. The JNDs and weights were submitted to a within-subjects one-way analysis of variance (ANOVA). For all analyses, the type-I error rate was set at 0.05.

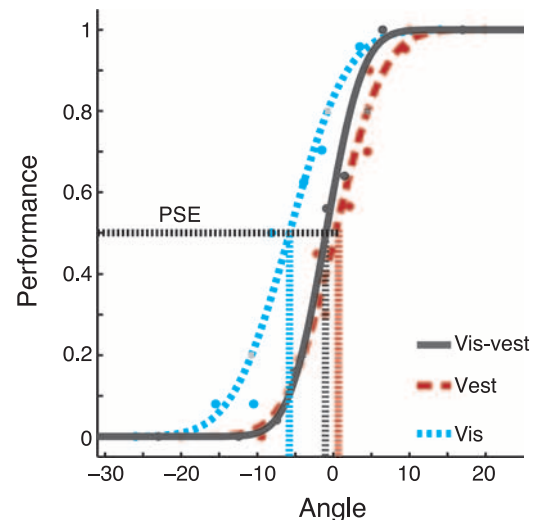


Figure 3. Data for one representative participant for Vis alone ( $\Theta = -6^\circ$ ; blue), Vest alone (red), and Vis–Vest ( $\Delta = -6^\circ$ ; gray) conditions. The data show the proportion of rightward judgments as a function of heading angle. Curved lines represent cumulative Gaussian functions fitted to the data.

## Maximum likelihood estimation (MLE)

We use the JND and PSE to estimate the likelihood distribution of both the unimodal cues, Vis,  $\hat{S}_{\text{Vis}}$  and Vest,  $\hat{S}_{\text{Vest}}$ , and the bimodal Vis–Vest cues. If visual and vestibular information combine in a statistically optimal fashion, we can predict the visual–vestibular likelihood,  $\hat{S}_{\text{Vis–Vest}}$ , using the a linear summation of the unimodal cues

$$\hat{S}_{\text{Vis–Vest}} = w_{\text{Vis}} \hat{S}_{\text{Vis}} + w_{\text{Vest}} \hat{S}_{\text{Vest}}, \quad (3)$$

where  $w_{\text{Vis}}$  and  $w_{\text{Vest}}$  are the weights predicted by the reliability of the unimodal cues,

$$w_{\text{Vis}} = \frac{1/\text{JND}_{\text{Vis}}^2}{1/\text{JND}_{\text{Vis}}^2 + 1/\text{JND}_{\text{Vest}}^2}, \quad (4)$$

$$w_{\text{Vest}} = 1 - w_{\text{Vis}}, \quad (5)$$

where  $\text{JND}_{\text{Vis}}$  and  $\text{JND}_{\text{Vest}}$  are the JNDs of the unimodal cues, Vis and Vest, respectively. The observed weights can be calculated using the observed PSE for the unimodal and bimodal conditions

$$w_{\text{Vis}} = \frac{\text{PSE}_{\text{Vis–Vest}} - \text{PSE}_{\text{Vest}}}{\text{PSE}_{\text{Vis}} - \text{PSE}_{\text{Vest}}}, \quad (6)$$

and

$$w_{\text{Vest}} = \frac{\text{PSE}_{\text{Vis–Vest}} - \text{PSE}_{\text{Vis}}}{\text{PSE}_{\text{Vest}} - \text{PSE}_{\text{Vis}}}. \quad (7)$$

Thus, we can compare the predicted weights, Equation 5, with the observed weights, Equation 6.

From Equation 1, we can also predict the JND of the Vis–Vest condition

$$\text{JND}_{\text{Vis–Vest}}^2 = \frac{\text{JND}_{\text{Vis}}^2 \text{JND}_{\text{Vest}}^2}{\text{JND}_{\text{Vis}}^2 + \text{JND}_{\text{Vest}}^2}, \quad (8)$$

where  $\text{JND}_{\text{Vis–Vest}}$  is the combined cue JND. From the Vis–Vest condition, we can extract the observed JND and compare it with the predicted JND (Equation 3). One last prediction is that the combined  $\text{JND}_{\text{Vis–Vest}}$  should be less than or equal to the lowest unimodal JND

$$\text{JND}_{\text{Vis–Vest}} \leq \min(\text{JND}_{\text{Vis}}, \text{JND}_{\text{Vest}}). \quad (9)$$

We first determined the unimodal reliabilities in order to make predictions about the combined cue conditions using Maximum Likelihood Estimation. We then used the results from the visual–vestibular spatial conflict trials to calculate the observed weights for the individual sensory cues. Finally, we compared the observed and predicted values.

## Pre-study results (congruent conditions)

In the pre-study trials, the average and standard error unimodal Vest and Vis thresholds were  $\text{JND}_{\text{Vest}} = 7.9^\circ \pm 1.1^\circ$  and  $\text{JND}_{\text{Vis}} = 6.5^\circ \pm 0.6^\circ$ , respectively. The combined threshold was  $\text{JND}_{\text{Vis–Vest}} = 4.63^\circ \pm 0.6^\circ$ . The JND values for the unimodal conditions and the bimodal condition were submitted to a one-way ANOVA. The analysis revealed that there was a significant difference between the JNDs ( $F(2,12) = 3.88$ ,  $\text{MSE} = 3.404$ ,  $p < 0.05$ ). Post-hoc two-tailed  $t$ -tests comparing the two unimodal and the bimodal conditions revealed that the Vis condition was statistically different from the Vis–Vest condition ( $p < 0.05$ ). The difference between the Vest condition and the Vis–Vest condition approached significance ( $p = 0.07$ ). The observed and predicted Vis–Vest JNDs were submitted to a two-tailed  $t$ -test, which revealed no statistical difference ( $p = 0.84$ ).

