

Self-motion leads to mandatory cue fusion across sensory modalities

Mario Prsa, Steven Gale and Olaf Blanke

J Neurophysiol 108:2282-2291, 2012. First published 25 July 2012;
doi: 10.1152/jn.00439.2012

You might find this additional info useful...

This article cites 50 articles, 11 of which you can access for free at:

<http://jn.physiology.org/content/108/8/2282.full#ref-list-1>

Updated information and services including high resolution figures, can be found at:

<http://jn.physiology.org/content/108/8/2282.full>

Additional material and information about *Journal of Neurophysiology* can be found at:

<http://www.the-aps.org/publications/jn>

This information is current as of January 14, 2013.

Self-motion leads to mandatory cue fusion across sensory modalities

Mario Prsa,^{1,2} Steven Gale,^{1,2} and Olaf Blanke^{1,2,3}

¹Laboratory of Cognitive Neuroscience, Brain Mind Institute, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland; ²Center for Neuroprosthetics, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland; and

³Department of Neurology, University Hospital Geneva, Geneva, Switzerland

Submitted 25 May 2012; accepted in final form 20 July 2012

Prsa M, Gale S, Blanke O. Self-motion leads to mandatory cue fusion across sensory modalities. *J Neurophysiol* 108: 2282–2291, 2012. First published July 25, 2012; doi:10.1152/jn.00439.2012.—When perceiving properties of the world, we effortlessly combine multiple sensory cues into optimal estimates. Estimates derived from the individual cues are generally retained once the multisensory estimate is produced and discarded only if the cues stem from the same sensory modality (i.e., mandatory fusion). Does multisensory integration differ in that respect when the object of perception is one's own body, rather than an external variable? We quantified how humans combine visual and vestibular information for perceiving own-body rotations and specifically tested whether such idiothetic cues are subjected to mandatory fusion. Participants made extensive size comparisons between successive whole body rotations using only visual, only vestibular, and both senses together. Probabilistic descriptions of the subjects' perceptual estimates were compared with a Bayes-optimal integration model. Similarity between model predictions and experimental data echoed a statistically optimal mechanism of multisensory integration. Most importantly, size discrimination data for rotations composed of both stimuli was best accounted for by a model in which only the bimodal estimator is accessible for perceptual judgments as opposed to an independent or additive use of all three estimators (visual, vestibular, and bimodal). Indeed, subjects' thresholds for detecting two multisensory rotations as different from one another were, in pertinent cases, larger than those measured using either single-cue estimate alone. Rotations different in terms of the individual visual and vestibular inputs but quasi-identical in terms of the integrated bimodal estimate became perceptual metamers. This reveals an exceptional case of mandatory fusion of cues stemming from two different sensory modalities.

mandatory fusion; metamers; multisensory integration; self-motion; vestibular

SENSORY CUES either within (Hillis et al. 2004; Knill and Saunderson 2003) or across senses (Alais and Burr 2004; Butler et al. 2010; Ernst and Banks 2002; Fetsch et al. 2009; Gu et al. 2008; Mendonca et al. 2011; van Beers et al. 1996, 1999) are often combined to produce the final percept according to statistical optimality. Until recently, this framework has been exclusively tested in situations where the object of perception is external to the body of the observer. However, the observer's body is also a multisensory object subjected to perceptual processes (Ionta et al. 2011). This is particularly apparent in passive whole body displacements that are perceived using mainly vision and the vestibular organs (Butler et al. 2010; Buttner and Henn 1981; Fetsch et al. 2009; Gu et al. 2008; Young et al. 1973). The associated cues are said to be idiothetic

in nature because they are derived from the observer's own displacements. Does the brain process and integrate two idiothetic signals differently from two externally generated signals? For example, an observer can simultaneously and independently use vision and audition to estimate the position of two different objects. In other words, sensory cues making the same physical measurement are often attributed to different causes and are not integrated (Koerding et al. 2007; Parise et al. 2012; Shams and Beierholm 2010). It seems that such dissociations cannot be made in the case of idiothetic cues in most ecological conditions. Visual and vestibular information relevant for perceiving self-motion are necessarily redundant and therefore always integrated. Optic flow incongruent with vestibular input does indeed arise but is in such cases caused by constituents of the visual surrounding that are not world-stationary. These visual cues are not idiothetic; they are attributed to external objects and thus most likely vetoed when estimating self-motion. Even in the presence of eye movements, optic flow information corresponding to the actual motion of the body is extracted to guarantee perceptual stability (Haarmeier et al. 1997, 2001; Royden et al. 1992). Because, for the purpose of estimating self-motion, visuovestibular integration involves sensory cues providing two ecologically nondissociable signals, its neural underpinnings must in some respect differ from those underlying nonidiothetic cue integration.

Extending recent work in heading perception involving linear translation stimuli (Butler et al. 2010; Fetsch et al. 2009; Gu et al. 2008) as well as theoretical models of the optimal use of vestibular signals in general (Laurens and Droulez 2007; Mac-Neilage et al. 2007; Zupan et al. 2002), we first show that the statistically optimal model of multisensory integration also applies to visual and vestibular cues when perceiving passive self-rotation. Crucially, we subsequently test whether the associated cues are subjected to mandatory fusion. Mandatory fusion entails that once integrated to produce a more reliable bisensory percept, perceptual access to unisensory estimates is lost and has been claimed not to occur across different sensory modalities (Hillis et al. 2002). However, we demonstrate that cross-modal mandatory fusion can ensue from idiothetic sensory input. When that situation arises, different cue combinations can theoretically give rise to the same fused percept, since they would differ only in terms of information that is lost. For example, a perceived rotation size S borne out by a whole body rotation of size $S + \Delta$ paired with an equally reliable visual cue simulating a rotation of size $S - \Delta$ can be indistinguishable from a true rotation size S produced by both stimuli. Such physically different but perceptually indistinguishable stimuli have been called metamers (Richards 1979) and can be compared to lights of the same color but different spectral compo-

Address for reprint requests and other correspondence: M. Prsa, Laboratory of Cognitive Neuroscience, Brain-Mind Institute, Ecole Polytechnique Fédérale de Lausanne, Station 19, AI 2101, 1015 Lausanne, Switzerland (e-mail: mario.prsa@mail.mcgill.ca).

sition. We argue that cues providing two ecologically nondissociable signals about own-body displacements account for this phenomenon.

MATERIALS AND METHODS

The optimal observer model. Based on how probable perceiving an own-body rotation of a certain size is given the visual and vestibular stimuli individually, the multisensory estimate can be predicted from probability theory. To this end, we describe each perceptual estimate in the form of a probability distribution (likelihood). The likelihood provides a probabilistic measure of the estimate: its most likely value and the uncertainty associated with this value. Bayesian statistics then formalize the optimal strategy for combining likelihoods arising from multiple sensory cues (and prior beliefs) to form the a posteriori estimate, the end result of the perceptual process. We describe an optimal observer model, show how it predicts both the variance and the mean of the posterior when human subjects integrate visual and vestibular cues for perceiving whole body rotations, and finally describe within the same framework how to test whether a mandatory fusion of the two cues occurs.

When the body is passively rotated around its yaw axis, both the visual and vestibular senses provide independent information (I_{vi} and I_{ve}) about the rotatory stimulus S . Probabilistic descriptions of the subjects' ensuing perception of rotation \hat{S} can be derived from visual and vestibular cues only [i.e., the likelihoods $P(I_{vi}|S)$ and $P(I_{ve}|S)$]

and from the bimodal pairing [i.e., the posterior $P(S|I_{vi}, I_{ve})$]. Each distribution recounts how likely it is to perceive any given rotation size. The value corresponding to the peak of the distribution is the most likely estimate, and its standard deviation captures how uncertain/reliable the estimate is (small values indicate low uncertainty and high reliability) (Fig. 1).

