Combination of noisy directional visual and proprioceptive information

Sascha Serwe Department of Psychology, Justus Liebig University Giessen, Germany

Knut Drewing Department of Psychology, Justus Liebig University Giessen, Germany

Julia Trommershäuser Department of Psychology, Justus Liebig University Giessen, Germany

We present experimental and computational evidence for the estimation of visual and proprioceptive directional information during forward, visually driven arm movements. We presented noisy directional proprioceptive and visual stimuli simultaneously and in isolation midway during a pointing movement. Directional proprioceptive stimuli were created by brief force pulses, which varied in direction and were applied to the fingertip shortly after movement onset. Subjects indicated the perceived direction of the stimulus after each trial. We measured unimodal performance in trials in which we presented only the visual or only the proprioceptive stimulus. When we presented simultaneous but conflicting bimodal information, subjects' perceived direction fell in between the visual and proprioceptive directions. We find that the judged mean orientation matched the MLE predictions but did not show the expected improvement in reliability as compared to unimodal performance. We present an alternative model (probabilistic cue switching, PCS), which is consistent with our data. According to this model, subjects base their bimodal judgments on only one of two directional cues in a given trial, with relative choice probabilities proportional to the average stimulus reliability. These results suggest that subjects based their decision on a probability mixture of both modalities without integrating information across modalities.

Keywords: optimality, statistical decision theory, cue combination, direction estimation, multimodal perception, haptic perception

Citation: Serwe, S., Drewing, K., & Trommershäuser, J. (2009). Combination of noisy directional visual and proprioceptive information. *Journal of Vision, 9*(5):28, 1–14, http://journalofvision.org/9/5/28/, doi:10.1167/9.5.28.

Introduction

Human behavior benefits from a robust combination of sensory evidence across modalities. The human brain effortlessly processes mostly ambiguous sensory information and combines it into a stable and unambiguous percept of the world. It is an open question how the brain achieves this computation with such ease and efficiency. Recent behavioral evidence indicates that the different redundant cues available in a scene are combined and weighted according to their reliability, e.g., different depth cues such as texture or disparity are combined to form a consistent percept of depth in a given scene (Hillis, Ernst, Banks, & Landy, 2002; Hillis, Watt, Landy, & Banks, 2004; Landy, Maloney, Johnston, & Young, 1995). This rule also seems to hold for the integration of redundant sensory cues across modalities (Ernst & Banks, 2002). In the presence of inconsistent sensory cues with large differences in reliability, the sensory system recalibrates and reweighs sensory evidence taking the reliable cue as a standard (Atkins, Fiser, & Jacobs, 2001; Atkins, Jacobs, & Knill, 2003).

A variety of computational approaches has modeled the problem of optimal cue integration. One successful approach is to use Ideal Observer Models to define the optimal weighting and combination of redundant visual cues against which behavior of human observers can then be compared (e.g., Adams & Mamassian, 2004; Ernst & Bülthoff, 2004; Knill & Pouget, 2004; Knill & Richards, 1996; Landy et al., 1995; Whiteley & Sahani, 2008; Yuille & Kersten, 2006). This framework has also been successfully applied to cue integration across modalities, e.g., modeling the spatial integration of auditory and visual information (Alais & Burr, 2004), visual-haptic size perception (Ernst & Banks, 2002; Gepshtein & Banks, 2003), and recalibration of visual perception by haptic feedback (Adams, Graf, & Ernst, 2004; Atkins, Fiser, & Jacobs, 2001; Atkins, Jacobs, & Knill, 2003). The general idea behind Ideal Observer approaches is that the brain combines sensory cues to form the *most reliable* estimate of the state of the world, i.e., the estimate for which the variance of the combination of both cues is as low as possible. In the case of independent, Gaussian-distributed sensory cues, this combination rule corresponds to a weighted linear combination of the individual cues according to their

doi: 10.1167/9.5.28 Received September 18, 2008; published May 28, 2009 ISSN 1534-7362 © ARVO

relative reliabilities, i.e., cues with higher reliability contribute more to the combined percept.

However, there are known limitations of cue combination. Obviously, if two signals do not belong together they should not be combined into a common percept but be processed separately. This hypothesis is supported by the work of Körding, Beierholm, Ma, Quartz, Tenenbaum, and Shams (2007) who demonstrated that the likelihood of cue integration depends on the spatial offset of auditory and visual information. The data of Körding et al. (2007) fit the predictions of a causal inference model of cue integration; according to the causal inference model, two cues are only integrated if the nervous system infers that they belong together. This inference can be modeled probabilistically and is, for example, influenced by spatial proximity.

Here we asked whether the perception of noisy directional information during a pointing movement benefits from the combination of directional cues provided in more than one sensory modality. Vision and proprioception have previously been found to play a crucial role for the execution of goal-directed pointing movements (van Beers, Baraduc, & Wolpert, 2002). Proprioception has been demonstrated to be sufficient for the learning of the dynamics of force perturbations (Franklin, So, Burdet, & Kawato, 2007; Scheidt, Conditt, Secco, & Mussa-Ivaldi, 2005); additional visual feedback provided about hand or final target position prior to or during the movement increased the success rate, straightened the movement path, and reduced the overall movement error. We used a directional force pulse and a noisy directional visual stimulus to present redundant information. Most other cue-integration studies involve judgments of object shape, object size, or object surface properties where in the absence of obvious cue conflicts vision and haptics are intuitively perceived as belonging together. In contrast, we used cues that were related to each other just on an abstract symbolic level. These cues are normally not combined, but we created conditions that should allow for integration of information across senses. We presented noisy visual and proprioceptive cues that coincided spatially and temporally and both redundantly represent a stimulus direction that is to be judged.

There are several reasons to integrate information across senses in this situation. First, we told subjects that both cues represent the same direction and that they—though their task is to judge the force pulse direction in the bimodal situation—should use the visual information as an additional hint. Second, subjects trained the task in an environment where visual and proprioceptive information was completely redundant without any cue conflicts. In addition, during training feedback about the correct direction on a trial-by-trial basis emphasized correlations between visual and proprioceptive cues. Finally, in all conditions of our experiment, both visual and haptic stimuli are very unreliable and noisy on their own. Thus, the need for noise reduction should be obvious and the benefit of integration in terms of absolute noise reduction is very high. This benefit is maximized if the information is integrated using optimal weights. However, any other integration across senses using arbitrary suboptimal weights also results in an absolute reduction of noise as compared to unimodal situation.

We measured the perception of noisy directional information for unimodal cues presented in the visual and in the proprioceptive modality in isolation. We used these unimodal estimates to predict the perception of directional information in conditions with both consistent and inconsistent visual and proprioceptive cues. We find that the most common model of optimal cue integration (maximum likelihood estimation, MLE) successfully predicts the observed biases in mean perceived direction in trials with inconsistent visual and proprioceptive cues. Surprisingly, we find that the MLE model fails to account for the observed changes in reliability. While the MLE model predicts an increase in bimodal reliability of judgments relative to the unimodal reliability of judgments for both inconsistent and consistent visual and proprioceptive cues, we observe a decrease in reliability. We present an alternative model (probabilistic cue switching, PCS), which accounts for both the observed biases in mean direction and the observed change in reliability under bimodal conditions. These results suggest that subjects base their judgment in a given trial on only one of two possible cues with relative choice probabilities proportional to the individual cue reliabilities.

Materials and methods

Subjects

Six healthy right-handed subjects (3 women, 3 men, age 21–24 years) participated in this experiment. All were naive to the purpose of the experiment and were paid for their participation. All subjects had normal or corrected-to-normal vision.