Therefore, the pre-study result examining unimodal and bimodal conditions suggests that the combination of visual and vestibular information yielded an optimal reduction of the JND consistent with predictions based on MLE.

## Unimodal conditions

The average unimodal vestibular threshold ( $\text{JND}_{\text{Vest}}$ ) and  $\text{PSE}_{\text{Vest}}$  were  $\Theta = 0^\circ$ :  $7.2^\circ \pm 1.08^\circ$  and  $0.1^\circ \pm 0.26^\circ$  (see Figure 4). The average unimodal visual thresholds ( $\text{JND}_{\text{Vis}}$ ) for each of the standard offsets (from straight ahead;  $\Theta = 0^\circ$ ) were  $\Theta = -10^\circ$ :  $6.4^\circ \pm 0.5^\circ$ ,  $\Theta = -6^\circ$ :  $5.8^\circ \pm 0.45^\circ$ ,  $\Theta = 6^\circ$ :  $5.8^\circ \pm 0.43^\circ$ , and  $\Theta = 10^\circ$ :  $7.6^\circ \pm 0.79^\circ$ . The average unimodal  $\text{PSE}_{\text{Vis}}$  values for each of the standard offsets were  $\Theta = -10^\circ$ :  $-9.9^\circ \pm 0.2^\circ$ ,  $\Theta = -6^\circ$ :  $-5.9^\circ \pm 0.2^\circ$ ,  $\Theta = 6^\circ$ :  $5.4^\circ \pm 0.4^\circ$ , and  $\Theta = 10^\circ$ :  $9.5^\circ \pm 0.3^\circ$ .

The four visual alone conditions (i.e., four standard heading angles) and the vestibular alone condition were submitted to a one-way ANOVA. The analysis revealed that there was no significant difference between the JNDs across any of the unimodal conditions ( $F(4,28) = 1.530$ ,  $\text{MSE} = 3.3$ ,  $p = 0.22$ ).

## Visual–vestibular combined conditions

The average observed JND values for the visual–vestibular combined condition ( $\text{JND}_{\text{Vis–Vest}}$ ) for the four different spatial offsets were  $\Delta = -10^\circ$ :  $4.51^\circ \pm 0.31^\circ$ ,  $\Delta = -6^\circ$ :  $4.35^\circ \pm 0.37^\circ$ ,  $\Delta = 6^\circ$ :  $4.4^\circ \pm 0.38^\circ$ , and  $\Delta = 10^\circ$ :  $4.8^\circ \pm 0.4^\circ$ . In these cases, the standard heading was an

## Results

In this study, we systematically investigated the weights of visual and vestibular cues during heading estimation.

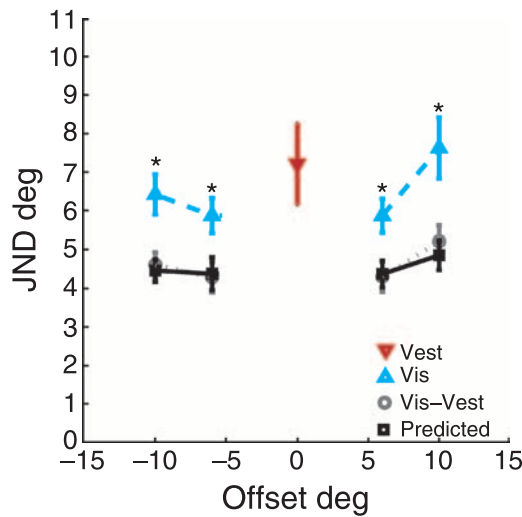


Figure 4. Unimodal and bimodal averaged JNDs. Error bars represent standard error of the mean.

incongruent stimulus consisting of a straight-ahead ( $\Theta = 0^\circ$ ) vestibular cue with a visual cue offset by  $\Delta = \pm 6^\circ$  or  $\pm 10^\circ$ . The average observed  $PSE_{Vis-Vest}$  for the four different spatial offsets were  $\Delta = -10^\circ$ :  $-3.25^\circ \pm 0.53^\circ$ ,  $\Delta = -6^\circ$ :  $1.7^\circ \pm 0.25^\circ$ ,  $\Delta = 6^\circ$ :  $1.6^\circ \pm 0.21^\circ$ , and  $\Delta = 10^\circ$ :  $3.5^\circ \pm 0.54^\circ$ .

The predicted JNDs were calculated from the unimodal JNDs using Equation 8. A 2 condition (Vis and Vis-Vest)  $\times$  4 offset ( $\Delta = -10^\circ$ ,  $\Delta = -6^\circ$ ,  $\Delta = 6^\circ$ ,  $\Delta = 10^\circ$ ) repeated-measures ANOVA was performed. The analysis revealed that there was a significant main effect of condition ( $F(1,7) = 54.276$ ,  $MSE = 1.737$ ,  $p < 0.002$ ) but no significant main effect of offset ( $F(3,21) = 3.904$ ,  $MSE = 1.628$ ,  $p = 0.187$ ) and no significant interaction effect ( $F(1,7) = 0.590$ ,  $MSE = 1.056$ ,  $p = 0.629$ ). It should also be noted that there was a significant quadratic effect of offset ( $F(1,7) = 7.872$ ,  $MSE = 1.595$ ,  $p < 0.05$ ), which was expected, as the reliability of the visual cue decreased as a function of the eccentricity of the heading offset (Crowell & Banks, 1993). Post-hoc  $t$ -tests comparing the visual alone and visual-vestibular conditions for each individual offset showed significant differences at all offset values ( $\Delta = -10^\circ$ :  $p < 0.05$ ,  $\Delta = -6^\circ$ :  $p < 0.01$ ,  $\Delta = 6^\circ$ :  $p < 0.01$ ,  $\Delta = 10^\circ$ :  $p < 0.01$ ).