Maximum likelihood estimation (MLE), derived from Bayes' rule (Landy et al. 1995; Yuille and Bulthoff 1996), predicts that the optimal way for the brain to combine sensory cues will result in the posterior distribution being a normalized product of the visual and vestibular likelihoods

$$P(S|I_{vi}, I_{ve}) \propto P(I_{vi}|S)P(I_{ve}|S) \quad (1)$$

given the assumed Gaussian distributions are produced by independent sources of visual and vestibular noise. The predicted variance of \hat{S}_p , the value of S that maximizes $P(S|I_{vi}, I_{ve})$, is then

$$\sigma_p^2 = \frac{\sigma_{vi}^2 \sigma_{ve}^2}{\sigma_{vi}^2 + \sigma_{ve}^2}, \quad (2)$$

where σ_{vi}^2 and σ_{ve}^2 are the variances of rotation estimates that maximize $P(S|I_{vi})$ and $P(S|I_{ve})$, respectively. $P(S|I_{vi})$ and $P(S|I_{ve})$ are the single-cue posteriors and are equal to the respective likelihoods under the assumption that all sensory signals are equally likely to occur. The maximum a posteriori estimate (value corresponding to the peak of

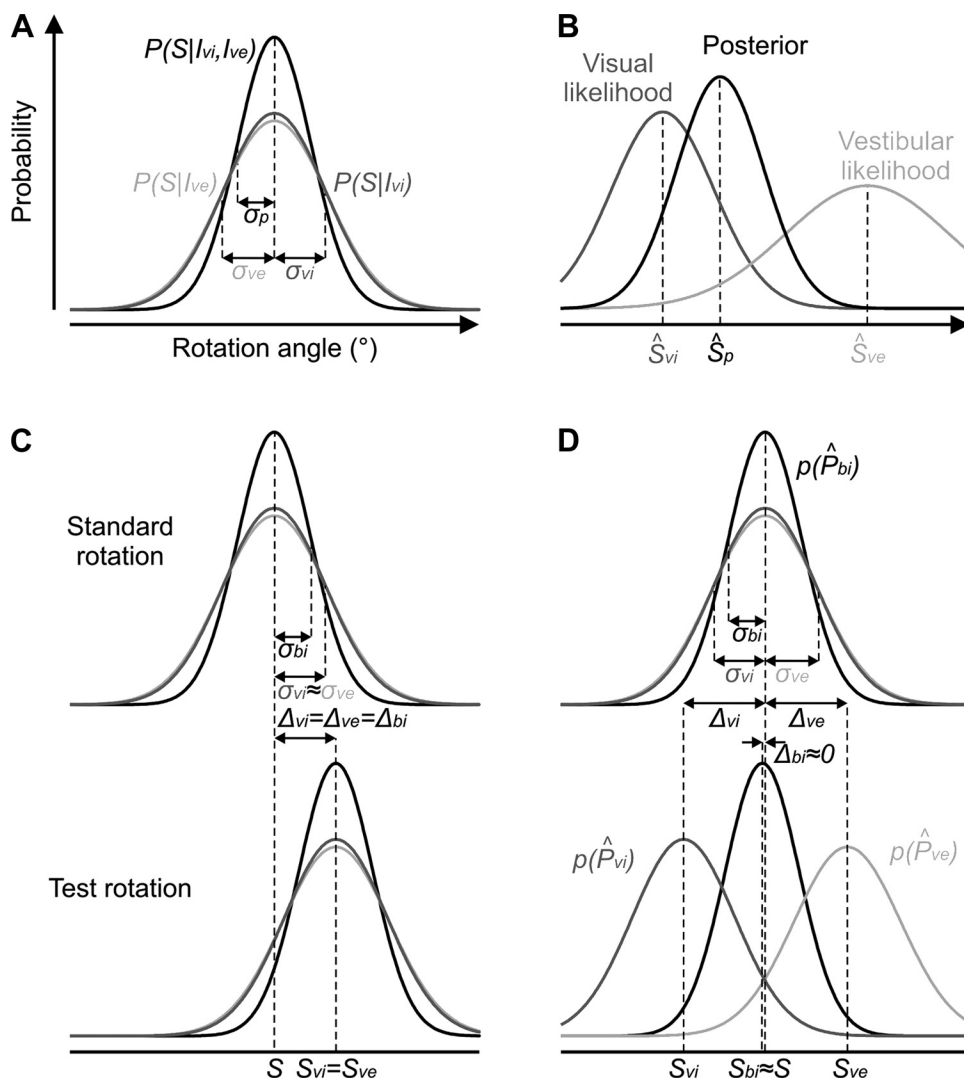


Fig. 1. Predictions of the optimal observer framework. **A** and **B**: a statistically optimal integration of sensory likelihoods results in a reduction of perceptual uncertainty according to Eq. 2 (**A**) and a reliability-based cue reweighting according to Eq. 3 (**B**). **C** and **D**: for rotations comprising both stimuli, optimal integration of the visual and vestibular likelihoods results in the posterior estimate according to Eq. 1. Discrimination between a standard and a test rotation can be based on either the single-cue estimators \hat{P}_{vi} and \hat{P}_{ve} (light and dark traces) or the bisensory estimator \hat{P}_{bi} (black trace). Traces symbolize the distribution of the estimates over trials. Successful discrimination is achieved when the physical difference between the estimates of the 2 rotations (Δ_{vi} , Δ_{ve} , and Δ_{bi}) is greater than the respective discrimination thresholds (defined as σ_{vi} , σ_{ve} , and σ_{bi}). Use of the bisensory estimator can either improve discriminability, as in **C**, or not, as in **D**. Empirical discrimination thresholds can therefore reveal which estimators the subjects have access to for making perceptual judgments. See MATERIALS AND METHODS for definition of all symbols and terms.

the posterior) will be a weighted average of the two most likely single-cue estimates

$$\hat{S}_p = w_{vi}\hat{S}_{vi} + w_{ve}\hat{S}_{ve}, \quad (3)$$

where the weights are equal to the normalized reliabilities (inverse of variance)

$$W_{vi} = \frac{1/\sigma_{vi}^2}{1/\sigma_{vi}^2 + 1/\sigma_{ve}^2} \quad W_{ve} = \frac{1/\sigma_{ve}^2}{1/\sigma_{vi}^2 + 1/\sigma_{ve}^2} \quad (4)$$

It follows that optimal integration always reduces the variability of the posterior relative to the individual likelihoods (Fig. 1A), and the estimate of the more reliable likelihood weighs more heavily on the posterior estimate (Fig. 1B).

To discriminate between a standard rotation of size S and a test rotation of size $S + \Delta$, the probabilistic observer can compare their visual and vestibular estimates \hat{P}_{vi} and \hat{P}_{ve} on any given trial. We refer to the two as the single-cue estimators, and we define the single-cue discrimination thresholds (T_{vi} and T_{ve}) as equal to one standard deviation over trials of \hat{P}_{vi} and \hat{P}_{ve} (i.e., σ_{vi} and σ_{ve}). When the rotations are composed of both stimuli, sensory integration produces an additional bisensory estimator \hat{P}_{bi} with discrimination threshold T_{bi} , estimated as in Eq. 2. Mandatory fusion can be tested by assessing which of the three estimators the subjects use to perceive the bimodal pairs ($S_{vi} = S$, $S_{ve} = S$) and ($S_{vi} = S + \Delta_{vi}$, $S_{ve} = S + \Delta_{ve}$) as different (Fig. 1, C and D). If all three estimators are available for perceptual judgments, then the two rotations can be perceived as different if either of the following inequalities is satisfied:

$$|\Delta_{vi}| > T_{vi} \quad (5)$$

$$|\Delta_{ve}| > T_{ve} \quad (6)$$

$$|w_{vi}\Delta_{vi} + w_{ve}\Delta_{ve}| > T_{bi} \quad (7)$$

where the cue weights are estimated as in Eq. 4. The left-hand side of Eq. 7 corresponds to the size difference Δ_{bi} between the bisensory estimates of the two rotations. Therefore, if all three estimators are accessible, successful discrimination can be achieved whenever the estimated difference between the two rotations using any of the estimators reaches its own discrimination threshold $\pm T_{vi}$, $\pm T_{ve}$, or $\pm T_{bi}$. When the visual and vestibular stimuli rotate by the same amount ($\Delta_{vi} = \Delta_{ve}$), an optimal observer can achieve better discrimination performance if using the bisensory estimator (Fig. 1C). However, with different combinations of Δ_{vi} and Δ_{ve} , the two rotations can be close to metamers (e.g., if $\Delta_{vi} = -\Delta_{ve}$, Fig. 1D), indistinguishable in terms of \hat{P}_{bi} (the estimated difference using \hat{P}_{bi} will not reach $\pm T_{bi}$), and successful discrimination can be achieved only if access to the unimodal estimators is still available. Empirical assessments of the subjects' discrimination thresholds and a comparison with theoretical predictions for different Δ_{vi}/Δ_{ve} values can therefore reveal which estimators the subjects have access to and use for perceiving the two rotations as different.