Task

Subjects performed pointing movements toward a visually specified circular target within a visuo-haptic environment (Figure 1). Targets were always presented at a distance of 35 cm from the start position of the movement, centered in front of the body along the sagittal plane. Subjects initiated a trial by placing their fingertip at the start position; the start position was a visual-haptic platform at 20 cm below eye level and centered at 20 cm in front of the body. After resting at the start position for 400–600 ms (randomized, uniformly distributed), auditory and visual "go" signals (beep and display of the word "GO!") appeared. As long as the fingertip rested on the

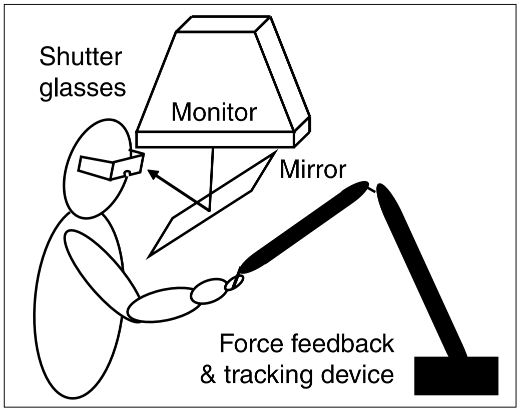


Figure 1. Experimental apparatus. Visual stimuli were displayed on a CRT screen suspended from above. Subjects viewed the stereoscopically displayed visual stimulus in a mirror using CrystalEyes liquid-crystal stereo glasses. A PHANToM 3.0 force-feedback device tracked the 3-d position of the right index finger and was used to apply the directional proprioceptive cue.

start position, the position of the fingertip was indicated by a red sphere. The whole visual scene disappeared upon movement onset, i.e., subjects performed the movement under open-loop conditions. Early in the movement, when the subject reached a trigger position 5 cm beyond the start position, a directional stimulus was presented for 50 ms. This stimulus was either a directional noisy visual stimulus, a directional proprioceptive stimulus (force pulse), or a combination of both (Figure 2). To constrain the range of possible movements, we facilitated accuracy as follows: After completion of the movement, subjects received feedback about whether they had successfully hit inside the target circle (visual display of the word "hit" or "miss"). If the subject responded too slowly (i.e., did not hit the wall or target within 1200 ms following the go signal) or too fast (i.e., initiated the movement prior to presentation of the go signal or within 100 ms after the go signal), feedback was provided (beep + "too fast", "too slow") and the trial was repeated later during the experiment. Only if subjects hit the target or the wall surrounding the target within the time limit they were allowed to do their main task: judging the direction of the perceived stimulus on a continuous scale by pointing at an annulus of 6-cm diameter presented at the same reaching distance as the target circle.

In the conditions in which both visual and proprioceptive information was presented, subjects were instructed to indicate the perceived direction of the proprioceptive cue. Subjects were told that in all trials the visual lines represent the direction of the proprioceptive cue. We also told our subjects that the lines were drawn from a distribution around the correct direction, i.e., that they therefore always contained some uncertainty and should only be considered as an additional hint. Note that even though this visual cue for the force pulse is of a rather abstract, symbolic nature, subjects were instructed and trained that it is a redundant cue and that it indicates the same direction as the force pulse. As long as cue conflicts are not noticed, this should motivate our subjects to integrate across the senses if possible. In all conditions of our experiment, both visual and haptic stimuli are very unreliable and similar in reliability. Therefore, the need for noise reduction and the possible benefit of integration is high.

Apparatus

Participants sat in front of a visuo-haptic setup in a dimly lit, quiet room (Figure 1). The apparatus consisted of a PHANToM 3.0 L haptic force-feedback device (temporal resolution = 1000 Hz, spatial resolution = ~0.03 mm, maximum force = 22 N, force feedback in the three translatory directions) and a 22″ computer screen (Iiyama Vision Master Pro 514, 120 Hz, 960 × 1280 pixels). The right index finger was connected to the PHANToM via a thimble-like holder, allowing for free movements with all six degrees of freedom in a workspace of about 40 × 60 × 80 cm^3 (approximately full arm movement pivoting at the shoulder). The haptic workspace of the PHANToM was spatially aligned with a 3-dimensional visual scene. The visual scene was created by viewing through stereoscopic shutter glasses (CrystalEyes 3 liquid-crystal shutter glasses, 120 Hz) onto a mirror that reflected the images on the computer screen. The shutter glasses were fixed relative to the screen location, and in addition, we used a chin rest to constrain head movements during the experiment. White noise presented via headphones masked possible sources of noise created by the mechanics of the PHANToM. The experiment was run on a PC (2.8 GHz; 1 GB RAM) using C++ code to control the apparatus, present the stimuli, and track the finger.

Stimuli

Proprioceptive stimulus

A weak force pulse of 1-N strength was applied to the index finger. The total force pulse duration was 50 ms including a 5-ms onset and 5-ms offset ramp (sinusoidal increase and decrease). Force direction was always orthogonal to the line connecting the start point and target; the force was therefore approximately orthogonal to the direction of the finger movement. Forces were applied in 8 directions, i.e., from above (0°), from the right (90°), from below (180°), from the left (270°), or from 45°, 135°, 225°, or 315°.

In the trials in which the proprioceptive cue was present, the applied force pulse led to a small deviation of the hand movement in the direction of the force pulse, i.e., opposite of the force pulse origin (see Figure 3). To