In order to compare the calculated predicted values with the actual observed values, we conducted a 2 (observed Vis-Vest vs. predicted Vis-Vest)  $\times$  4 (offset;  $\Delta = -10^\circ$ ,  $\Delta = -6^\circ$ ,  $\Delta = 6^\circ$ ,  $\Delta = 10^\circ$ ) repeated-measures ANOVA, which revealed no significant differences ( $F(1,7) = 0.886$ ,  $MSE = 0.137$ ,  $p = 0.378$ ). In Figure 5, individual participant's observed JNDs for all conflict levels are plotted as a function of the predicted JNDs. This figure nicely illustrates that each individual's observed JNDs are consistent with what would be predicted given an optimal combination of unimodal cues.

Overall, the results demonstrate that heading discrimination performance improves when visual and vestibular

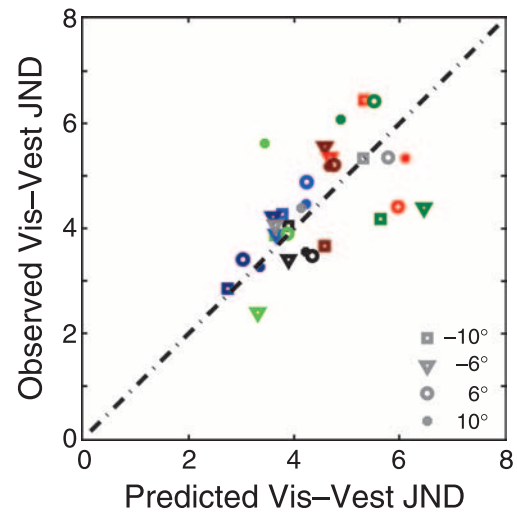


Figure 5. Scatter plot of observed Vis-Vest JNDs versus Vis-Vest predicted JNDs. Different symbols represent different conflict levels. Different colors represent individual participants. The dashed line indicates the ideal data.

stimuli are presented together. Furthermore, the discrimination improvement in the bimodal condition was predicted by the unimodal reliabilities using Maximum Likelihood Estimation.

## Weights

The observed weights were determined by introducing a conflict between the heading direction specified by vision and that specified by vestibular information (Equation 6). Assuming that the combination of senses is linear, we predicted the visual weights using Equation 4. The average predicted and observed visual weights for each of the four conflict levels are illustrated in Figure 6.

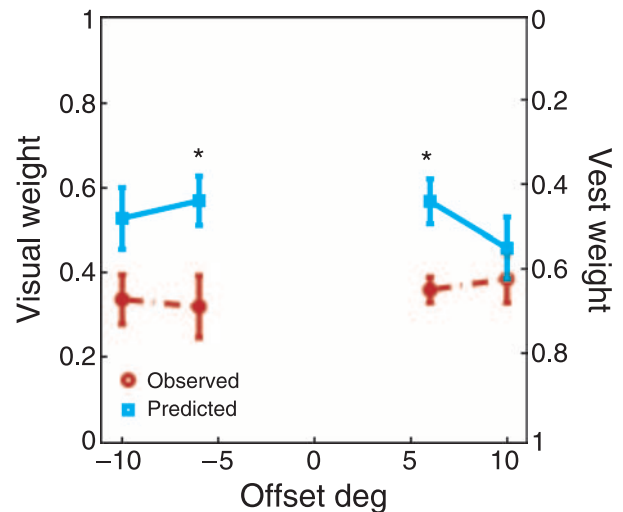


Figure 6. Predicted (blue squares) and observed (red circles) visual weights.

A 2 (observed vs. predicted)  $\times$  4 (offsets;  $\Delta = -10^\circ$ ,  $\Delta = -6^\circ$ ,  $\Delta = 6^\circ$ ,  $\Delta = 10^\circ$ ) repeated-measures ANOVA was performed on the visual weights. The analysis revealed no significant main effect of offset ( $F(3,21) = 0.402$ ,  $\text{MSE} = 0.013$ ,  $p = 0.753$ ), but a significant main effect was observed when comparing the actual and predicted visual weights ( $F(1,7) = 5.823$ ,  $\text{MSE} = 0.047$ ,  $p < 0.05$ ). There was also a significant interaction effect observed ( $F(3,21) = 3.550$ ,  $\text{MSE} = 0.007$ ,  $p < 0.05$ ). The interaction effect is a result of the quadratic interaction that was significant ( $F(1,7) = 7.847$ ,  $\text{MSE} = 0.005$ ,  $p < 0.05$ ). We conducted two-tailed paired-sample  $t$ -tests comparing the predicted and observed visual weights for each of the four offset levels separately. Two of the four  $t$ -tests revealed significant differences ( $\Delta = -10^\circ$ :  $p = 0.06$ ,  $\Delta = -6^\circ$ :  $p < 0.05$ ,  $\Delta = 6^\circ$ :  $p < 0.05$ ,  $\Delta = 10^\circ$ :  $p = 0.5$ ).

These results show that the observed weights are biased toward the vestibular cue and that these weights are not predicted by a linear combination of the two senses based on the unisensory responses. In [Appendix A](#), we briefly describe a simple Bayesian model that can partially account for the offset in the predicted weights through the inclusion of a prior.

## General discussion

The current experiment sought to more precisely quantify the relative weights of optic flow and vestibular cues during egocentric heading estimation. The results support previous findings that demonstrated a statistically optimal combination of visual and vestibular information, such that the reliabilities of the two unimodal conditions accurately predicted the reduction in variance observed in the combined cue conditions (Fetsch et al., 2009a; Gu et al., 2008). However, when a spatial conflict was introduced as a way of quantifying the specific cue-weighting values, results showed that vestibular cues were consistently weighted higher, despite the fact that the reliabilities observed in the unimodal vestibular condition were generally lower (see also Fetsch et al., 2009a). In an effort to address this, we propose a simple, preliminary, Bayesian model in which we introduced a prior to account for the bias toward the vestibular cue. This prior was able to account for the higher weighting of the vestibular cues but at the expense of accurately predicting the JNDs (see [Appendix A](#)).