Experimental setup. Vestibular stimuli were delivered in complete darkness by a centrifuge cockpit-style chair digitally servo-controlled (PCI-7352) with highly precise positioning ($\pm 0.1^\circ$). The chair was centered on the rotation axis so that only angular and no linear stimuli were provided to the vestibular organs. Subjects were comfortably restrained with a five-point racing harness, feet straps, and additional cushioning. Head movements were minimized by using a head pillow and face paddles pressed against the cheek bones. Rotation profiles were precomputed and specified the chair's instantaneous angular position at a rate of 100 Hz. The rotations' velocity profile $v(t)$ was a single cycle of a 0.77-Hz raised cosine function (Fig. 2B)

$$v(t) = \frac{A}{T} \left[1 - \cos\left(\frac{2\pi t}{T}\right) \right], \quad (8)$$

where A is rotation size and T is its duration ($T = 1.3$ s in this case). Instantaneous angular position $p(t)$ is then specified as

$$p(t) = A \left[\frac{t}{T} - \frac{1}{2\pi} \sin\left(\frac{2\pi t}{T}\right) \right]. \quad (9)$$

Such transient stimuli fall into the pass-band range of the transfer function describing the cascade of semicircular canal dynamics and the velocity storage mechanism (Bertolini et al. 2011). The signal encoding the rotation at the first stage of neural processing is therefore quasi-identical to actual head velocity in this case. Visual-only stimuli were simulated by a moving stereoscopic random dot pattern presented on a 22-in. display fixed to the chair and facing the subject at a distance of ~ 29 cm (Fig. 2A). The limited visual field therefore covered $\sim 80^\circ$ of horizontal and 56° of vertical visual angle. The subject and the display were physically enclosed to eliminate any visual cues that might emanate from the stationary surroundings during rotations. The visual scene was constructed as an almost infinite three-dimensional volume of randomly distributed dots of different size. Rotations were then simulated by having the observer's point of view placed in the middle of this space and rotated around the yaw axis. The generated scene therefore simulated retinal optic flow information that would ensue from actual rotation. The stereoscopic stimulus was generated by the Nvidia Quadro FX 3800 graphics card using the OpenGL quad-buffer mechanism. The stimulus was programmed with the Python language and viewed with the Nvidia 3D Vision kit (active shutter glasses) paired with a Samsung Syncmaster 2233RZ display (120-Hz refresh rate) via an infrared transmitter. Subjects were required to maintain visual fixation on a stationary target in the middle of the display in all conditions, and masking white noise was delivered over earphones at all times. The fixation dot was of different color than the random dot pattern and appeared with zero binocular disparity.

Participants. Eight healthy adults (MP, SG, and 6 subjects naive to the aims of the experiment) with normal or corrected vision and no history of inner ear disease participated in each experiment (optimal integration: 1 female, mean age 27 ± 5.8 yr; mandatory fusion: 4 females, mean age 24 ± 4.2 yr). In both experiments, an additional subject completed the task but had to be discarded because the performance did not surpass chance level in a number of conditions. All participants gave informed consent and received monetary retribution at 20 CHF/h. The studies were approved by a local ethics committee and were conducted in accordance with the Declaration of Helsinki.

Experimental paradigms. To test for optimal integration, subjects judged the relative size of two successive rotations (the standard and the test) in a two-alternative forced-choice task (Fig. 2C). The size of the standard was 15° , and the test was any of 7 equally spaced angles in the interval 10° – 20° tested using the method of constant stimuli. These values were chosen on the basis of preliminary tests conducted on two subjects and such that they would include one point on each end of the psychometric fitting curve where size discrimination could be achieved with nearly 100% certainty. The two rotations were preceded, followed, and separated by an interval of 0.5 s. A 2-s period followed during which the subject had to answer, via a button press, whether the second rotation was bigger or smaller than the first. The standard rotation was randomly assigned to come either first or second, but measured responses were always those comparing the test with the standard. Each comparison was repeated 10 times per subject, and no feedback was given. Trials in which the subjects failed to give an answer were discarded; this constituted 0.2% of all trials. The relative reliability of the visual and vestibular cues was manipulated by changing the coherence of the visual motion (number of dots simulating rotation/number of dots moving randomly) from trial to trial between four different levels. The random motion varied in both direction and speed, but the stereoscopic depth of the individual dots did not change during the motion. The four coherence levels used were 100%, 75% (for 2 subjects) or 65% (for 6 subjects), 50%, and 25%. For bimodal comparisons, visual and vestibular stimuli were temporally synchronized and occurred simultaneously in congruent

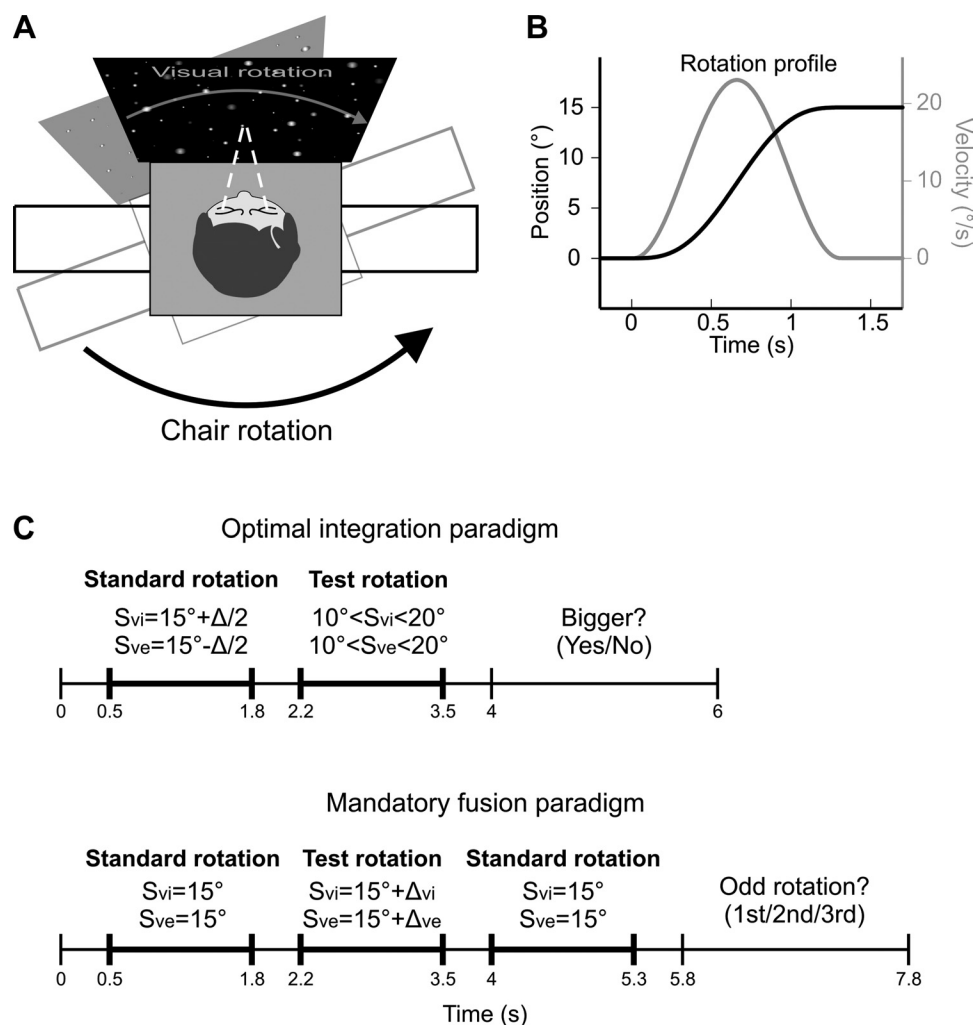


Fig. 2. Experimental setup and paradigms. *A*: schematic of the vestibular and visual stimuli delivered to the subjects for perceiving whole body yaw rotations. *B*: position and velocity profiles of the rotation stimulus. *C, top*: trial timeline of the optimal integration paradigm used to determine the sensory likelihoods and posterior. Subjects made size comparisons between a standard and a test stimulus. Visual- and vestibular-only conditions involved stimuli within a single modality. In the bimodal condition, the conflict angle Δ between the visual and vestibular rotations was 0° , 4° , or -4° . *Bottom*: trial timeline in the oddity detection task used to test for mandatory fusion. Subjects had to detect the odd stimulus among 3 successive rotations. The odd stimulus could be either the standard or the test and could come first, second, or third.

(i.e., opposite) directions. Three conflict angles were tested in the bimodal case ($S_{vi} - S_{ve} = 0^\circ$, 4° , or -4°) and were applied to the standard rotation. This produced 17 conditions in total (1 vestibular, 4 visual, and 12 bimodal) giving rise to 1,190 trials per subject (7 angles \times 10 repetitions \times 17 conditions). The experiment was divided into 14 sessions of 10 min each during which all the conditions were randomly intermingled with an intertrial interval of 0.9 s. Visual-only, vestibular-only, and bimodal stimuli were therefore alternated on a trial-by-trial basis, and predicting the nature of the stimulus was impossible. The subjects took regular breaks between sessions and completed the entire experiment in ~ 4 h. The direction of rotation (left or right) was randomly chosen on each trial.