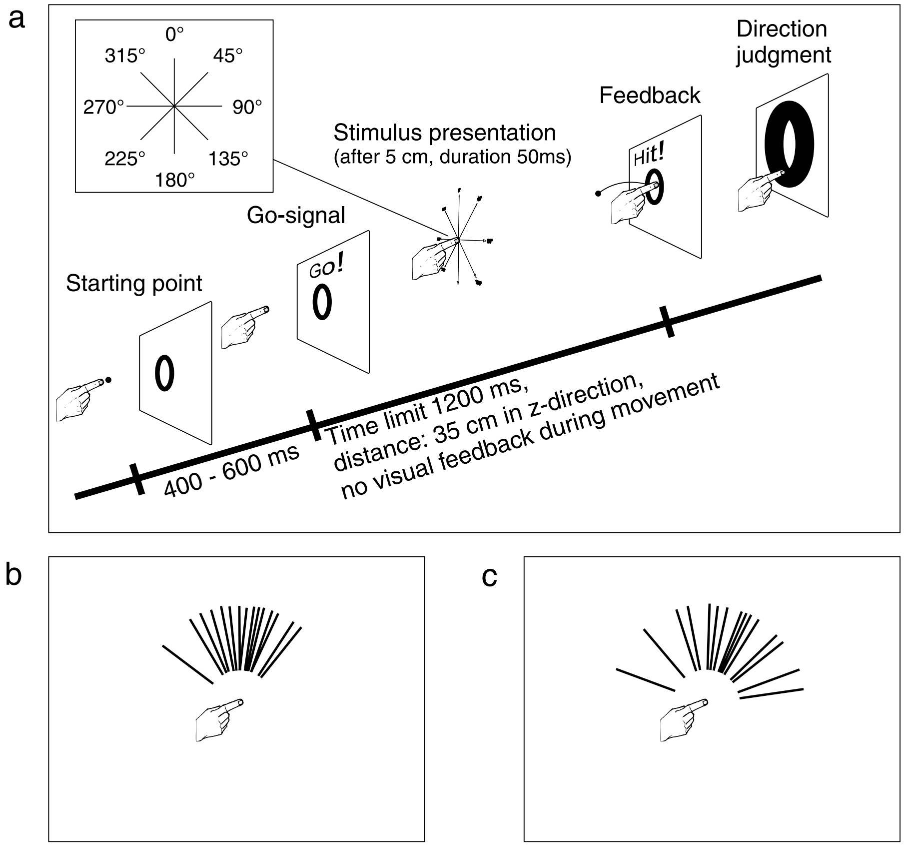


Figure 2. Time course of a trial and description of directional stimulus. (a) Time course of a trial. Subjects had to hit a circular target 35 cm away from their starting position within 1200 ms. After subjects covered a distance of 5 cm (see text for detailed description), either a visual or a proprioceptive directional stimulus or a combination of both stimuli was presented for 50 ms. The proprioceptive stimulus was a force pulse in one out of 8 possible directions toward the index finger applied via the PHANToM haptic interface. The visual stimulus was a set of 15 lines that were arranged on a virtual annulus around the index finger (see (b) and (c) for more details). In bimodal trials, the visual stimulus direction differed from the proprioceptive stimulus direction by −30°, 0°, or 30°. After completion of the movement, subjects received feedback about whether they hit inside the target circle. Then they judged the direction of the perceived stimulus by pointing at an annulus of 6-cm diameter presented at the same reaching distance as the target circle. Except for the training trials, there was no feedback about the correct stimulus direction. (b) Example for a visual stimulus in the V60 condition. The stimulus direction (in this example 0° or "from top") is chosen and 15 lines were arranged around this direction. The direction of each line is drawn from a Gaussian distribution with the stimulus direction as mean angle and a standard deviation of 60°. (c) Example for a visual stimulus in the V90 condition. Same as (b) except that the standard deviation of the underlying Gaussian distribution is 90°.

illustrate the effect of the applied force pulse, we measured the maximum vertical distance between the mean trajectory for force pulses from above (0°) as compared to the mean trajectory for force pulses from below (180°). The mean across subjects for this maximum distance in vertical direction was 27 mm (ranging from 19 mm to 37 mm) and was reached about 124 ms after force onset (ranging from 86 ms to 166 ms). For the maximum horizontal distance, i.e., for the mean trajectory for force pulses from right (90°) as compared to the mean trajectory for force pulses from the left (270°), the mean across subjects was slightly less, i.e., 16 mm (ranging from 12 mm to 22 mm for individual subjects) and was reached slightly earlier, i.e., after 85 ms (ranging from 66 ms to 106 ms for individual subjects).

Visual stimulus

Visual stimuli were temporally and spatially aligned with the proprioceptive stimulus to promote cue integration. That is, they were presented at the same distance from the start position as the proprioceptive stimulus and also approximately orthogonal to the movement direction. The center of the visual stimulus was matched to the position of the index finger at stimulus onset. Unlike the finger itself and the resulting force pulse position, the visual stimuli did not move during presentation. The stimulus consisted of 15 radial lines (width = 1 mm, length = 10 mm, distance from center randomized between 8 and 12 mm). Lines were sampled, by randomly drawing 15 times from a Gaussian around a mean direction with a standard deviation of either 60° (low visual noise,

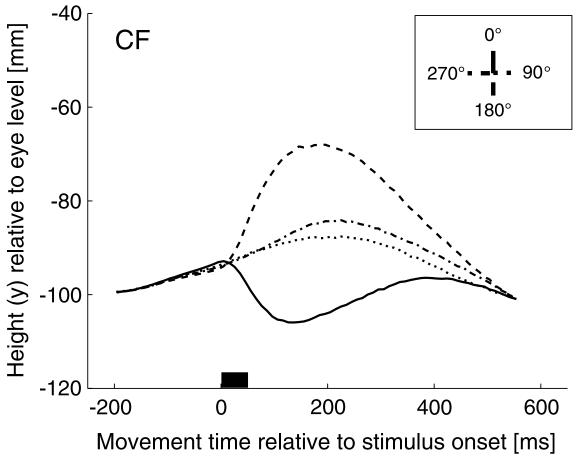


Figure 3. Side view on mean trajectories for one subject (CF) and four different force pulse directions (0°, 90°, 180°, 270°). All trajectories for one subject (here: CF) were aligned to force pulse onset and then averaged across stimulus directions to produce mean trajectories. The maximum vertical distance between the mean trajectory for force pulses from above (0°) as compared to force pulses from below (180°) was 37 mm for this subject and was reached 146 ms after force pulse onset. Start and target position was 100 mm below eye level. The black bar located at the x-axis represents the force pulse duration.

Figure 2b) or 90° (high visual noise, Figure 2c). We chose the standard deviation of the Gaussian such that unimodal visual judgments exhibited either a slightly lower or a slightly higher reliability as compared to the expected range of reliabilities for proprioceptive judgments.

Experimental sessions and design

All subjects performed an initial training session to get accustomed with the setup, the unimodal and bimodal tasks. During the training trials, subjects received feedback about the correct direction of the presented stimulus. There was no cue conflict in bimodal training trials. We presented a broad range of stimulus directions between 0° and 350° (with a step size of 10°) in the training trials to prevent the subjects from learning that the stimulus set that was later used in the experiment was limited. Subjects first repeated blocks of 88 visual trials until the mean absolute distance between perceived and correct directions was less than 30°. They then repeated blocks of proprioceptive trials until they reached the same criterion. One subject failed to learn the task (i.e., did not reach criterion) and was excluded from further participation in the main experiment. After learning and understanding both unimodal tasks, subjects trained in the bimodal condition. They ran one block of 88 trials with visual and proprioceptive information presented simultaneously and without cue conflict.

Following the training session, data were collected across five experimental sessions of 88 warm-up plus 480 experimental trials each. To further emphasize the close relationship between visual and proprioceptive stimuli, the warm-up sequences were identical to the bimodal training trials described above (no cue conflict, feedback about the correct stimulus direction). The experimental trials were also bimodal but without feedback about the correct stimulus direction. The proprioceptive stimulus pushed the finger in one out of eight directions (0°, 45°, 90°, 135°, 180°, 225°, 270°, or 315°). The visual stimulus direction either coincided with the proprioceptive stimulus direction or was rotated by 30° clockwise or 30° counterclockwise with respect to the proprioceptive stimulus, i.e., this constituted three conditions of cue conflict: −30°, 0°, and 30°. The visual stimulus differed in reliability, i.e., the lines were sampled around the mean direction with a standard deviation of either 60° (V60) or 90° (V90). This resulted in a total of 48 different types of trials. Each trial type was repeated 50 times resulting in a total of 2400 trials.

In the seventh and last session, subjects performed one block of unimodal proprioceptive trials and one block of unimodal visual trials. The proprioceptive stimuli were repeated 50 times for each of the eight directions (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°), resulting in a total of 400 trials. The visual stimulus with a standard deviation of 60° was shown in the same eight directions and was intermixed with the visual stimulus with a standard deviation of 90° (representative directions only: 0°, 90°, 180°, 270°). Each condition was repeated 50 times for a total of 600 trials.

To facilitate cue integration, we matched stimulus presentation in space and time for visual and proprioceptive stimulus and told subjects that the signals belong together; following data collection, we asked all subjects whether they had noticed a cue conflict and if so asked to estimate the number of trials in which they had noticed a cue conflict.

Model of optimal directional cue combination based on maximum likelihood estimation (MLE)

We next briefly explain how the predictions of the most commonly used model of cue combination, maximum likelihood estimation (MLE), apply to modeling the combination of noisy directional information. The general idea behind this model is that the brain combines sensory information such that the resulting estimate is as reliable as possible (e.g., Ernst and Banks, 2002; Ernst and Bülthoff, 2004; Landy et al., 1995). In the case of independent, Gaussian-distributed sensory cues, this combination rule corresponds to a weighted combination of estimates derived from the individual cues with weights relative to each estimate's reliability, i.e., its inverse variance.