Considering that a large emphasis of past behavioral and neurophysiological research on heading perception has focused mainly on visual inputs, the higher weighting of vestibular information is particularly interesting. It is clear, however, that vision also contributed significantly to heading discrimination in the combined cue conditions, evidenced by the reduction in variance that was consistently observed even during relatively large spatial offsets

(i.e.,  $\pm 10^\circ$ ). This indicates that optimal visual–vestibular cue integration is quite robust even during cue-conflict conditions. Furthermore, the quadratic nature of the visual alone and combined JNDs indicates that visual eccentricities play a role in the way in which the cues are combined (i.e., at larger eccentricities the change in reliability of the visual cues is taken into account in the combination of cues).

Some previous experiments evaluating multisensory heading perception using pointing tasks have, in fact, reported a visual dominance (Ohmi, 1996; Telford et al., 1995). For instance, Ohmi (1996) reported that unisensory vestibular conditions exhibited the lowest accuracy and highest variance of any other unimodal or bimodal condition. Further, combining visual with vestibular information did not significantly improve accuracy or precision above unisensory visual conditions. When visual and vestibular information provided conflicting directional information, visual cues were weighted higher for smaller directional conflicts ( $30^\circ$  and  $150^\circ$  offsets) and vestibular cues were weighted higher for the largest conflict ( $180^\circ$  offset). Even the smallest conflict introduced by Ohmi (1996) was three times as large as the largest conflict introduced in the current study. Conflicts of the size used in Ohmi (1996) are likely highly detectable by the participant, and therefore, the two cues may not be interpreted as arising from a single source but rather as signaling two separate events (Körding et al., 2007). This would likely have a strong impact on how the two cues are combined in the perception of heading.

Other studies using more sensitive psychophysical measures have in fact reported optimal visual–vestibular integration during very similar heading discrimination tasks as the one described here (Fetsch et al., 2009a; Gu et al., 2008). Interestingly, the unexpected vestibular bias observed in the current study is consistent with that recently reported by Fetsch et al. (2009a), in which they observed this effect in both humans and non-human primates. Continued work is being conducted to determine the neurophysiological underpinnings of these effects (Fetsch, Turner, DeAngelis, & Angelaki, 2009b; Gu et al., 2010).

Due to the inconsistencies observed in these few studies, it is important to reconcile these differences and also to better define what the limits of visual–vestibular integration are when discrepant information is provided by each cue. In line with this, recent work by our group has explicitly tested whether visual and vestibular cues are optimally integrated under different types of conflicts, aside from spatially defined discrepancies. These results have shown that visual–vestibular integration is observed, to some extent, when a temporal offset is introduced between the onset of the visual motion and the onset of the vestibular motion (Campos, Butler, & Bühlhoff, 2009), and also when the visual acceleration profile differs from the vestibular motion profile (Butler, Campos, & Bühlhoff, 2008). Further,



evidence indicates that relative cue weighting appears to change dynamically over trials as a function of increased exposure to discrepant cue information (Campos et al., 2009). Overall, these findings are somewhat different from other categories of cue combination in which relatively large temporal and spatial offsets often do not lead to optimal cue combination (e.g., Gepshtein, Burge, Ernst, & Banks, 2005; Roach, Heron, & McGraw, 2006).

The higher weighting of vestibular information observed here could be the result of several different factors. For instance, the optic flow stimuli used here was itself sufficient for estimating heading and resulted in more precise discrimination capabilities than the unisensory vestibular condition. However, it is possible that, in order to be weighted higher when combined with vestibular information, certain characteristics of dynamic visual information are necessary. For instance, in a previous experiment, we have shown that optic flow must be presented stereoscopically in order for optimal cue integration to be observed during combined visual–vestibular conditions (Butler et al., [submitted](#); although see Fetsch et al., 2009a). Importantly, in that study, the psychophysical data obtained during the unisensory visual conditions did *not* differ as a function of whether optic flow was presented in stereo, but rather it was only during the combined condition that this effect exhibited itself. Others have also described an important role for stereovision in the perception of optic flow (e.g., Lappe & Grigo, 1999; van den Berg & Brenner, 1994). While the visual stimulus used in the current study was, in fact, in stereo, there could be other characteristics of the visual display that might lead to a higher weighting of visual information during cue integration. Considering that much of the claims regarding the capacity for observers to use optic flow alone to estimate heading has relied on computer-simulated, random dot flow fields, it will now be important to define precisely which characteristics of dynamic visual stimuli contribute to optimal cue integration during more natural, multimodal scenarios.

As in most experiments evaluating relative cue weighting in the context of self-motion perception, there is the inherent concern that the visual alone condition in this experiment introduced a sensory conflict. Specifically, in this condition, the vestibular system is specifying a stationary position, while the visual system detects self-motion. Therefore, perhaps the visual alone condition is not entirely representative as a unisensory comparison. We do not believe that this is a strong concern here. Support for this is provided by the fact that Gu et al. (2007) demonstrated that in non-human primates there was no difference in visual heading estimation between labyrinthectomized animals and animals with fully functioning vestibular systems. In other words, the removal of this potentially “conflicting” vestibular information did not seem to change visual heading estimates.

There has also been a great deal of discussion in the literature regarding the interpretation of optic flow in the

presence of eye movements (e.g., Royden et al., 1992; Warren & Hannon, 1990; Warren et al., 1988). Because retinal flow is generated by a combination of the optic flow that occurs during head and eye movements, in order to use this information to determine heading, one must cancel out particular components of the flow field. Because, in the current experiment, observers did not fixate during the movement, it might seem possible that the optic flow presented by the forward motion of the display image was compounded by eye movements, thus leading to increased processing requirements. That said, Fetsch et al. (2009a) used a very similar paradigm as the one used here, except that they required participants to fixate during presented movements and yet they found a similar overweighting of vestibular information. Further, Gu et al. (2007) explicitly tested whether fixating during a heading task affected performance measures on a psychophysical task and demonstrated that they did not. Other studies have also indicated that humans appear to be able to effectively account for the different sources of retinal flow using other non-visual, oculomotor cues (e.g., Royden et al., 1992), and therefore, overall, this is likely not the reason for the vestibular dominance seen here.