To test for mandatory fusion, subjects had to pick out the odd stimulus among three successive rotations in a three-alternative forced-choice task (Fig. 2C). The same standard rotation size and the same trial timings were used as in the first experiment. The test rotation sizes were 7 equally spaced angles in the interval 9° – 21° . The sizes were again chosen on the basis of preliminary tests conducted on two subjects and such that they would include one point on each end of the fitting curve where size discrimination could be achieved with nearly 100% certainty. Two of the rotations were the test and one the standard (or vice versa), and all were presented in random order. For bimodal stimuli, the test rotation included conflicting visual and vestibular cue pairs ($S_{vi} = S + \Delta_{vi}$, $S_{ve} = S + \Delta_{ve}$). We tested eight conditions of bimodal pairings, each corresponding to a different Δ_{vi}/Δ_{ve} ratio (1, 0.5, 2, 0, ∞ , -1 , -0.5 , and -2) (see RESULTS for details). The different ratios yield different amounts of conflict between S_{vi} and S_{ve} , with no conflict in the 1 condition and maximum

conflict in the -1 condition. For the 1 ratio, S_{vi} and S_{ve} were always equal and therefore correspond to the easiest condition for picking out the odd stimulus (tested Δ_{ve} and Δ_{ve} values were 0, ± 1.71 , ± 3.43 , and ± 5.14). For the -1 ratio, the average of S_{vi} and S_{ve} was always equal to the standard rotation size S and, in the case of equal cue reliabilities, makes the identification of the odd stimulus theoretically impossible if single cues remain inaccessible. For the ∞ ratio, $\Delta_{ve} = 0$, meaning that S_{ve} was always equal to the standard rotation size, whereas S_{vi} took on the 7 equally spaced angles in the interval 9° – 21° . The opposite was true for the 0 condition. For the 2 and -2 ratios, tested Δ_{vi} values were 0, ± 1.71 , ± 3.43 and ± 5.14 and the Δ_{ve} values were calculated from the Δ_{vi}/Δ_{ve} ratio. The opposite was true for the -0.5 and 0.5 ratios. The subject had to answer whether the first, second, or third rotation was different from the other two on any basis. Each comparison was repeated 10 times across the 10 conditions (2 unimodal and 8 bimodal), producing a total of 700 trials per subject, randomly intermixed in 10 sessions of 10 min each. In both experiments, subjects initially underwent multiple short training sessions to familiarize themselves with the task and to ensure better than chance performance. Data collected on these sessions were not used for analysis.

Data analysis. All data analyses were performed off-line with custom programs compiled in MATLAB (The MathWorks). For each test angle, individual answers were pooled across all subjects to obtain a probabilistic measure of the response and yield a sufficient sample set for the statistical comparisons. This consisted of calculating the proportion of “bigger” (see Fig. 3) or “incorrect” responses (see Fig. 6) based on 80 answers (10 from each subject) for every test angle.

Using the method of least squares, the proportions of “bigger” responses in the first experiment were fit with a cumulative Gaussian function and the proportions of incorrect responses in the second experiment were fit with a Gaussian. Measures of the mean, variance, and discrimination threshold were then extracted from the obtained fits in each condition. A bootstrap analysis provided standard errors for each measure and allowed statistical comparison between the experimentally measured values and model predictions. This consisted of repeating the data fit for each condition 9,999 times on a different subset of responses each time. The different subsets were formed by taking at random, with replacement, N trials from the total set of N for each test angle (for $N = 80$, 10^{23} such combinations are possible). The standard deviation of 9,999 repeated measures is then the standard error of the measure obtained using the original data set. Statistical tests were made by assessing the amount of overlap between the bootstrap iterations of two measures. If the measure of interest is σ , and σ_{ex}^j and σ_{pr}^j are its experimental and predicted estimates obtained from the j th bootstrap sample, then the one-tailed bootstrap probability of ($\sigma_{\text{ex}} > \sigma_{\text{pr}}$) is

$$p = \frac{1}{B} \sum_{j=1}^B I(\sigma_{\text{ex}}^j - \sigma_{\text{pr}}^j > 0),$$

where $B = 9,999$ and $I()$ is the indicator function, which is equal to 1 when its argument is true and 0 otherwise. The inequality would be reversed for the probability of ($\sigma_{\text{ex}} < \sigma_{\text{pr}}$). The one-tailed bootstrap P value is therefore simply the proportion of ($\sigma_{\text{ex}}^j - \sigma_{\text{pr}}^j$) values that are more extreme than 0. We prefer this approach to parametric testing because it provides a direct computation of the cumulative distribution of a test statistic instead of having to use an asymptotic approximation.

RESULTS

Optimal integration of visual and vestibular cues. To experimentally obtain the visual and vestibular likelihoods and the posterior resulting from their integration when perceiving own-body rotations, subjects made relative judgments about the size of two consecutive rotations: a standard rotation of fixed angular displacement and a test rotation of variable size (Fig.

2C and MATERIALS AND METHODS). The probabilities of perceiving the test rotation as bigger than the standard were fit by a cumulative Gaussian function for the tested range of angles, and the Gaussian likelihoods and posterior can be derived for each condition by taking the mathematical derivative of the fits.

Variance measures extracted from the experimentally obtained sensory likelihoods were used to predict the variance of the posteriors according to MLE and Eq. 2 (see MATERIALS AND METHODS). Because these were evaluated by comparing the size of two successive rotations, they actually correspond to the true likelihood variances scaled by a factor of 2. The variances measured experimentally using the bimodal stimuli (means \pm SE: $\sigma_p = 2.7 \pm 0.2$, 3.1 ± 0.24 , 3.3 ± 0.26 , 3.9 ± 0.35 for the 4 coherence levels) were reduced relative to using either single cue alone ($\sigma_{\text{ve}} = 5.1 \pm 0.54$; $\sigma_{\text{vi}} = 3.5 \pm 0.28$, 3.7 ± 0.3 , 5.0 ± 0.53 , 7.3 ± 1.1) and closely matched the predictions for all four visual coherence levels (Fig. 3A). The result was readily reproducible across individual participants. The standard deviations of the psychometric fits to the bimodal data of individual subjects across coherence levels were not statistically different between the measured and MLE-predicted values (paired t -test, $P = 0.56$) and are shown in Fig. 4A. The same bimodal thresholds were, however, significantly reduced relative to the smallest single-cue thresholds (paired t -test, $P < 0.001$), confirming that integrating information from visual and vestibular senses reduces perceptual uncertainty in a statistically optimal manner when estimating rotatory self-motion. As the critical test for demonstrating cue integration, we furthermore provide a comparison between the bimodal and single-cue thresholds for each subject in the case where the visual and vestibular weights were closest to being equal (Fig. 4B).

To test for the predictions of Eq. 3, size comparisons were made against a standard stimulus in which the chair and the visual scene rotated by different amounts so that $\hat{S}_{\text{vi}} \neq \hat{S}_{\text{ve}}$. When \hat{S}_{ve} signaled a bigger rotation than \hat{S}_{vi} , the subjects' most