In a single trial of our experiment, two sources of sensory information are available: the proprioceptive estimate of the direction $\hat{S}_P$ and the visual estimate of the direction $\hat{S}_V$. Both estimates are noisy and the amount of noise can be quantified by measuring the variance of answers toward repetitions of the same unimodal stimulus. We told our subjects that the visual cue represents the same direction as the proprioceptive cue and all subjects believed in this instruction until the end of the experiment. If both estimates are noisy estimates of the same direction, a less noisy combined estimate $\hat{S}_{MLE}$ can be computed by weighted averaging:

$$\hat{S}_{MLE} = w_P\hat{S}_P + w_V\hat{S}_V. \tag{1}$$

For two estimates with independent Gaussian noise, the variance of the combined estimate is minimized if each single cue estimate is weighted according to its relative reliability R, i.e., the relative inverse variance of the corresponding single cue estimates:

$$w_P = \frac{R_P}{R_P + R_V} = \frac{1/\sigma_P^2}{1\sigma_P^2 + 1/\sigma_V^2}, \tag{2}$$

and

$$w_V = \frac{R_V}{R_P + R_V} = \frac{1/\sigma_V^2}{1/\sigma_P^2 + 1/\sigma_V^2}. \tag{3}$$

Here, we assume that the noise distributions are independent. Both mean and variance of the resulting combined estimate can then be calculated for a given mean and variance of the unimodal estimates. The mean μ_{MLE} of the combined estimate is always between either of the two individual estimates, namely

$$\mu_{MLE} = w_P\mu_P + w_V\mu_V. \tag{4}$$

The variance of the combined estimate σ_{MLE}^2 is always less than either of the two individual variances, namely

$$\sigma_{MLE}^2 = w_P^2\sigma_P^2 + w_V^2\sigma_V^2. \tag{5}$$

Note that for our data the measured and predicted perceived angles are measured on a circular scale with values between 0° and 360°. One could use specialized models for circular data here, for example the wrapped normal distribution (Jammalamadaka, Sengupta, & Sengupta, 2001). However, for the range of variances tested here and for maximum cue differences of 30°, the wrapped normal distribution is nearly identical to the Gaussian distribution that we chose for simplification. In addition, note that the measure of interest is the angular difference between the true stimulus direction and the perceived stimulus direction. This difference is in principle periodic and varies from −180° to 180° where negative numbers indicate a deviation in counterclockwise direction and positive numbers indicate a clockwise deviation. For the range of conditions in our experiment, periodicity should not matter: we neither expected nor found absolute differences close to 180°, and therefore the fit of the unimodal data with a linear Gaussian is a good approximation.

Data analysis

Comparison with MLE predictions

To test whether subjects integrated visual and proprioceptive information as predicted by MLE, we computed MLE estimates according to Equations 2–4 based on the cue estimates from the unimodal conditions as follows: We measured the mean bias, i.e., the mean difference between the visually presented stimulus direction, the judged (as a measure of perceived) stimulus direction μ_V, and the judgment variability (as a measure of perceptual variability) σ_V^2 in visual-only trials, and respectively, μ_P and σ_P^2 of judged directions in proprioceptive-only trials. The unimodal performance was computed for each subject individually for the proprioceptive stimulus and both visual reliabilities (V60 vs. V90), averaged across all stimulus directions. Based on these unimodal data estimates, we computed the expected bimodal bias μ_{MLE} (Equation 3) relative to the presented proprioceptive stimulus direction and the expected variability σ_{MLE}^2 (Equation 4) of perceived angles for each cue conflict (−30°, 0°, 30°) and for both conditions of visual reliability (V60 vs. V90), individually for each subject. We then compared the computed MLE predictions, μ_{MLE} and σ_{MLE}^2 for all subjects and conditions with the observed data in the corresponding bimodal trials, μ_{VP} and σ_{VP}^2.

Comparison with PCS predictions

Because the results were not fully consistent with the MLE predictions, we tested post hoc whether subjects integrated visual and proprioceptive information according to an alternate, multisensory probabilistic choice model. This model (probabilistic cue switching, PCS) makes two basic assumptions. First, in a single bimodal trial subjects base their directional judgment on either the visual estimate or the proprioceptive estimate. Second, subjects do not choose randomly between the available estimates but choose the more reliable estimate more often. Exact choice probabilities are based on the individual relative reliabilities of each estimate. We computed PCS estimates as follows: Exact choice probabilities are determined by the individual relative reliabilities of each estimate, i.e., the weights w_i in Equation 2 denote the choice probability $p(i)$ for PCS. Thus, in every single trial, the subject perceives the stimulus either in the direction of the visual estimate with probability $p(V) = w_V$ and will have the visual bias μ_V and variance σ_V^2; or the subject perceives the stimulus in the direction of the proprioceptive estimate with probability $p(P) = w_P = 1 - p(V)$ and will have the proprioceptive bias μ_P and variance σ_P^2. We modeled PCS by computing unimodal estimates for each subject individually based on the data measured in the unimodal conditions. We sampled 100,000 times either from a Gaussian with μ_V and σ_V with probability $p(V) = w_V$, or from a Gaussian with μ_P and σ_P with probability $p(P) = w_P$. The mean and standard deviation of the resulting distribution is the prediction of the PCS model, i.e., μ_{PCS} and σ_{PCS}. We then compared the computed PCS predictions, μ_{PCS} and σ_{PCS} for all subjects and conditions with the observed data in the corresponding bimodal trials, μ_{VP} and σ_{VP}.

Results

Combination of visual and proprioceptive directional information

Figure 4 shows each of the six subject's mean bias for the 2 visual and the proprioceptive unimodal stimuli (Figure 4a) as well as the corresponding standard deviation of reported directions (Figure 4b). Subjects correctly reported the direction of both the visual and proprioceptive cue with only small unsystematic biases (Figure 4a). For 5 out of 6 subjects, the variance of reported directions increased from the visual stimulus with 60° standard deviation of lines (V60) through the proprioceptive stimulus toward the visual stimulus with 90° standard deviation of lines (V90, Figure 4b). Thus, in the bimodal conditions, these subjects should weight the proprioceptive estimate more than the visual estimate if combined with the visual stimulus with high noise (V90) and less than the visual estimate if combined with the visual stimulus with low noise (V60).

Figure 5 shows the mean perceived angle relative to the force pulse direction for all cue conflicts and both visual noise levels in the bimodal condition. Without cue conflict, there is no deviation from the applied force pulse direction ($p > 0.5$). For any cue conflict situation of either 30° or −30°, the mean perceived angle differs from the force pulse direction for both visual noise levels (all $ps < 0.05$). For V90, they also differ significantly from the visual stimulus direction in conflict trials (both $p < 0.01$). For V60, they do not differ significantly from the visual stimulus direction but show a strong trend toward the force pulse direction ($p = 0.10$ for −30° and $p = 0.12$ for 30°). This suggests that subjects rely on both visual and proprioceptive information in bimodal trials. Furthermore, the V60 trials are closer to the visually perceived direction than the V90 trials. Thus, subjects seem to adjust their weights relative to the visual cue reliability.