The conflicts used in this study, while small relative to several other studies (e.g., offsets of 30°–150° used by Ohmi, 1996), were still relatively large considering that the highest unimodal JND (i.e., approximately 7°) was lower than the largest conflict levels of  $\pm 10^\circ$ . Despite the fact that the participants did not report being consciously aware of the discrepancies, perhaps the perceptual awareness of the conflict changed relative cue weighting. Berger and Bühlhoff (2009) recently evaluated the effects of attention and of having an awareness of a cue conflict on visual–vestibular cue weighting during a rotation reproduction task. They showed that attended cues were weighted higher only when a cue conflict was detected. However, in the current experiment, there was no evidence that the relative cue weighting changed with greater conflicts. Therefore, either the conflict was not consciously detected in this case, or attentional factors directed toward the physical movements during conflict scenarios are not likely the reason for the higher weighting of vestibular cues.

In general, the results observed in the current study suggest that the main non-visual cue contributing to passive heading perception is vestibular. However, there are other sources of sensory and motor information that are available during passive movements in a vehicle or motion simulator, including auditory cues, proprioceptive information provided by muscles in the neck, and somatosensory information caused by changes in body pressure during accelerations, decelerations, and vibrations (Seidman, 2008). In our case, we used vibrational masking to avoid providing somatosensory information from the movement of the platform and we also required that participants wear noise-canceling headphones that presented sound masking. Gu et al. (2007) also showed that monkeys having

undergone a labyrinthectomy were no longer able to perform passive heading detection in the absence of vision. These results, therefore, strongly implicate the vestibular organs—the otoliths in particular—as being the critical source of non-visual information in this task. Further studies will be required to parse out the contributions of additional sources of inertial information.

There are important applied questions in which a greater understanding of the interactions between visual and vestibular information could be quite revealing. For instance, there is some evidence to suggest that relative cue weighting changes during aging. This is either due to sensory deficiencies (e.g., loss of vestibular function) or changes in cue integration. For instance, Berard, Fung, McFadyen, and Lamontagne (2009) asked older and younger participants to walk along straight paths while the visual focus of expansion presented in a head-mounted display was manipulated. They reported that younger participants directed their walking in accordance with the direction of the visuals, while older adults tended to direct their walking using a body-based frame of reference. Further, Mapstone, Logan, and Duffy (2006) showed that older adults and those with Alzheimer’s disease demonstrated deficits in the perception of optic flow during object motion and simulated self-motion. However, Chou et al. (2009) demonstrated that changing the properties of an optic flow stimulus (i.e., speed and forward/backward direction) during walking did not differentially affect locomotor parameters (e.g., walking speed and stride length) in older and younger participants. Therefore, additional research is needed to more precisely understand multi-modal self-motion perception during aging. Because these potential age-related changes related to visual–vestibular integration could have serious implications with respect to incidences of accidents such as falls and motor vehicle collisions in the elderly, this is an important applied question to be addressed in the future.

## Conclusions

In this study, we found that even when there were relatively large spatial conflicts between the egocentric heading directions specified by optic flow and that specified by vestibular cues, participants exhibited a statistically optimal reduction of variance under combined cue conditions. However, performances in the unimodal conditions did not predict the weights in the combined cue conditions. Therefore, we conclude that visual and vestibular cue integration is not predicted solely by the reliability of each individual cue but rather there is a higher weighting of vestibular information in this task. Future work and modeling will help define the cause of the higher vestibular weighting and reveal the mechanisms underlying visual and vestibular heading perception.

## Appendix A

### Supplementary materials

#### Bayesian model

The MLE model described in the [Results](#) section nicely predicts the appropriate reduction in variance in the combined cue condition based on the variances in each of the unisensory conditions. However, when using the estimates in the conflict trial to predict relative cue weighting, a systematic higher weighting of vestibular cues was observed. This bias is not unlike other similar biases that have been reported in other types of sensory-integration experiments. Specifically, several studies have now shown that, in some situations, the variance and weights are not purely predicted by a linear weighting of each cue (Battaglia, Jacobs, & Aslin, 2003; Roach et al., 2006) or demonstrate a breakdown of optimal integration (Gepshtein et al., 2005; Körding et al., 2007; Wallace et al., 2004). In these cases, a prior is introduced into the MLE model.

The MLE model (Equation 3) is expanded upon by the introduction of a prior distribution  $\hat{S}_{\text{Prior}}$

$$\hat{S}_{\text{Vis-Vest}} = w_{\text{Vis}} \hat{S}_{\text{Vis}} + w_{\text{Vest}} \hat{S}_{\text{Vest}} + w_{\text{Prior}} \hat{S}_{\text{Prior}}, \quad (\text{A1})$$

where  $w_{\text{Vis}}$ ,  $w_{\text{Vest}}$ , and  $w_{\text{Prior}}$  are the weights (Körding et al., 2007). From this, we get the formula for the Bayes predicted visual weights

$$w_{\text{Vis}} = \frac{1/\text{JND}_{\text{Vis}}^2}{1/\text{JND}_{\text{Vis}}^2 + 1/\text{JND}_{\text{Vest}}^2 + 1/\text{JND}_{\text{Bayes-Prior}}^2}, \quad (\text{A2})$$

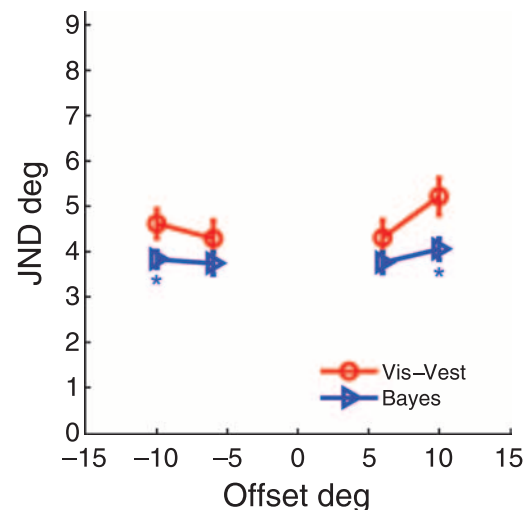


Figure A1. Observed JNDs (red circles) and predicted JNDs from the Bayes model (blue triangles).