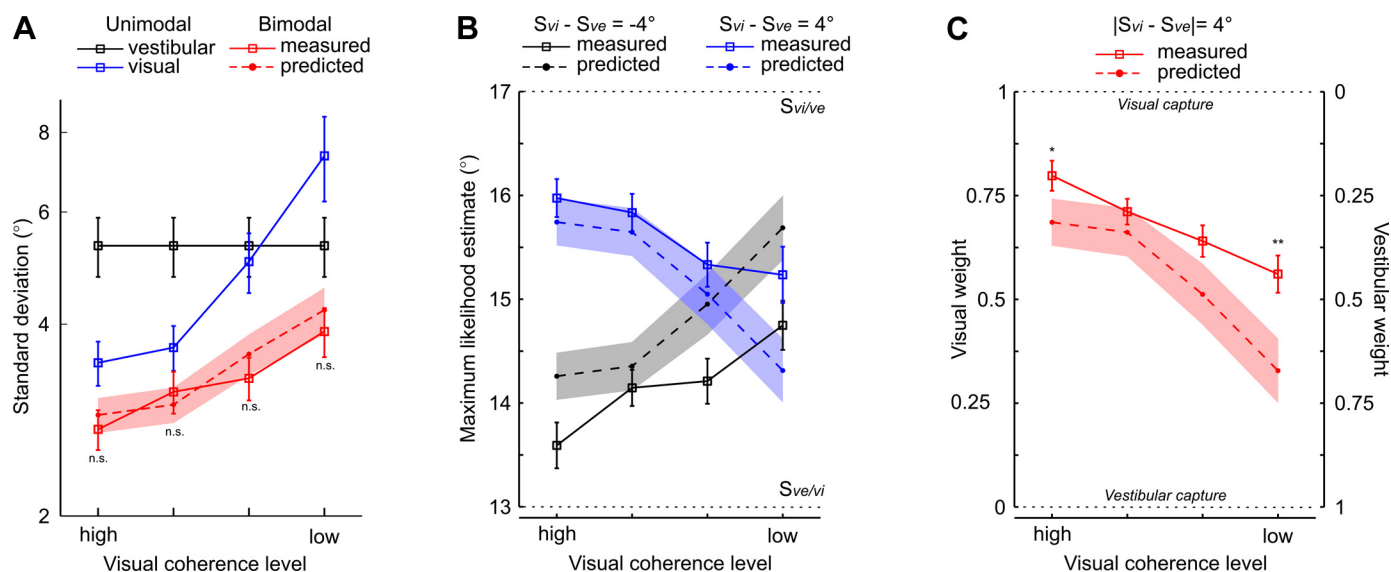


Fig. 3. Optimal integration of visual and vestibular cues. A: standard deviations obtained from the measured visual, vestibular, and bisensory likelihoods for 4 different levels of visual coherence. The empirical bisensory values were not significantly different (n.s.: $P > 0.2$, 1-tailed bootstrap test) from the maximum likelihood estimation (MLE) predictions. B: MLEs extracted from the measured posterior distributions and compared with the MLE predictions for the 2 conflicting visual and vestibular stimuli ($S_{\text{vi}} = 13^\circ$, $S_{\text{ve}} = 17^\circ$) and ($S_{\text{vi}} = 17^\circ$, $S_{\text{ve}} = 13^\circ$). C: averaged data from the 2 conditions in B and expressed in terms of weights show a suboptimal cue reweighting compared with model predictions (* $P < 0.05$, ** $P < 0.01$, 1-tailed bootstrap test). All delimiters correspond to bootstrap standard error.

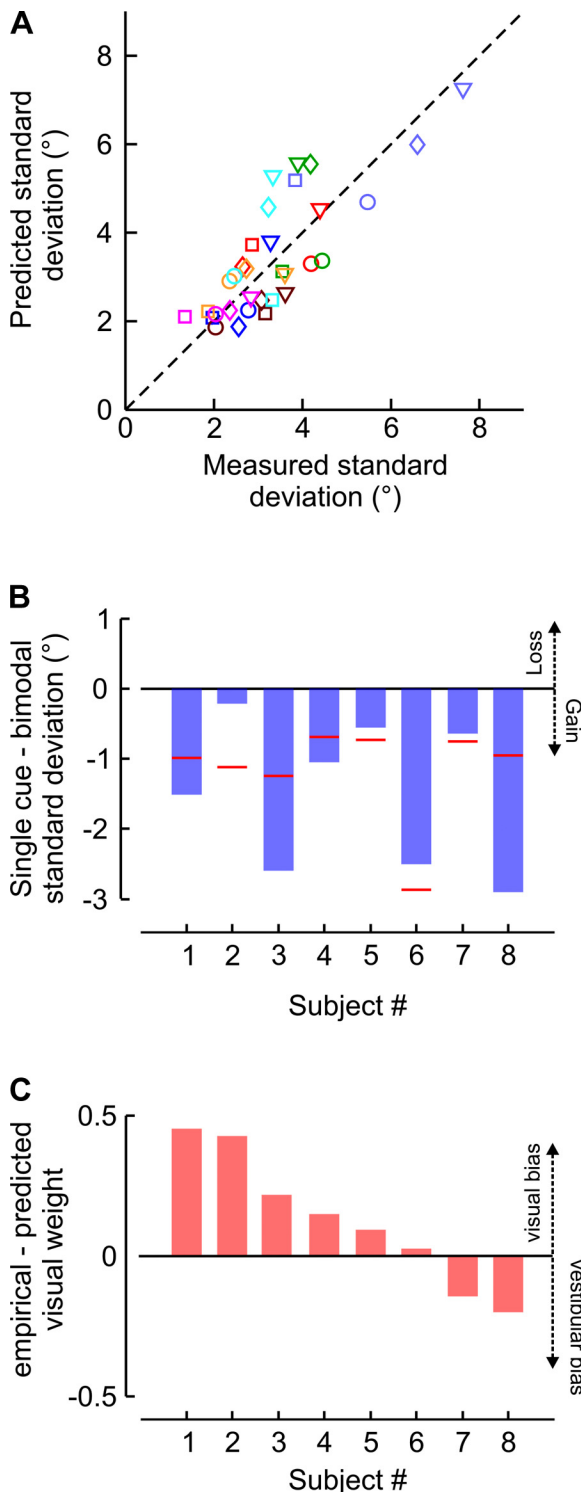


Fig. 4. Individual subject data. A: similarity between the predicted and measured standard deviations of the posterior across all subjects and for all 4 coherence levels. Colors denote different subjects, and symbols correspond to the different coherence levels. B: difference between the lowest single-cue standard deviation and the measured (blue bars) and predicted (red lines) bimodal standard deviations for the coherence level yielding best-matched visual and vestibular weights for each subject. C: difference between predicted and empirical visual weights averaged across the 4 coherence levels and sorted in descending order for the 8 subjects.

likely posterior estimate \hat{S}_p increased with decreasing visual coherence level (black trace in Fig. 3B). The opposite was observed when $\hat{S}_{vi} > \hat{S}_{ve}$ (blue trace in Fig. 3B). This is what is predicted if the subjects dynamically attribute more weight to the vestibular cue than the visual, according to Eqs. 3 and 4, as the reliability of the latter is reduced (dotted lines in Fig. 3B). Pooling the data from the two conflict conditions and expressing it in terms of visual and vestibular weights (Fig. 3C) shows that cue reweighting occurs and follows the MLE-predicted trend but deviates from optimality because subjects tended to significantly overweigh the visual cue. The extent of the visual bias was, however, variable across individuals (Fig. 4C).

Mandatory fusion. We next addressed the issue of mandatory cue fusion by adopting a paradigm (Hillis et al. 2002) measuring an observer's ability to perceive two rotations as different. Subjects were asked to detect the odd stimulus among three successive rotations, two of which were the standard and one the test (or vice versa) (Fig. 2C and MATERIALS AND METHODS). This oddity task is advantageous to outright asking the subjects whether two rotations are different, because it forces them to adopt the same decision criterion and thus eliminates response bias from the observer's actual discriminability (Swets 1961) (i.e., the subjects stay unaware of the task's real purpose; perceptual responses are not contaminated by higher level cognitive strategies).

In the unimodal cases, comparisons were made between a standard rotation of size S and a test of size $S + \Delta$. The proportions of incorrect responses for each Δ were fit by a Gaussian function (Fig. 5A). The fits symbolize the distribution over trials of visual and vestibular estimators \hat{P}_{vi} and \hat{P}_{ve} , the probabilistic descriptions of the observer's discrimination ability using each sense alone, and the associated single-cue discrimination thresholds T_{vi} and T_{ve} that we define here as equal to one standard deviation of the Gaussian fits. The coherence level used for the visual stimulus was the one that yielded $\sigma_{vi} \approx \sigma_{ve}$ for the average subject when optimal integration was tested in the first experiment (see Fig. 3A). In the bimodal case, discriminability between the pairs ($S_{vi} = S$, $S_{ve} = S$) and ($S_{vi} = S + \Delta_{vi}$, $S_{ve} = S + \Delta_{ve}$) was assessed. We tested eight conditions of bimodal pairings, each corresponding to a different Δ_{vi}/Δ_{ve} ratio. A ratio of 1 generates visual and vestibular stimuli that always rotate by the same amount, and the bisensory estimator \hat{P}_{bi} should provide better discrimination ability than either single-cue estimator in that case (Fig. 1C). A ratio of -1 produces conflicting visual and vestibular cues but identical bisensory estimates between the standard and the test according to Eq. 3 when $w_{vi} \approx w_{ve}$. Theoretically, the ability to discriminate the two metameric stimuli will be compromised if the observer only uses \hat{P}_{bi} and does not retain the unimodal estimators \hat{P}_{vi} and \hat{P}_{ve} (Fig. 1D). The remaining six conditions corresponded to Δ_{vi}/Δ_{ve} values (0.5, 2, 0, ∞ , -0.5 , -2) that produce bimodal stimuli with varying amounts of cue conflict that lie between the latter two extremes.