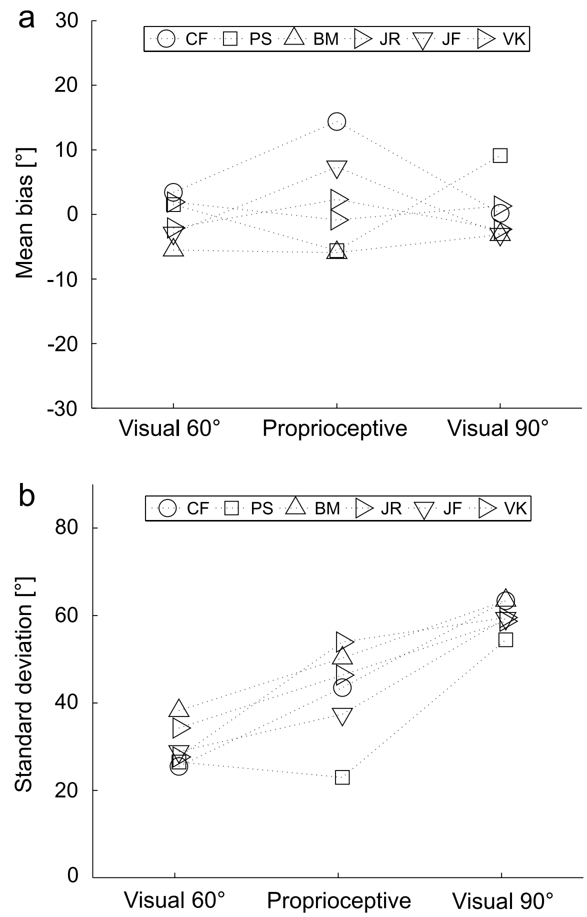


Figure 4. Unimodal biases and reliabilities. The unimodal bias (a) is the difference between the true stimulus direction and the mean perceived stimulus direction that is close to zero for all subjects and both modalities. The standard deviations σ of perceived angles (b) are lowest for the better visual stimulus with 60° spread of lines, slightly higher for the proprioceptive stimulus and highest for the visual stimulus with 90° spread of lines. Data are averaged across all directions for each subject.

However, as described in the data analysis section, we predicted each subject's bimodal biases (Equation 4) and reliabilities (Equation 5) quantitatively using his/her unimodal data. Model predictions were computed for each subject individually, using each subject's individual estimates of visual bias and reliability and proprioceptive bias and reliability. We predicted bimodal bias and reliability for each cue condition (differences of −30°, 0°, 30°) and for both conditions of visual reliability (V60 vs. V90), pooled across all eight force pulse directions. Figure 6a shows the predictions of the mean direction of

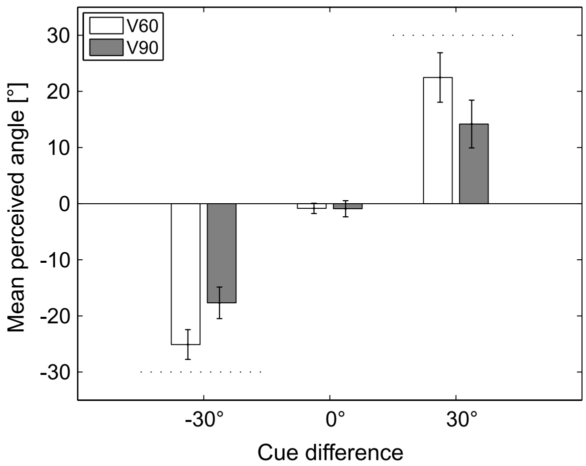


Figure 5. Mean perceived angle for the different visual conditions across all subjects (N = 6). For cue differences, the bimodal perceived directions are between the proprioceptively presented direction (set to 0°) and the visually presented direction (either 30° or −30°). Without cue difference, the bimodal perceived direction does not differ from the presented directions. Consistent with reliability-based weighting for the more reliable visual stimuli (black bars), the bimodal percept is closer to the visual presented direction than for the less reliable visual stimuli (white bars).

the combined bimodal estimate as predicted by the MLE model (μ_{MLE}) compared to the observed direction judgments in the corresponding bimodal condition μ_{VP}. A linear regression of the MLE prediction fitted to the observed bimodal data yielded an R squared correlation of R^2 = 0.90. However, the estimated slope of the linear correlation between MLE prediction and observed data was 1.25 (confidence interval from 1.10 to 1.39), i.e., larger than 1, indicating that subjects shifted slightly more toward the visual cue than predicted by MLE.

In addition to the mean perceived angle, the MLE model predicts an increase in reliability of bimodal judgments as compared to unimodal judgments. An increase in reliability corresponds to a reduction in the standard deviation of the directional judgments. Figure 6b shows the expected standard deviation according to MLE and the measured standard deviation of the perceived directions in the corresponding bimodal conditions. As

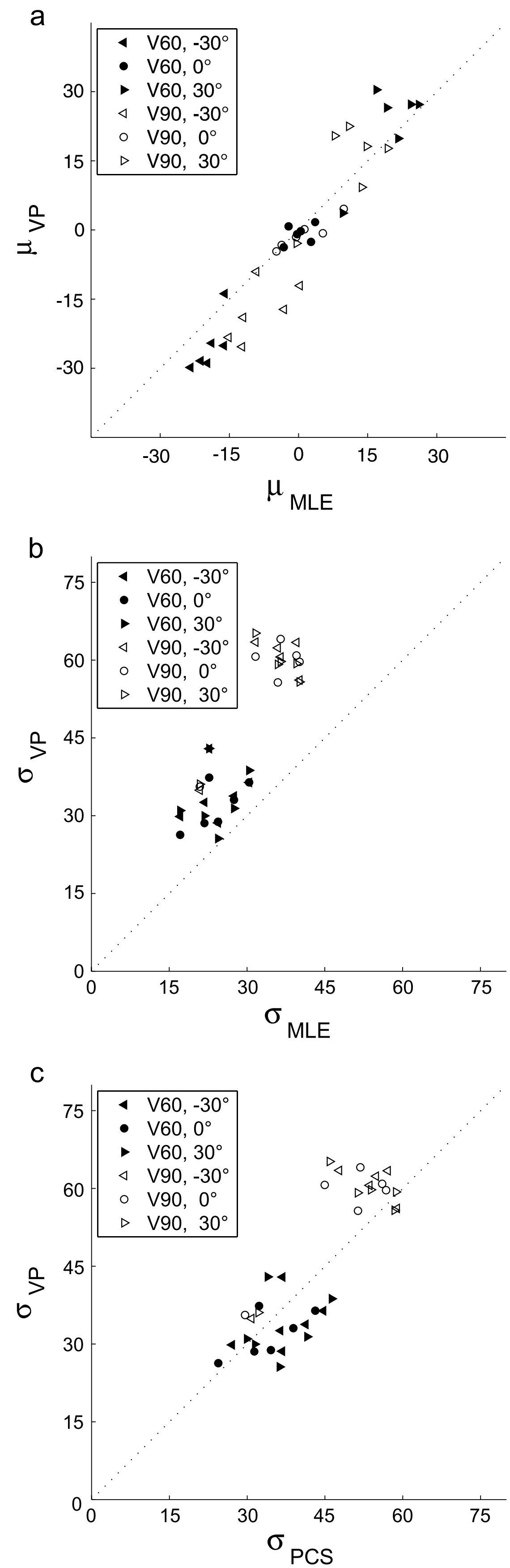


Figure 6. Comparison between model prediction and data. For every subject and every condition, a prediction for the combined multimodal estimate was computed on the basis of single cue data, using each subject's own average estimates for each cue condition and for both conditions of visual reliability. (a) Predicted mean perceived directions μ_{MLE} relative to the proprioceptive stimulus direction (0°) compared to mean perceived direction in the bimodal condition μ_{VP}. The prediction is identical for MLE and PCS. The observed standard deviation σ_{VP} is compared to the corresponding predictions of MLE in (b) and PCS in (c). Each data point represents one out of 6 subjects in 1 out of 6 conditions (2 visual noise levels, 3 cue differences).

