nature neuroscience

Neural correlates of reliability-based cue weighting during multisensory integration

Christopher R Fetsch[1], Alexandre Pouget[2,3], Gregory C DeAngelis[1,2,5] & Dora E Angelaki[1,4,5]

Integration of multiple sensory cues is essential for precise and accurate perception and behavioral performance, yet the reliability of sensory signals can vary across modalities and viewing conditions. Human observers typically employ the optimal strategy of weighting each cue in proportion to its reliability, but the neural basis of this computation remains poorly understood. We trained monkeys to perform a heading discrimination task from visual and vestibular cues, varying cue reliability randomly. The monkeys appropriately placed greater weight on the more reliable cue, and population decoding of neural responses in the dorsal medial superior temporal area closely predicted behavioral cue weighting, including modest deviations from optimality. We found that the mathematical combination of visual and vestibular inputs by single neurons is generally consistent with recent theories of optimal probabilistic computation in neural circuits. These results provide direct evidence for a neural mechanism mediating a simple and widespread form of statistical inference.

The information received through our senses is inherently probabilistic, and an important challenge for the brain is to construct an accurate representation of the world despite this uncertainty. This problem is particularly relevant when considering the integration of multiple sensory cues, as the reliability associated with each cue can vary rapidly and unpredictably. Numerous psychophysical studies[1–8] have shown that human observers combine cues by weighting them in proportion to their reliability, consistent with statistically optimal (for example, Bayesian or maximum-likelihood) schemes. This solution is optimal because it generates unbiased perceptual estimates with the lowest possible variance[9,10], leading to an improvement in psychophysical performance beyond what can be achieved with cue alone or with any *ad hoc* weighting scheme.

The neural basis of optimal cue integration is not well understood, in part because of a lack of neurophysiological measurements in behaving animals. Recently, we developed a behavioral task in which macaque monkeys discriminate their direction of translational self-motion (heading) from visual and/or vestibular cues[11]. We found that monkeys can integrate these cues to improve psychophysical performance, which is consistent with a key prediction of optimal cue integration models. We also identified a population of multimodal neurons in the dorsal medial superior temporal area (MSTd) that likely form part of the neuronal substrate for this integration. MSTd neurons with congruent visual and vestibular heading tuning show increased sensitivity during presentation of multimodal stimuli, analogous to the perceptual improvement[11]. Our findings also revealed significant trial-by-trial correlations between neuronal activity and monkeys' perceptual decisions (choice probabilities), suggesting a functional link between MSTd activity and performance in this task[11,12].

This earlier study[11] revealed neural correlates of increased sensitivity during cue integration (with cue reliabilities fixed), but did not explore how multisensory neurons could account for the second prediction of optimal cue integration: weighting cues according to their relative reliabilities. Behaviorally, reliability-based cue weighting can be measured by introducing a small conflict between cues and examining the extent to which subjects' perceptual choices favor the more reliable cue[1–6]. Using this approach, we showed that monkeys, similar to humans, can reweight cues (that is, adjust weights from trial to trial) according to their reliability in a near-optimal fashion[13]. We then addressed two fundamental questions regarding the neural basis of reliability-based cue weighting. First, can the activity of a population of multisensory neurons predict behavioral reweighting of cues as reliability varies? We found that a simple decoding of MSTd neuronal activity, recorded during a cue-conflict task, can account fairly well for behavioral reweighting, including some modest deviations from optimality and individual differences between subjects. Second, what mathematical operations need to take place in single neurons, such that a simple population code can account for behavioral cue reweighting? We found that neurons combined their inputs linearly with weights dependent on cue reliability in a manner that is broadly consistent with the theory of probabilistic population codes[14]. These findings establish for the first time, to the best of our knowledge, a link between empirical observations of multisensory integration in single neurons and optimal cue integration at the level of behavior.

RESULTS

Theoretical predictions and behavioral performance

We begin by outlining the key predictions of optimal cue integration theory and how we tested these predictions in behaving monkeys.