If all three estimators are accessible, in the cue space depicted in Fig. 5B, the predicted discrimination thresholds (blue dots) would lie on the contour defined by the red (Eqs. 5 and 6) and green lines (Eq. 7). That is the case if the subjects do not make an additive use of all three estimators. Probability summation is actually unrealistic, because it assumes that the three estimators can be used independently. Given the fact that \hat{P}_{bi} is a weighted average of \hat{P}_{vi} and \hat{P}_{ve} , the independence

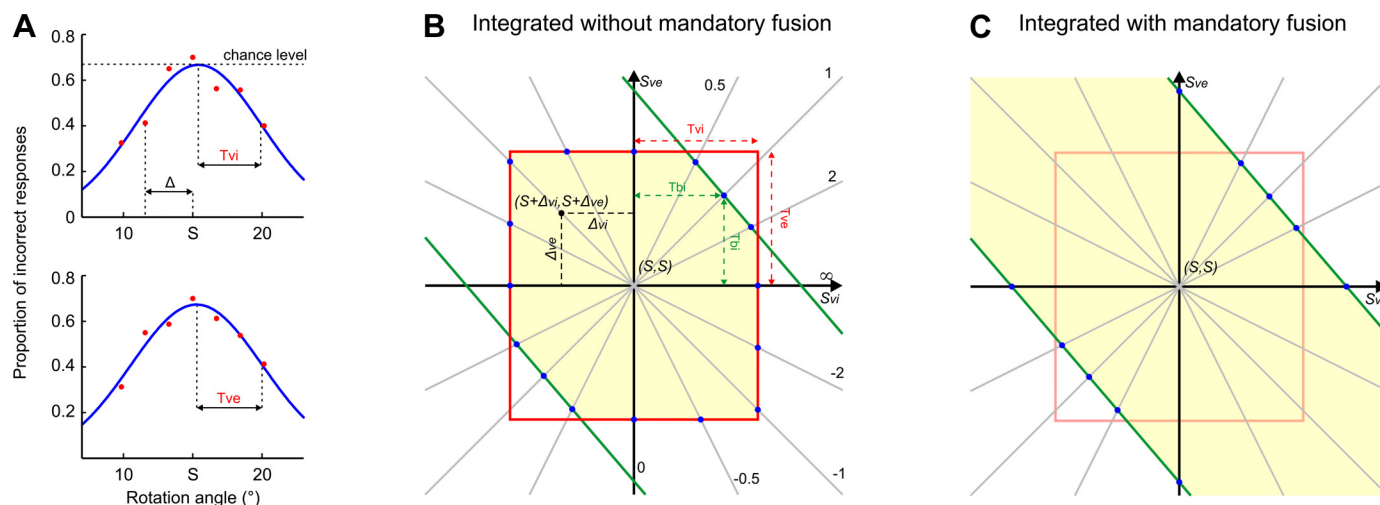


Fig. 5. Theoretical predictions for rotation size discrimination. **A**: mean responses for choosing the odd visual or vestibular stimulus (i.e., perceiving $S + \Delta$ as different from S) obtained by pooling raw choices from all subjects. The Gaussian fits symbolize the visual and vestibular estimators and provide the unimodal discrimination thresholds T_{vi} and T_{ve} . **B**: cue space where each point corresponds to a bimodal pairing (abscissas and ordinates indicate rotation sizes signaled by the visual and vestibular stimuli, respectively). When perceiving $(S + \Delta_{vi}, S + \Delta_{ve})$ as different from (S, S) , if the cues are integrated without mandatory fusion, discrimination thresholds (blue dots) are predicted to lie on the contour defined by the red (Eqs. 5 and 6) and green lines (Eq. 7). Gray lines and main axes each correspond to a Δ_{vi}/Δ_{ve} value representing one of the tested conditions. The yellow area comprises data points theoretically indistinguishable from (S, S) . **C**: predictions in the case of mandatory fusion.

assumption is clearly invalid. The probability summation model will thus not be considered further and in any case makes a less conservative prediction than the model in Fig. 5B. If mandatory fusion was to occur, perceptual access to \hat{P}_{vi} and \hat{P}_{ve} would be lost and discrimination performance would uniquely be based on the bisensory estimator. (S, S) can be perceived as different from $(S + \Delta_{vi}, S + \Delta_{ve})$ only if Eq. 7 is satisfied, giving rise to the theoretical prediction shown in Fig. 5C.

The prediction depicted in Fig. 5B was compared with the mandatory fusion prediction (Fig. 5C) to see which best accounts for the discrimination thresholds obtained experimentally in the eight tested conditions. The predicted contours were

constructed using the measured T_{vi} and T_{ve} (means \pm SE: 4.54 ± 0.42 and 4.74 ± 0.45 , respectively) and the estimated T_{bi} . Our results show that the subjects' discrimination thresholds (again defined as equal to 1 standard deviation of the obtained Gaussian fits) for perceiving (S, S) as different from $(S + \Delta_{vi}, S + \Delta_{ve})$ were in many cases larger than those predicted from an independent use of all three estimators (Fig. 6A). Individual subjects indeed showed consistent losses in discrimination performance compared with prediction with the use of single-cue estimators, even without adding probabilistic responses (Fig. 6B). Threshold predictions were derived for each subject individually from the corresponding measured single-

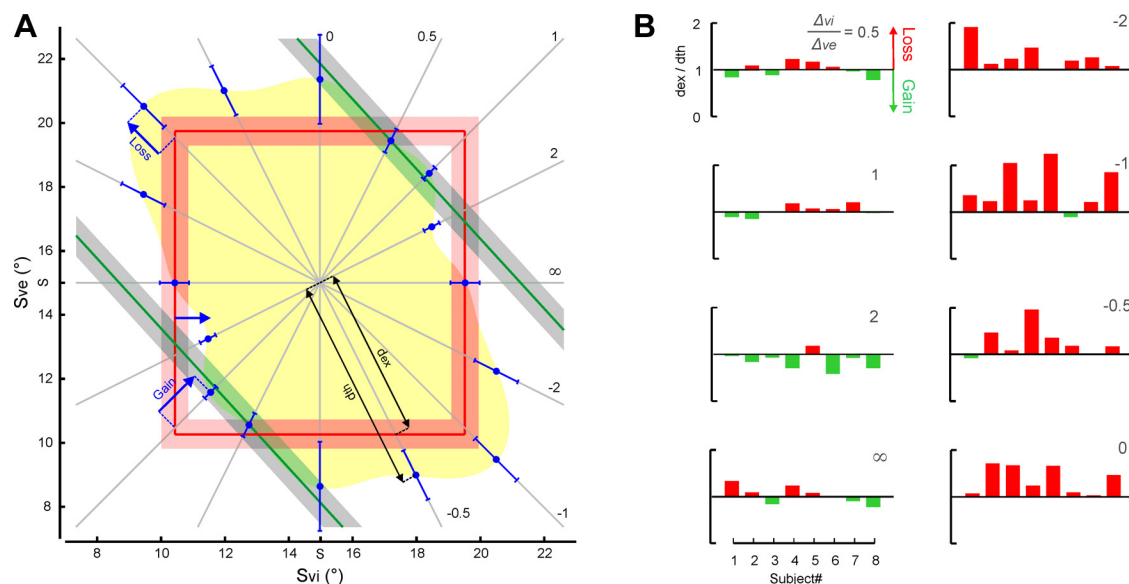


Fig. 6. Comparison between predicted and measured discrimination thresholds. **A**: compared with single-cue thresholds (red), pooled data from all subjects show improved discrimination ability ("gain") when using the bisensory estimator (green) is advantageous (0.5, 1, and 2 conditions) and worse discrimination ("loss") when it is not (∞ , -2 , -1 , -0.5 , and 0 conditions). **B**: loss or gain in discrimination ability compared with the theoretical prediction of Fig. 5B for individual subject data as measured by the d_{ex}/d_{th} ratio, where d_{ex} is the distance between the cue space origin and experimental data (blue), and d_{th} is the distance between the origin and the theoretical limits (red or green), as depicted in **A**. All shadings and delimiters correspond to bootstrap standard error.

cue thresholds. For Δ_{vi}/Δ_{ve} ratios (0.5, 1 and 2) where the use of \hat{P}_{bi} leads to a gain in performance over the use of \hat{P}_{vi} and \hat{P}_{ve} , single subject thresholds (Fig. 6B) were not statistically different ($P > 0.05$, t -test) from the identical predictions of Fig. 5, B and C. On the other hand, in the remaining conditions where mandatory fusion leads to a worse performance, subjects showed losses in discrimination ability (Fig. 6B) relative to use of the single-cue estimators ($P < 0.05$, for $\Delta_{vi}/\Delta_{ve} = -2, -1, -0.5$ and 0, average power = 0.81, and $P > 0.3$ for ∞ , 1-tailed t -test). The results are therefore best accounted for by assuming that only the bisensory estimator stemming from sensory fusion is available for making perceptual judgments. Conditions where very large, even infinite, discrimination thresholds are predicted ($\Delta_{vi}/\Delta_{ve} = -0.5, -1, -2$) are those that involve stimuli with the largest conflict angles between the visual and vestibular cues. Because the subjects were free to detect the odd stimulus on any basis, it is reasonable to assume that detection was based on sometimes perceiving the conflict in trials corresponding to the biggest Δ values. The absence of optimal cue integration in those cases explains the noninfinite or lower than theoretically expected thresholds. Also, a perceptual aspect other than rotation size or conflict detection might have transpired in cases of large conflict and been used by the subjects to do the task. We finally note that visuovestibular integration resulted in overweighing of the visual cue (Fig. 3C): including a visual bias in the model would increase w_{vi} and decrease w_{ve} in Eq. 7, translate into the green lines being rotated clockwise in Fig. 6B, and therefore yield an even better correspondence between experimental results and the mandatory fusion prediction.