obvious from Figure 6b, the MLE model systematically underestimates the measured standard deviation. The deviations from the MLE prediction are particularly evident in the trials with large visual stimulus variability (open symbols in Figure 6b). Subjects do not exhibit the expected benefit in performance in the bimodal conditions, rather the opposite—the directional judgments are even more variable in the bimodal conditions than in the more reliable of the two unimodal conditions (average estimates across all subjects and cue difference conditions; for V60: bimodal variability estimate $\sigma_{VP} = 33°$, unimodal variability estimate $\sigma_V = 30°$ and $\sigma_P = 42°$; for V90: bimodal variability $\sigma_{VP} = 56°$, unimodal variability estimates, $\sigma_V = 60°$ and $\sigma_P = 42°$; see also Figure 2b). A linear regression of the MLE prediction fitted to the observed reliability data yielded an R squared correlation of $R^2 = 0.72$. However, the estimated slope of the linear correlation between MLE prediction and observed data was 1.56 (95% confidence interval from 1.23 to 1.92), i.e., larger than 1, indicating that subjects' variability was higher than predicted by MLE. Taken together, the results for mean and variability of judgments in the bimodal conditions indicate that subjects seem to combine visual and proprioceptive information somehow and adjust the weights relative to stimulus reliability. However, they do not show the expected increase in reliability in bimodal as compared to unimodal conditions as predicted by the MLE model.

Probabilistic cue switching

The probabilistic cue switching model (PCS, see data analysis section for more details) is a model that can describe the observed change in mean perceived direction and predicts an increase in variability in the bimodal conditions. According to this model, subjects base their decision in a single trial either on the visual cue or on the proprioceptive cue but do not integrate information across senses. The choice probabilities for each modality match the relative reliabilities and hence the weights in the MLE model. Thus for repeated trials, this model yields the same prediction for mean direction as the MLE model: the MLE weights for the visual cue and the proprioceptive cue are the same as the corresponding choice probabilities for PCS. The difference between the two models is that MLE predicts that a weighted average is built in every single trial, whereas for PCS a single trial is the result of a probabilistic choice. The predicted mean direction is the average across multiple trials and therefore identical if the weights are used on a trial-by-trial basis as for MLE, or if the weights influence the mixture of the whole set of trials as for PCS ($\mu_{PCS} = \mu_{MLE}$). However, if every single trial is a weighted average, the expected deviation of a single trial from the mean direction is lower, i.e., averaged over all trials, the predicted variability is higher for PCS than for MLE ($\sigma_{PCS} > \sigma_{MLE}$). More precise, whereas the reliability for the MLE estimate is *better* than or *at least* as reliable as the most reliable unimodal estimate, the combined estimate for PCS is expected to be *worse than* or *at most* as reliable as the most reliable unimodal estimate.

We compare the predicted PCS variabilities σ_{PCS} in the bimodal condition to the observed bimodal variabilities σ_{VP} in Figure 6c. In contrast to the variability predictions made by the MLE model (Figure 6b), our fit of the PCS model cannot be rejected in the bimodal conditions. A linear regression of the PCS prediction fitted to the observed reliability data yielded an R squared correlation of $R^2 = 0.72$. However, the estimated slope of the linear correlation between MLE prediction and observed data was 1.14 (95% confidence interval from 0.89 to 1.39), i.e., it includes 1. Though there is still unexplained variance in the data, the PCS predictions fit to the data clearly better than the MLE predictions.

Effect of force pulse direction

As shown in Table 1, biases and variability of perceived proprioceptive directions vary considerably not only between subjects but as well within subjects for different directions of the force pulse. An optimal observer should in principle use this information and adjust the weights according to the reliability of the directional force pulse in a given single trial (Equation 2). However, in the analysis presented above, we assumed for simplicity that the variability of the proprioceptive estimate is the same for all directions. That means that we used the averaged biases and reliabilities reported in the last column of Table 1. To account for the observed differences in force pulse reliability and to check whether a direction specific MLE or PCS model fits our data, we repeated the model comparisons for MLE and PCS in the bimodal case, but with the eight different proprioceptive biases and weights for the eight different force pulse directions that are also reported in Table 1. Figure 7 shows the results of this more detailed and direction specific analysis. We still assume that the visual estimate is the same for all directions. In fact, a few subjects showed significant differences (subjects CF, BM, JF, and VK for V60 and JR, JF, and VK in V90). However, these were small in magnitude. The model computations were based on 50 observed trials in each of the 48 conditions per subject (eight directions, two visual cue reliabilities, three levels of cue conflict).

As shown in Figure 7, the PCS model is a better predictor of the perceived direction than the MLE model. The predicted mean direction in the bimodal conditions is the same for both PCS and MLE models and accounts well for the observed mean direction (Figure 7a). A linear regression of the MLE prediction fitted to the measured bimodal mean direction yielded an R squared of $R^2 = 0.6$. The estimated slope of the linear correlation between MLE prediction and observed data across subjects and

Subject		Force pulse direction								Pooled across directions
		0°	45°	90°	135°	180°	225°	270°	315°	
CF	μ	13.7	10.8	12.4	31.5	15.8	3.5	−1.9	33.1	14.4
	σ	25.3	38.7	40.5	30.3	24.9	60.0	58.4	50.9	43.5
PS	μ	4.1	6.7	10.2	4.5	−6.1	−24.1	−17.6	−16.8	−5.6
	σ	14.2	24.6	12.9	18.7	10.9	13.1	13.9	32.2	22.6
BM	μ	−9.4	−4.2	−1.9	4.3	6.4	−21	−6.7	−13.3	−5.9
	σ	47	48.9	54.3	64.2	28.3	35.3	48.7	64.8	50.4
JR	μ	0.8	22.3	22.4	33.8	−2.4	−16.3	−25.5	−6.3	2.3
	σ	52.7	76.2	47.3	36.1	40.4	43	31.4	61.3	54.1
JF	μ	7.8	3.7	5.3	34.5	−0.1	−19.4	2.1	30	7.4
	σ	37	42.8	26.1	20.6	17.8	15.2	41.4	36.8	37.3
VK	μ	−2.9	−10.9	2.4	8.9	−14.4	−26	13.7	11.7	−0.8
	σ	30.3	43.1	43.1	21.1	28.4	67.7	54.7	46	46.4

Table 1. Individual parameters of perceived direction following application of the directional force perturbation (mean bias μ and variability σ, measured in degrees).

conditions was 1.07 (95% confidence interval from 0.97 to 1.17), i.e., overall the MLE prediction neither over- nor underestimated the shift in perceived mean direction. The comparison of predicted and observed variabilities shows that the variability in the bimodal conditions does not exhibit the expected decrease as predicted by MLE (Figure 7b, $R^2 = 0.49$, slope 1.31, 95% confidence interval from 1.16 to 1.47) but closely matches the PCS predictions (Figure 7c, $R^2 = 0.49$, slope 0.93, 95% confidence interval from 0.81 to 1.04), though again the amount of unexplained variance or noise is large and similar for both models.