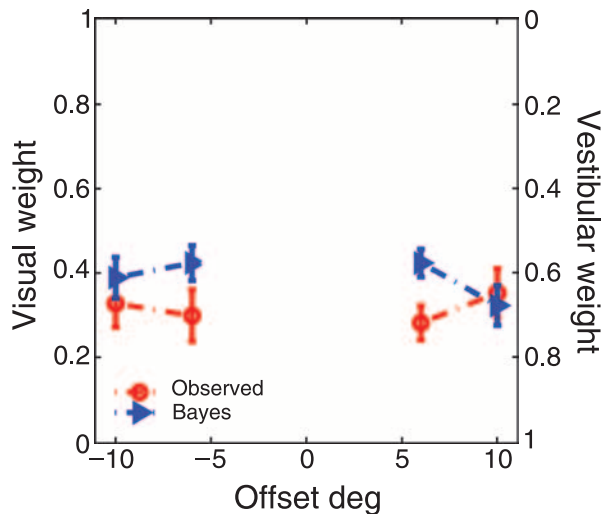


Figure A2. Observed visual weights (red circles) and predicted weights from the Bayes model (blue triangles).

and for the Bayes predicted JND, we have

$$\text{JND}_{\text{Vis-Vest}}^2 = \frac{1}{1/\text{JND}_{\text{Vis}}^2 + 1/\text{JND}_{\text{Vest}}^2 + 1/\text{JND}_{\text{Bayes-Prior}}^2}. \quad (\text{A3})$$

As a preliminary approach, we have implemented a simple Bayesian prior model, which has a Gaussian distribution prior where the parameters  $\text{JND}_{\text{Bayes-Prior}}$  and  $\text{PSE}_{\text{Bayes-Prior}}$  define the reliability and mean of the prior. The parameters of the prior distribution were fit using a least-squares method to the whole data set, therefore fitting two parameters for all participants. The average predicted and observed JNDs and weights for this model are illustrated for the four conflict levels in Figures A1 and A2. The fitted parameters for the priors were  $\text{JND}_{\text{Bayes-Prior}} = 7.21^\circ$  and  $\text{PSE}_{\text{Bayes-Prior}} = 0.32^\circ$ .

A 2 (observed vs. Bayes model prediction)  $\times$  4 (offsets;  $\Delta = -10^\circ$ ,  $\Delta = -6^\circ$ ,  $\Delta = 6^\circ$ ,  $\Delta = 10^\circ$ ) repeated-measures ANOVA was performed on the JNDs. The analysis revealed no significant main effect of offset ( $F(3,21) = 1.336$ ,  $\text{MSE} = 0.464$ ,  $p = 0.06$ ), but a significant main effect was observed when comparing the actual and predicted JNDs ( $F(1,7) = 9.312$ ,  $\text{MSE} = 0.141$ ,  $p < 0.05$ ). We conducted two-tailed paired-sample  $t$ -tests comparing the predicted and observed JNDs for each of the four offset levels separately. Two of the four  $t$ -tests revealed significant differences ( $\Delta = -10^\circ$ :  $p < 0.05$ ,  $\Delta = -6^\circ$ :  $p = 0.1$ ,  $\Delta = 6^\circ$ :  $p = 0.09$ ,  $\Delta = 10^\circ$ :  $p < 0.05$ ).

A 2 (observed vs. Bayes model prediction)  $\times$  4 (offsets;  $\Delta = -10^\circ$ ,  $\Delta = -6^\circ$ ,  $\Delta = 6^\circ$ ,  $\Delta = 10^\circ$ ) repeated-measures ANOVA was performed on the weights. The analysis revealed no significant main effect of offset ( $F(3,21) = 0.242$ ,  $\text{MSE} = 0.007$ ,  $p = 0.866$ ) and no significant main effect when comparing the actual and predicted JNDs ( $F(1,7) =$

$3.048$ ,  $\text{MSE} = 0.029$ ,  $p = 0.124$ ). Therefore, the Bayesian model provided a good fit with respect to the weights but not with respect to the JNDs. Specifically, the predicted JNDs were statistically different from the observed JNDs for the two largest conflict conditions ( $\Delta = \pm 10^\circ$ ).

## Acknowledgments

This research was supported by the Max Planck Society and an Enterprise Ireland International Collaboration Grant and by the World Class University (WCU) program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (R31-2008-000-10008-0). We would like to thank Julian Hofmann, Michael Weyel, and the participants for help with the data collection. We are grateful to Edel Flynn, Daniel Berger, John Foxe, Marc Ernst, Martin Banks, and the reviewers for their helpful suggestions, advice, and support.

Commercial relationships: none.

Corresponding authors: John S. Butler and Heinrich H. Bühlhoff.

Emails: john.butler@einstein.yu.edu; heinrich.buelthoff@tuebingen.mpg.de.

Address: Max Planck Institute for Biological Cybernetics, Tübingen, Germany.