DISCUSSION

When perceiving properties of the world, observers combine redundant information from different sensory modalities. Perceptual estimates derived from each of these sensory cues will naturally exhibit variability across repeated observations of the same stimulus. It has been repeatedly shown that the variability is reduced in a statistically optimal manner if cues are combined to produce the final percept. The neural mechanism underlying multisensory integration is therefore likely to be a process of probabilistic inference seeking to reduce perceptual uncertainty (Knill and Pouget 2004). Our results demonstrate, and extend what has recently been reported for the perception of heading (Butler et al. 2010; Fetsch et al. 2009), that the central nervous system integrates multisensory idiothetic information according to the same laws of probability. Indeed, the same optimal reduction of variance and cue reweighting in proportion to relative reliability occurs as for nonidiothetic cues (Fig. 3). When perceiving whole body rotations, however, subjects tended to overweigh the visual cue (Figs. 3C and 4C), which is in contrast with heading perception, where an excessive reliance on the vestibular cue is observed (Butler et al. 2010; Fetsch et al. 2009). A selective overweighing of otolith signals and an underweighing of semicircular canal signals might thus be occurring when integrated with vision. Similarly suboptimal cue weights have been observed for other sensory modalities as well (Battaglia et al. 2003; Knill and Saunders 2003), and why such deviations from optimality occur is not fully understood. One possibility is that the biases might be reflective of a recalibration mechanism, where one estimate is

remapped and realigned with the other in an attempt to internally correct the inconsistencies in multisensory input. Such a process is indeed known to occur when the observer is constrained to integrate conflicting information (Adams et al. 2001; Block and Bastian 2011), even after only a very short exposure (in the milliseconds range) to the conflict (Wozny and Shams 2011).

When sensory cues are combined, perception is governed by the posterior, the product of their integration. Because they are individually less reliable, neural signals underlying the two isolated sensory likelihoods are therefore not used at the perceptual level. It has been demonstrated that when the two cues originate from different sensory modalities (vision and touch, for example), observers still have access to this unused unimodal information (Hillis et al. 2002). Why might the ability to access these “useless” signals be preserved? In laboratory settings, experimental conditions can be constructed where the use of individual likelihoods, if available, is advantageous in terms of better task performance, even in the presence of a more reliable estimate (Fig. 1D). Such conditions, however, have little ecological validity. They can only serve as a useful experimental tool for testing whether likelihood signals are accessible but do not provide a valid explanation for their accessibility. The reasons might instead be rooted in the causal inference process (Koerding et al. 2007; Parise et al. 2012; Shams and Beierholm 2010) since observers often effortlessly attribute separate causes to simultaneously received cross-modal cues. The dissociated cues are not integrated, and each gives rise to an independent percept. Preserving unused information when cues are integrated might thus be reflective of a neural organization sculptured by the experience that most cues simply do not have the same cause in everyday life. Actually, the MLE model of Eq. 1 is a simplification, under the common cause assumption, of a more generalized hierarchical causal inference (HCI) model (Shams and Beierholm 2010). In the full HCI model, the a posteriori estimate is a weighted average of the MLE estimate of Eq. 3 and one of the single-cue estimates, the one that dominates when separate causes are inferred. The weighting is determined by the probability of the common cause scenario. Therefore, whenever the latter probability is not equal to one, the individual likelihoods must remain accessible for one of them to be combined with the product of their initial integration.

Our results show that this generalization does not apply to cross-modal integration when the object of perception is the perceiver's own body. Visual and vestibular idiothetic cues are individually discarded after being fused into a single percept (Fig. 6) similarly to nonidiothetic cues within the same sense. Indeed, mandatory fusion has so far only been observed for the integration of unimodal cues: binocular disparity and texture gradients when perceiving the slant of a surface (Hillis et al. 2002; Nardini et al. 2010). These two cases share the characteristic of cues not being dissociable in natural conditions; they are necessarily redundant and always integrated. The probability of the common cause scenario is therefore never different from unity. Even when observers are instructed to actively attend either the visual or vestibular stimulus and ignore the other, providing the cue conflicts go unnoticed, they seem incapable of weighing the two cues independently (Berger and Buelthoff 2009). Because there is a cost, in terms of brain resources, associated with carrying two neural representations

of the same rotation stimulus, mandatory fusion likely results from the absence of an evolutionary pressure to preserve the individual channels for stimuli including both information sources. Sensory signals for perceiving self-motion thus seem to be processed more like unimodal cues rather than as originating from separate senses. In addition, the particular rotations we used (Fig. 2B) are faithfully encoded by both visual and vestibular systems. For a broader range of stimuli, however, both systems are individually deficient. Low-frequency, low-acceleration vestibular stimuli and high-frequency, high-acceleration visual stimuli are inaccurately encoded at early stages of neural processing (Waespe and Henn 1977, 1979). Only neural signals generated by rotations including both sensory inputs give a truthful account of the actual motion of the body over the entire operating range (Dichgans et al. 1973; Waespe and Henn 1977, 1979). A loss of inaccurate unisensory information might be incurred for the benefit of a more accurate bisensory signal, thus providing a potential neurophysiological basis for the mandatory fusion that we observed. We postulate that the same is also likely to be true for heading perception (Butler et al. 2010; Fetsch et al. 2009) but not in other instances where optimal cross-modal cue integration has been previously demonstrated (Alais and Burr 2004; Ernst and Banks 2002; Mendonca et al. 2011). This also applies to own hand position perception using visual and proprioceptive cues (van Beers et al. 1996, 1999), despite the idiothetic nature of the latter, because it is identical, in all things that matter, to estimating the position of an external object held in that hand.

It follows that passive whole body rotations can be simulated by combinations of visuovestibular stimuli that, like color metamers, differ physically but “look” the same. Our result is also akin to examples of visual metamers such as, for instance, two verniers with opposite offsets flashed in quick succession that are fused and become indistinguishable from a single, almost aligned vernier (Scharnowski et al. 2009). Color metamers occur because spectral information is lost by cone photoreceptors in the retina (Wandell 1995) and therefore cannot be recovered in the brain. In the case of visual and visuovestibular metamers, information loss necessarily occurs at a neural level upstream of where the metameric stimuli would yield identical neural responses. We envision that our findings might guide future electrophysiology and modeling approaches to elucidate where and how visuovestibular metamers are formed, similar to recent attempts at explaining visual metamerism (Freeman and Simoncelli 2011) and at examining the dynamics of the fusion process (Scharnowski et al. 2009). Finally, mandatory fusion of visual and vestibular signals can explain why neurons in cortical and subcortical vestibular centers are highly multimodal (Akbarian et al. 1988; Bremmer et al. 2002; Buttner and Buettner 1978; Butter and Henn 1976; Dichgans et al. 1973; Duffy 1998; Grusser et al. 1990; Henn et al. 1974; Meng and Angelaki 2010; Page and Duffy 2003; Schlack et al. 2002; Takahashi et al. 2007; Waespe et al. 1981; Waespe and Henn 1977, 1979) and most often visually responsive, and it hints at developmental reasons for the absence of a centralized vestibular cortex (Fukushima 1997; Guldin and Grusser 1998).

ACKNOWLEDGMENTS

We thank Bruno Herbelin for technical assistance and Danilo Jimenez Rezende for advice on data analysis, as well as Michael Herzog and Wulfram Gerstner for comments on an earlier version of the manuscript.

GRANTS

This study was supported by Swiss National Science Foundation Grant SINERGIA CRSIII-125135/1.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS

M.P. and O.B. conception and design of research; M.P. and S.G. performed experiments; M.P. analyzed data; M.P., S.G., and O.B. interpreted results of experiments; M.P. prepared figures; M.P. drafted manuscript; M.P., S.G., and O.B. edited and revised manuscript; M.P., S.G., and O.B. approved final version of manuscript.