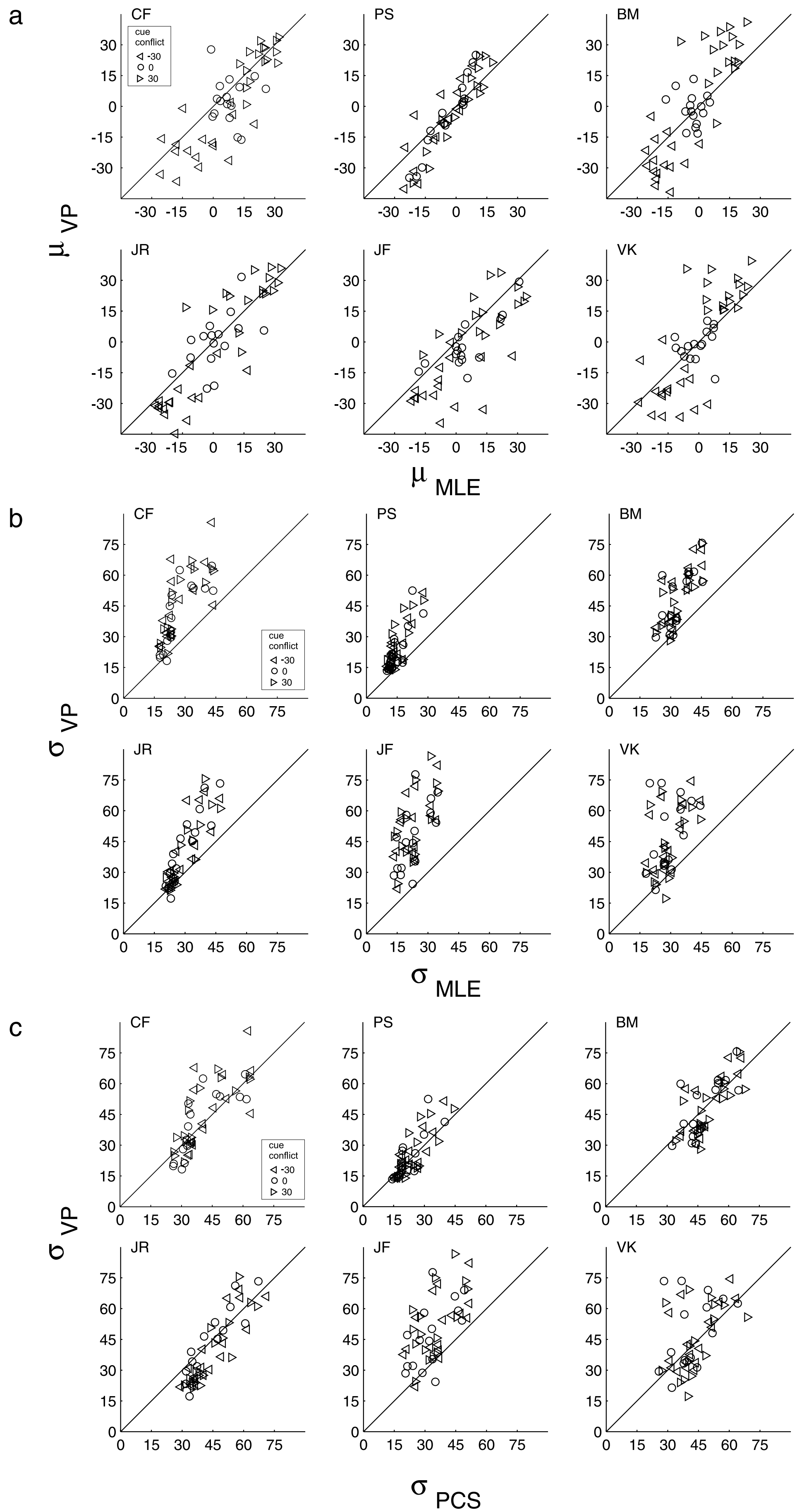
Discussion

In the study presented here, we asked whether the perception of noisy directional information during a pointing movement benefits from the combination of cues provided in more than one sensory modality. Unlike other cue-integration studies, the cues in our study were related to each other on an abstract symbolic level. These cues are normally not combined, but we created conditions that should allow for integration of information across senses. Visual and haptic cues coincided spatially and temporally and our subjects received instructions and learned during training that they belong together.

Both cues were unreliable on their own such that any integration of information would result in a significant reduction of noise as a direct consequence of the averaging even if suboptimal weights were used. This applies to any two cues, even if they do not provide entirely redundant information. The reduction is maximal if the weight is chosen according to MLE. However, any fixed weight will give some value between the optimum and the value for the less reliable cue. Only if the weight is not fixed but switches from trial to trial (as in our PCS model between zero and one) the error can be larger (see, e.g., Brenner, Granzier, & Smeets, 2007, for a simulation of unfixed weights in a color matching task).

However, we find that subjects consistently failed to integrate visual and proprioceptive information as expected according to the predictions of Maximum Likelihood Estimation (MLE), the most common model of cue integration (e.g., Ernst and Banks, 2002; Ernst and Bülthoff, 2004; Gepshtein, Burge, Ernst, & Banks, 2005; Helbig and Ernst, 2007; Hillis et al., 2002, 2004; Knill and Saunders, 2003; Landy et al., 1995). While simultaneous presentation of inconsistent visual and proprioceptive information biased our subjects' mean judgments in a manner consistent with MLE, subjects never reached the expected gain in reliability predicted by MLE. This is surprising, especially since the biased mean judgments across trials suggested close to optimal weights. Further-

Figure 7. Comparison between model prediction and data with direction-dependent model. For every subject (N = 6), we modeled all 8 force pulse directions independently using each subject's own average estimates of bias and reliability for every force pulse direction (see Table 1). The visual reliability was still assumed to be constant across directions. For each force pulse direction, both visual reliabilities (V60 vs. V90), and each cue conflict (−30°, 0°, 30°), we predicted the bimodal mean answer and standard deviation on the basis of single cue data for both modalities. (a) Predicted mean answer μ_{MLE} compared to mean perceived angle in the bimodal condition μ_{VP} (prediction is identical for MLE and PCS). The observed standard deviation σ_{VP} is compared to the corresponding predictions of MLE in (b) and PCS in (c). Each data point represents one out of 48 conditions (8 force pulse directions, 2 visual noise levels, 3 cue differences), each symbol represents one of the three cue differences (−30°, 0°, 30°) and each graph represents one of six subjects.


a
CF
PS
BM
JR
JF
VK
cue conflict
-30
0
30
μ_{VP}
μ_{MLE}
b
σ_{VP}
σ_{MLE}
c
σ_{VP}
σ_{PCS}

more, reliability is not only suboptimal but seems to decrease as compared to the most reliable unimodal situation.

The results of our post-hoc analysis suggest that subjects did not integrate visual and proprioceptive information, but in a single trial based their judgment on either the visual or the proprioceptive estimate. Performance as observed in our experiment was consistent with the predictions of a model of probabilistic cue switching (PCS), which assumes that on every trial subjects reported the perceived visual direction or the perceived proprioceptive direction according to choice probabilities identical to the corresponding relative weights for MLE. As a result, the PCS model predicts the same mean perceived direction as the MLE model but yields a higher overall variance for the combined estimate, i.e., a less reliable combined estimate. This model fits our data well.