[1]Department of Anatomy and Neurobiology, Washington University School of Medicine, Saint Louis, Missouri, USA. [2]Department of Brain and Cognitive Sciences, University of Rochester, Rochester, New York, USA. [3]Department of Basic Neuroscience, University of Geneva, Geneva, Switzerland. [4]Department of Neuroscience, Baylor College of Medicine, Houston, Texas, USA. [5]These authors contributed equally to this work. Correspondence should be addressed to C.R.F. (crfetsch@u.washington.edu).

Received 23 June; accepted 18 October; published online 20 November 2011; doi:10.1038/nn.2983

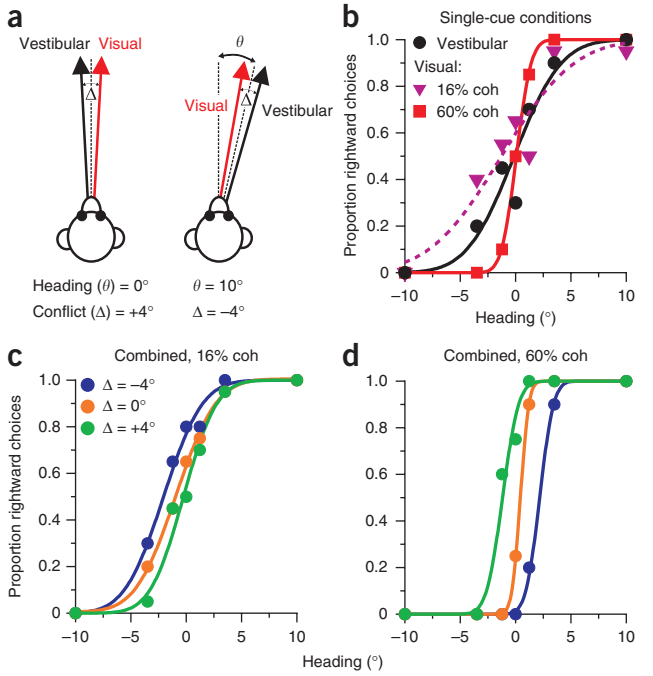


Figure 1 Cue-conflict configuration and example behavioral session. (**a**) Monkeys were presented with visual (optic flow) and/or vestibular (inertial motion) heading stimuli in the horizontal plane. The heading (θ) was varied in fine steps around straight ahead, and the task was to indicate rightward or leftward heading with a saccade after each trial. On a subset of visual-vestibular (combined) trials, the headings specified by each cue were separated by a conflict angle (Δ) of ±4°, where positive Δ indicates visual to the right of vestibular, and vice versa for negative Δ. Schematic shows two possible combinations of θ and Δ. (**b**) Psychometric functions for an example session showing the proportion of rightward choices as a function of heading for the single-cue conditions. Psychophysical thresholds were taken as the s.d. (σ) of the best-fitting cumulative Gaussian function (smooth curves) for each modality. Single-cue thresholds were used to predict (via equation (2)), the weights that an optimal observer should assign to each cue during combined trials. (**c**) Psychometric functions for the combined modality at low (16%) coherence, plotted separately for each value of Δ. The shifts of the PSEs during cue-conflict were used to compute observed vestibular weights (equation (4)). (**d**) Data are presented as in **c**, but for the high (60%) coherence combined trials.

Following the standard (linear) ideal-observer model of cue integration[15], we postulate an internal heading signal S_{comb} that is a weighted sum of vestibular and visual heading signals S_{ves} and S_{vis} (where $w_{vis} = 1 - w_{ves}$)

$$S_{comb} = w_{ves}S_{ves} + w_{vis}S_{vis} \quad (1)$$

If each S is considered to be a Gaussian random variable with mean μ and variance σ^2, the optimal estimate of μ_{comb} (minimizing its variance while remaining unbiased) is achieved by setting the weights in equation (1) proportional to the reliability (that is, inverse variance) of S_{ves} and S_{vis}[6,10]

$$w_{ves-opt} = \frac{1/\sigma_{ves}^2}{1/\sigma_{ves}^2 + 1/\sigma_{vis}^2}; w_{vis-opt} = \frac{1/\sigma_{vis}^2}{1/\sigma_{ves}^2 + 1/\sigma_{vis}^2} \quad (2)$$