REFERENCES

- Adams WJ, Banks MS, van Ee R. Adaptation to three-dimensional distortions in human vision. *Nat Neurosci* 4: 1063–1064, 2001.
- Akbarian S, Berndt K, Grusser OJ, Guldin W, Pause M, Schreier U. Responses of single neurons in the parietoinsular vestibular cortex of primates. *Ann NY Acad Sci* 545: 187–202, 1988.
- Alais D, Burr D. The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol* 14: 257–262, 2004.
- Battaglia PW, Jacobs RA, Aslin RN. Bayesian integration of visual and auditory signals for spatial localization. *J Opt Soc Am A* 20: 1391–1397, 2003.
- Berger DR, Buelthoff HH. The role of attention on the integration of visual and inertial cues. *Exp Brain Res* 198: 287–300, 2009.
- Bertolini G, Ramat S, Laurens J, Bockisch CJ, Marti S, Straumann D, Palla A. Velocity storage contribution to vestibular self-motion perception in healthy human subjects. *J Neurophysiol* 105: 209–223, 2011.
- Block HJ, Bastian AJ. Sensory weighting and realignment: independent compensatory processes. *J Neurophysiol* 106: 59–70, 2011.
- Bremmer F, Klam F, Duhamel JR, Ben Hamed S, Graf W. Visual-vestibular interactive responses in the macaque ventral intraparietal area (VIP). *Eur J Neurosci* 16: 1569–1586, 2002.
- Butler JS, Smith ST, Campos JL, Buelthoff HH. Bayesian integration of visual and vestibular signals for heading. *J Vis* 10: 2010.
- Buttner U, Buettner UW. Parietal cortex (2v) neuronal activity in alert monkey during natural vestibular and optokinetic stimulation. *Brain Res* 153: 392–397, 1978.
- Buttner U, Henn V. Circularvection—psychophysics and single-unit recordings in the monkey. *Ann NY Acad Sci* 374: 274–283, 1981.
- Buttner U, Henn V. Thalamic unit-activity in alert monkey during natural vestibular stimulation. *Brain Res* 103: 127–132, 1976.
- Dichgans J, Schmidt CL, Graf W. Visual input improves speedometer function of vestibular nuclei in goldfish. *Exp Brain Res* 18: 319–322, 1973.
- Duffy CJ. MST neurons respond to optic flow and translational movement. *J Neurophysiol* 80: 1816–1827, 1998.
- Ernst MO, Banks MS. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415: 429–433, 2002.
- Fetsch CR, Turner AH, DeAngelis GC, Angelaki DE. Dynamic reweighting of visual and vestibular cues during self-motion perception. *J Neurosci* 29: 15601–15612, 2009.
- Freeman J, Simoncelli EP. Metamers of the ventral stream. *Nat Neurosci* 14: 1195–U1130, 2011.
- Fukushima K. Corticovestibular interactions: anatomy, electrophysiology, and functional considerations. *Exp Brain Res* 117: 1–16, 1997.
- Grusser OJ, Pause M, Schreier U. Vestibular neurons in the parieto-insular cortex of monkeys (*Macaca fascicularis*)—visual and neck receptor responses. *J Physiol* 430: 559–583, 1990.
- Gu Y, Angelaki DE, DeAngelis GC. Neural correlates of multisensory cue integration in macaque MSTd. *Nat Neurosci* 11: 1201–1210, 2008.
- Guldin WO, Grusser OJ. Is there a vestibular cortex? *Trends Neurosci* 21: 254–259, 1998.
- Haarmeier T, Bunjes F, Lindner A, Berret E, Thier P. Optimizing visual motion perception during eye movements. *Neuron* 32: 527–535, 2001.
- Haarmeier T, Thier P, Repnow M, Petersen D. False perception of motion in a patient who cannot compensate for eye movements. *Nature* 389: 849–852, 1997.

- Henn V, Young LR, Finley C.** Vestibular nucleus units in alert monkeys are also influenced by moving visual-fields. *Brain Res* 71: 144–149, 1974.
- Hillis JM, Ernst MO, Banks MS, Landy MS.** Combining sensory information: mandatory fusion within, but not between, senses. *Science* 298: 1627–1630, 2002.
- Hillis JM, Watt SJ, Landy MS, Banks MS.** Slant from texture and disparity cues: optimal cue combination. *J Vis* 4: 967–992, 2004.
- Ionta S, Heydrich L, Lenggenhager B, Mouthon M, Fornari E, Chapuis D, Gassert R, Blanke O.** Multisensory mechanisms in temporo-parietal cortex support self-location and first-person perspective. *Neuron* 70: 363–374, 2011.
- Knill DC, Pouget A.** The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci* 27: 712–719, 2004.
- Knill DC, Saunders JA.** Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Res* 43: 2539–2558, 2003.
- Koording KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L.** Causal inference in multisensory perception. *PLoS One* 2: 2007.
- Landy MS, Maloney LT, Johnston EB, Young M.** Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Res* 35: 389–412, 1995.
- Laurens J, Droulez J.** Bayesian processing of vestibular information. *Biol Cybern* 96: 389–404, 2007.
- MacNeilage PR, Banks MS, Berger DR, Buelthoff HH.** A Bayesian model of the disambiguation of gravito-inertial force by visual cues. *Exp Brain Res* 179: 263–290, 2007.
- Mendonca C, Santos JA, Lopez-Moliner J.** The benefit of multisensory integration with biological motion signals. *Exp Brain Res* 213: 185–192, 2011.
- Meng H, Angelaki DE.** Responses of ventral posterior thalamus neurons to three-dimensional vestibular and optic flow stimulation. *J Neurophysiol* 103: 817–826, 2010.
- Nardini M, Bedford R, Mareschal D.** Fusion of visual cues is not mandatory in children. *Proc Natl Acad Sci USA* 107: 17041–17046, 2010.
- Page WK, Duffy CJ.** Heading representation in MST: sensory interactions and population encoding. *J Neurophysiol* 89: 1994–2013, 2003.
- Parise CV, Spence C, Ernst MO.** When correlation implies causation in multisensory integration. *Curr Biol* 22: 46–49, 2012.
- Richards W.** Quantifying sensory channels - generalizing colorimetry to orientation and texture, touch, and tones. *Sens Processes* 3: 207–229, 1979.
- Royden CS, Banks MS, Crowell JA.** The perception of heading during eye-movements. *Nature* 360: 583–587, 1992.
- Scharnowski F, Rueter J, Jolij J, Hermens F, Kammer T, Herzog MH.** Long-lasting modulation of feature integration by transcranial magnetic stimulation. *J Vis* 9: 2009.
- Schlack A, Hoffmann KP, Bremmer F.** Interaction of linear vestibular and visual stimulation in the macaque ventral intraparietal area (VIP). *Eur J Neurosci* 16: 1877–1886, 2002.
- Shams L, Beierholm UR.** Causal inference in perception. *Trends Cogn Sci* 14: 425–432, 2010.
- Swets JA.** Is there a sensory threshold? *Science* 134: 168–177, 1961.
- Takahashi K, Gu Y, May PJ, Newlands SD, DeAngelis GC, Angelaki DE.** Multimodal coding of three-dimensional rotation and translation in area MSTd: comparison of visual and vestibular selectivity. *J Neurosci* 27: 9742–9756, 2007.
- van Beers RJ, Sittig AC, Denier van der Gon JJ.** Localization of a seen finger is based exclusively on proprioception and on vision of the finger. *Exp Brain Res* 125: 43–49, 1999.
- van Beers RJ, Sittig AC, Denier van der Gon JJ.** How humans combine simultaneous proprioceptive and visual position information. *Exp Brain Res* 111: 253–261, 1996.
- Waespe W, Buttner U, Henn V.** Visual-vestibular interaction in the flocculus of the alert monkey. 1. Input activity. *Exp Brain Res* 43: 337–348, 1981.
- Waespe W, Henn V.** Neuronal activity in vestibular nuclei of alert monkey during vestibular and optokinetic stimulation. *Exp Brain Res* 27: 523–538, 1977.
- Waespe W, Henn V.** Velocity response of vestibular nucleus neurons during vestibular, visual, and combined angular-acceleration. *Exp Brain Res* 37: 337–347, 1979.
- Wandell BA.** *Foundations of Vision*. Sunderland, MA: Sinauer Associates, 1995.
- Wozny DR, Shams L.** Recalibration of auditory space following milliseconds of cross-modal discrepancy. *J Neurosci* 31: 4607–4612, 2011.
- Young LR, Dichgans J, Murphy R, Brandt T.** Interaction of optokinetic and vestibular stimuli in motion perception. *Acta Otolaryngol* 76: 24–31, 1973.
- Yuille AL, Bulthoff HH.** Bayesian decision theory and psychophysics. In: *Perception as Bayesian Inference*, edited by Knill DC and Richards W. Cambridge: Cambridge University Press, 1996.
- Zupan LH, Merfeld DM, Darlot C.** Using sensory weighting to model the influence of canal, otolith and visual cues on spatial orientation and eye movements. *Biol Cybern* 86: 209–230, 2002.