There are several possible reasons why cue integration according to statistically optimal MLE was not observed in our task. Unlike most other cue-integration studies, the stimuli in our task do not belong together naturally, i.e., do not originate from the same object and are normally not combined. The work by Körding et al. (2007) has demonstrated that humans can efficiently infer the causal structure of multisensory events and either integrate information across senses or process multisensory information independently. Following this argument, a missing obvious causal structure, as in our study, would hinder cue integration. On the other hand, the pure knowledge of a joint causal structure can be sufficient for cue integration: if subjects know that vision and touch provide redundant information about the same object, visual and haptic shape information is integrated despite spatial discrepancies (Helbig & Ernst, 2007). However, in the study by Helbig and Ernst, the knowledge about a joint causal structure was most likely due to the perceptual-motor coherence and it is unclear whether higher level cognitive knowledge is able to produce the same results. In a similar study, Ernst (2007) showed that subjects learned experimentally introduced artificial correlations between stiffness and luminance in a multisensory discrimination task; this demonstrates that it is possible to change the likelihood of integration of two arbitrary sensory signals by manipulating their statistical co-occurrence. Deviations from statistically optimal MLE cue integration have previously been observed under experimental conditions in which the single cue estimates were correlated (Oruç, Maloney, & Landy, 2003; Rosas, Wichmann, & Wagemans, 2007), in which subjects did not have access to the single cue reliability on a trial-by-trial basis (Rosas, Wagemans, Ernst, & Wichmann, 2005), and with increasing spatial discrepancy (Gepshtein et al., 2005).

Even if our results suggest that our subjects do not integrate visual and proprioceptive information according to MLE, one might still wonder why subjects use a strategy like PCS. If subjects have access to individual cue reliabilities, why do they not choose the more reliable cue all the time? At this point, we can only speculate about the reasons. One idea is that the subjects' estimates of cue reliability in itself might be noisy and therefore the less reliable cue might sometimes appear to be the more reliable. Subjects likely will not have access to this additional "decision noise" on a trial-by-trial basis but choose the more reliable cue estimate more often on average—as predicted by the PCS model.

It is worth noting that—as intended by our instruction—neither of our subjects reported to have noticed the cue conflicts. We asked our subjects after the experiment whether they had noticed a cue conflict and also asked them for an estimate of the number of trials in which they had noticed a cue conflict. Most subjects reported to not have perceived a cue conflict at all. Those two subjects who reported to have perceived a cue conflict reported a very small number of trials, i.e., maximum of 50 out of the total of 2400 trials (i.e., 3%).

We conclude that the simultaneous presentation of noisy visual and proprioceptive information does not automatically lead to information integration that follows statistically optimal principles as predicted by the Maximum Likelihood Estimation. Even though the mean perceived direction in our experiment suggests a weighted combination of information, this is most likely a result of strictly not integrating information. The probabilistic cue switching model is not new in the context of cue integration (e.g., Ernst & Bülthoff, 2004; Landy & Kojima, 2001; Rosas et al., 2005). The model is usually explained as a possible alternative to MLE that one wants to exclude explicitly. This is done by showing the increased reliability that is only predicted by MLE. Our data show that this test is indeed a necessary test for MLE since we found experimental evidence for a strategy that was so far only discussed for theoretical reasons.

Acknowledgments

We thank Tim Schönwetter for help with data collection and Marty Banks and two anonymous reviewers for helpful comments on a previous version of the manuscript. This work is funded by the Deutsche Forschungsgemeinschaft (Emmy-Noether-Programm, TR, 528/1-2, 1-3).

Commercial relationships: none.
Corresponding author: Sascha Serwe.
Email: Sascha.Serwe@psychol.uni-giessen.de.
Address: Justus Liebig University, Department of General Psychology, Otto-Behaghel-Str. 10F, Giessen, Germany.

References

Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the "light-from-above" prior. *Nature Neuroscience, 7,* 1057–1058. [PubMed]

Adams, W. J., & Mamassian, P. (2004). Bayesian combination of ambiguous shape cues. *Journal of Vision, 4*(10):7, 921–929, http://journalofvision.org/4/10/7/, doi:10.1167/4.10.7. [PubMed] [Article]

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14,* 257–262. [PubMed] [Article]

Atkins, J. E., Fiser, J., & Jacobs, R. A. (2001). Experience-dependent visual cue integration based on consistencies between visual and haptic percepts. *Vision Research, 41,* 449–461. [PubMed]

Atkins, J. E., Jacobs, R. A., & Knill, D. C. (2003). Experience-dependent visual cue recalibration based on discrepancies between visual and haptic percepts. *Vision Research, 43,* 2603–2613. [PubMed]

Brenner, E., Granzier, J. J., & Smeets, J. B. (2007). Combining local and global contributions to perceived colour: An analysis of the variability in symmetric and asymmetric colour matching. *Vision Research, 47,* 114–125. [PubMed]

Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision, 7*(5):7, 1–14, http://journalofvision.org/7/5/7/, doi:10.1167/7.5.7. [PubMed] [Article]

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415,* 429–433. [PubMed]

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, 8,* 162–169. [PubMed]

Franklin, D. W., So, U., Burdet, E., & Kawato, M. (2007). Visual feedback is not necessary for the learning of novel dynamics. *PLoS ONE, 2,* e1336. [PubMed] [Article]

Gepshtein, S., & Banks, M. S. (2003). Viewing geometry determines how vision and haptics combine in size perception. *Current Biology, 13,* 483–488. [PubMed] [Article]

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://journalofvision.org/5/11/7/, doi:10.1167/5.11.7. [PubMed] [Article]

Helbig, H. B., & Ernst, M. O. (2007). Optimal integration of shape information from vision and touch. *Experimental Brain Research, 179,* 595–606. [PubMed]

Hillis, J. M., Ernst, M. O., Banks, M. S., & Landy, M. S. (2002). Combining sensory information: Mandatory fusion within, but not between, senses. *Science, 298,* 1627–1630. [PubMed]

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision, 4*(12):1, 967–992, http://journalofvision.org/4/12/1/, doi:10.1167/4.12.1. [PubMed] [Article]

Jammalamadaka, S. R., Sengupta, A., & Sengupta, A. (2001). *Topics in circular statistics*. Singapore: World Scientific.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences, 27,* 712–719. [PubMed]

Knill, D. C., & Richards, W. (1996). *Perception as Bayesian inference*. Cambridge, UK: Cambridge University Press.

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research, 43,* 2539–2558. [PubMed]

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. [PubMed] [Article]

Landy, M. S., & Kojima, H. (2001). Ideal cue combination for localizing texture-defined edges. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 18,* 2307–2320. [PubMed]

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research, 35,* 389–412. [PubMed]

Oruç, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research, 43,* 2451–2468. [PubMed]

Rosas, P., Wagemans, J., Ernst, M. O., & Wichmann, F. A. (2005). Texture and haptic cues in slant discrimination: Reliability-based cue weighting without statistically optimal cue combination. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 22,* 801–809. [PubMed]

Rosas, P., Wichmann, F. A., & Wagemans, J. (2007). Texture and object motion in slant discrimination: Failure of reliability-based weighting of cues may be evidence for strong fusion. *Journal of Vision, 7*(6):3, 1–21, http://journalofvision.org/7/6/3/, doi:10.1167/7.6.3. [PubMed] [Article]

Scheidt, R. A., Conditt, M. A., Secco, E. L., & Mussa-Ivaldi, F. A. (2005). Interaction of visual and proprioceptive feedback during adaptation of human reaching movements. *Journal of Neurophysiology, 93,* 3200–3213. [PubMed] [Article]

van Beers, R. J., Baraduc, P., & Wolpert, D. M. (2002). Role of uncertainty in sensorimotor control. *Philosophical Transactions of the Royal Society of*

London B: Biological Sciences, 357, 1137–1145. [PubMed] [Article]

Whiteley, L., & Sahani, M. (2008). Implicit knowledge of visual uncertainty guides decisions with asymmetric outcomes. *Journal of Vision, 8*(3):2, 1–15, http://journalofvision.org/8/3/2, doi:10.1167/8.3.2. [PubMed] [Article]

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: Analysis by synthesis? *Trends in Cognitive Sciences, 10,* 301–308. [PubMed]