## References

- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*, 257–262.
- Angelaki, D. E., & Cullen, K. E. (2008). Vestibular system: The many facets of a multimodal sense. *Annual Review of Neuroscience*, *31*, 125–150.
- Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, *20*, 1391–1397.
- Benson, A. J., Spencer, M. B., & Stott, J. R. (1986). Thresholds for the detection of the direction of whole-body, linear movement in the horizontal plane. *Aviation Space & Environment Medicine*, *57*, 1088–1096.
- Berard, J. R., Fung, J., McFadyen, B. J., & Lamontagne, A. (2009). Aging affects the ability to use optic flow in the control of heading during locomotion. *Experimental Brain Research*, *194*, 183–190.
- Berger, D. R., & Bühlhoff, H. H. (2009). The role of attention on the integration of visual and inertial cues. *Experimental Brain Research*, *198*, 287–300.



- Bertin, R. J., & Berthoz, A. (2004). Visuo-vestibular interaction in the reconstruction of travelled trajectories. *Experimental Brain Research*, 154, 11–21.
- Britten, K. H., & van Wezel, R. J. (1998). Electrical microstimulation of cortical area MST biases heading perception in monkeys. *Nature Neuroscience*, 1, 59–63.
- Britten, K. H., & van Wezel, R. J. (2002). Area MST and heading perception in macaque monkeys. *Cerebral Cortex*, 12, 692–701.
- Bülthoff, H. H., Little, J., & Poggio, T. (1989). A parallel algorithm for real-time computation of optical flow. *Nature*, 337, 549–555.
- Bülthoff, H. H., & Yuille, A. (1991). Bayesian models for seeing shapes and depth. *Comments on Theoretical Biology*, 2, 283–314.
- Butler, J. S., Campos, J. L., & Bühlhoff, H. H. (2008). The robust nature of visual–vestibular combination for heading. *Perception*, 37 (ECVP Abstract Supplement), 40.
- Butler, J. S., Campos, J. L., Bühlhoff, H. H., & Smith, S. T. (submitted). Visual and Vestibular cue combination for perception of heading: The role of stereo.
- Butler, J. S., Smith, S. T., Beykirch, K., & Bühlhoff, H. H. (2006). Visual vestibular interactions for self motion estimation. *Proceedings of the Driving Simulation Conference Europe (DSC Europe 2006)*, 1, 1–10 (10 2006).
- Campos, J. L., Butler, J. S., & Bühlhoff, H. H. (2009). Visual–vestibular cue combination during temporal asynchrony. *IMRF*, 10, 198.
- Chou, Y. H., Wagenaar, R. C., Saltzman, E., Giphart, J. E., Young, D., Davidsdottir, R., et al. (2009). Effects of optic flow speed and lateral flow asymmetry on locomotion in younger and older adults: A virtual reality study. *Journals of Gerontology B: Psychological Sciences and Social Sciences*, 64, 222–231.
- Crowell, J. A., & Banks, M. S. (1993). Perceiving heading with different retinal regions and types of optic flow. *Perception & Psychophysics*, 53, 325–337.
- Duffy, C. J., & Wurtz, R. H. (1991). Sensitivity of MST neurons to optic flow stimuli: I. A continuum of response selectivity to large-field stimuli. *Journal of Neurophysiology*, 65, 1329–1345.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415, 429–433.
- Ernst, M. O., Banks, M. S., & Bühlhoff, H. H. (2000). Touch can change visual slant perception. *Nature Neuroscience*, 3, 69–73.
- Fetsch, C. R., Turner, A. H., DeAngelis, G. C., & Angelaki, D. E. (2009a). Dynamic reweighting of visual and vestibular cues during self-motion perception. *Journal of Neuroscience*, 29, 15601–15612.
- Fetsch, C. R., Turner, A. H., DeAngelis, G. C., & Angelaki, D. E. (2009b). Reliability-based cue re-weighting in rhesus monkeys: Behavior and neural correlates. *IMRF*, 10, 346.
- Frenz, H., & Lappe, M. (2005). Absolute travel distance from optic flow. *Vision Research*, 45, 1679–1692.
- Frenz, H., & Lappe, M. (2006). Visual distance estimation in static compared to moving virtual scenes. *Spanish Journal of Psychology*, 9, 321–331.
- Gepshtein, S., Burge, J., Ernst, M. O., & Banks, M. S. (2005). The combination of vision and touch depends on spatial proximity. *Journal of Vision*, 5(11):7, 1013–1023, <http://www.journalofvision.org/content/5/11/7>, doi:10.1167/5.11.7. [PubMed] [Article]
- Gibson, J. J. (1950). *The perception of the visual world*. Boston: Houghton Mifflin.
- Gu, Y., Angelaki, D. E., & Deangelis, G. C. (2008). Neural correlates of multisensory cue integration in macaque MSTd. *Nature Neuroscience*, 11, 1201–1210.
- Gu, Y., DeAngelis, G. C., & Angelaki, D. E. (2007). A functional link between area MSTd and heading perception based on vestibular signals. *Nature Neuroscience*, 10, 1038–1047.
- Gu, Y., Fetsch, C. R., Adeyemo, B., DeAngelis, G. C., & Angelaki, D. E. (2010). Decoding of MSTd population activity accounts for variations in the precision of heading perception. *Neuron*, 66, 596–609.
- Gu, Y., Watkins, P. V., Angelaki, D. E., & DeAngelis, G. C. (2006). Visual and nonvisual contributions to three-dimensional heading selectivity in the medial superior temporal area. *Journal of Neuroscience*, 26, 73–85.
- Harris, L. R., Jenkin, M., & Zikovitz, D. C. (2000). Visual and non-visual cues in the perception of linear self-motion. *Experimental Brain Research*, 135, 12–21.
- Heuer, H. W., & Britten, K. H. (2004). Optic flow signals in extrastriate area MST: Comparison of perceptual and neuronal sensitivity. *Journal of Neurophysiology*, 91, 1314–1326.
- Israel, I., Chapuis, N., Glasauer, S., Charade, O., & Berthoz, A. (1993). Estimation of passive horizontal linear whole-body displacement in humans. *Journal of Neurophysiology*, 70, 1270–1273.
- Israel, I., Grasso, R., Georges-Francois, P., Tsuzuku, T., & Berthoz, A. (1997). Spatial memory and path integration studied by self-driven passive linear displacement: I. Basic properties. *Journal of Neurophysiology*, 77, 3180–3192.
- Jurgens, R., & Becker, W. (2006). Perception of angular displacement without landmarks: Evidence for Bayesian fusion of vestibular, optokinetic, podokinesthetic, and cognitive information. *Experimental Brain Research*, 174, 528–543.
- Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS One*, 2, e943.



- Lappe, M., Bremmer, F., Pökel, M., Thiele, A., & Hoffmann, K. P. (1996). Optic flow processing in monkey STS: A theoretical and experimental approach. *Journal of Neuroscience*, 16, 6265–6285.
- Lappe, M., Bremmer, F., & van den Berg, A. V. (1999). Perception of self-motion from visual flow. *Trends in Cognitive Sciences*, 3, 329–336.
- Lappe, M., & Grigo, A. (1999). How stereovision interacts with optic flow perception: Neural mechanisms. *Neural Networks*, 12, 1325–1329.
- MacNeilage, P. R., Banks, M. S., Berger, D. R., & Bühlhoff, H. H. (2007). A Bayesian model of the disambiguation of gravito-inertial force by visual cues. *Experimental Brain Research*, 179, 263–290.
- Mapstone, M., Logan, D., & Duffy, C. J. (2006). Cue integration for the perception and control of self-movement in ageing and Alzheimer's disease. *Brain*, 129, 2931–2944.
- Ohmi, M. (1996). Egocentric perception through interaction among many sensory systems. *Brain Research and Cognitive Brain Research*, 5, 87–96.
- Orban, G. A., Lagae, L., Verri, A., Raiguel, S., Xiao, D., Maes, H., et al. (1992). First-order analysis of optical flow in monkey brain. *Proceedings of the National Academy of Sciences of the United States of America*, 89, 2595–2599.
- Perrone, J. A., & Stone, L. S. (1998). Emulating the visual receptive-field properties of MST neurons with a template model of heading estimation. *Journal of Neuroscience*, 18, 5958–5975.
- Redlick, F. P., Jenkin, M., & Harris, L. R. (2001). Humans can use optic flow to estimate distance of travel. *Vision Research*, 41, 213–219.
- Roach, N. W., Heron, J., & McGraw, P. V. (2006). Resolving multisensory conflict: A strategy for balancing the costs and benefits of audio-visual integration. *Proceedings of the Royal Society of London B: Biological Sciences*, 273, 2159–2168.
- Royden, C. S., Banks, M. S., & Crowell, J. A. (1992). The perception of heading during eye movements. *Nature*, 360, 583–585.
- Schaafsma, S. J., & Duysens, J. (1996). Neurons in the ventral intraparietal area of awake macaque monkey closely resemble neurons in the dorsal part of the medial superior temporal area in their responses to optic flow patterns. *Journal of Neurophysiology*, 76, 4056–4068.
- Seidman, S. H. (2008). Translational motion perception and vestibuloocular responses in the absence of non-inertial cues. *Experimental Brain Research*, 184, 13–29.
- Siegle, J. H., Campos, J. L., Mohler, B. J., Loomis, J. M., & Bühlhoff, H. H. (2009). Measurement of instantaneous perceived self-motion using continuous pointing. *Experimental Brain Research*, 195, 429–444.
- Sun, H. J., Campos, J. L., & Chan, G. S. (2004). Multi-sensory integration in the estimation of relative path length. *Experimental Brain Research*, 154, 246–254.
- Sun, H. J., Campos, J. L., Young, M., Chan, G. S., & Ellard, C. G. (2004). The contributions of static visual cues, nonvisual cues, and optic flow in distance estimation. *Perception*, 33, 49–65.
- Tanaka, K., & Saito, H. (1989). Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey. *Journal of Neurophysiology*, 62, 626–641.
- Telford, L., Howard, I. P., & Ohmi, M. (1995). Heading judgments during active and passive self-motion. *Experimental Brain Research*, 104, 502–510.
- van den Berg, A. V. (1992). Robustness of perception of heading from optic flow. *Vision Research*, 32, 1285–1296.
- van den Berg, A. V., & Brenner, E. (1994). Why two eyes are better than one for judgements of heading. *Nature*, 371, 700–702.
- Wallace, M. T., Roberson, G. E., Hairston, W. D., Stein, B. E., Vaughan, J. W., & Schirillo, J. A. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research*, 158, 252–258.
- Wallis, G., Chatziastros, A., & Bühlhoff, H. H. (2002). An unexpected role for visual feedback in vehicle steering control. *Current Biology*, 12, 295–299.
- Warren, W. H., Jr., & Hannon, D. J. (1990). Eye movements and optical flow. *Journal of the Optical Society of America A*, 7, 160–169.
- Warren, W. H., Jr., Morris, M. W., & Kalish, M. (1988). Perception of translational heading from optical flow. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 646–660.
- Wichmann, F. A., & Hill, N. J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, 63, 1293–1313.
- Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics*, 63, 1314–1329.
- Wilkie, R. M., & Wann, J. P. (2005). The role of visual and nonvisual information in the control of locomotion. *Journal of Experimental Psychology: Human Perception and Performance*, 31, 901–911.
- Yuille, A., & Bühlhoff, H. H. (1996). Bayesian decision theory and psychophysics. In D. Knill & W. Richards (Eds.), *Perception as Bayesian inference* (pp. 123–161). Cambridge, UK: Cambridge University Press.