The combined reliability is equal to the sum of the individual cue reliabilities

$$\frac{1}{\sigma_{comb}^2} = \frac{1}{\sigma_{ves}^2} + \frac{1}{\sigma_{vis}^2}$$

or, solving for σ_{comb}

$$\sigma_{comb-opt} = \sqrt{\frac{\sigma_{ves}^2\sigma_{vis}^2}{\sigma_{ves}^2 + \sigma_{vis}^2}} \quad (3)$$

Equation (3) formalizes the intuition that multisensory estimates should be more precise than single-cue estimates (that is, resulting in improved discrimination performance), with the maximum effect being a reduction in σ by a factor of $\sqrt{2}$ when single-cue reliabilities are matched. Note that the predictions specified by equation (2) and (3) are the same if optimality is defined in the Bayesian (maximum a posteriori) sense, assuming Gaussian likelihoods and a uniform prior[3,4,6,15,16].

To test these predictions, we trained two monkeys to report their heading relative to a fixed, internal reference of straight ahead (a one-interval, two-alternative forced-choice task). Heading stimuli could be presented in one of three modalities: vestibular (inertial motion delivered by a motion platform), visual (optic flow simulating observer movement through a random-dot cloud) or combined (simultaneous inertial motion and optic flow; see Online Methods and refs. 11–13,17). On combined trials, we pseudorandomly varied the conflict angle (Δ) between the headings specified by visual and vestibular cues (Δ = −4°, 0° or 4°; **Fig. 1a**). By convention, Δ = +4° indicates that the visual cue was displaced 2° to the right of the assigned heading value for that trial, whereas the vestibular cue was 2° to the left (and vice versa for Δ = −4°). To manipulate cue reliability, we also varied the percentage of coherently moving dots in the visual display (motion coherence; 16% or 60%). We chose these values for motion coherence to set the reliability of the visual cue (as measured by behavioral performance) above and below that of the vestibular cue, for which reliability was held constant.

Behavioral results (see also ref. 13) indicate that monkeys reweight visual and vestibular heading cues on a trial-by-trial basis in proportion to their reliability (**Fig. 1b–d**). Psychometric data (proportion rightward choices as a function of heading) were fit with cumulative Gaussian functions, yielding two parameters: the point of subjective equality (PSE, mean of the fitted cumulative Gaussian) and the threshold (defined as its s.d., σ). Similar to previous studies[1–6,8], thresholds from single-cue conditions (**Fig. 1b**) were used to estimate the relative reliability of the two cues, which specifies the weights that an optimal observer should apply to each cue (equation (2)). These optimal weights were computed from equation (2) by pairing the vestibular threshold (σ = 3.3° in this example; **Fig. 1b**) with each of the two visual thresholds (16% coherence (coh), σ = 5.1°; 60% coh, σ = 1.1°). Optimal vestibular weights for this session were 0.70 and 0.10 for low and high coherence, respectively (recall that $w_{vis} = 1 - w_{ves}$; we report only vestibular weights for simplicity).

On cue-conflict trials (**Fig. 1c,d**), the monkey's choices were biased toward the more reliable cue, as indicated by lateral shifts of the psychometric functions for Δ = ±4°, relative to Δ = 0°. At 16% coherence (**Fig. 1c**), when the vestibular cue was more reliable, the monkey made more rightward choices for a given heading when the vestibular cue was displaced to the right (Δ = −4°), and more leftward choices when the vestibular cue was to the left (Δ = +4°). This pattern was reversed when the visual cue was more reliable (60% coherence; **Fig. 1d**). We used the shifts in the PSEs to compute 'observed' vestibular weights, according to a formula

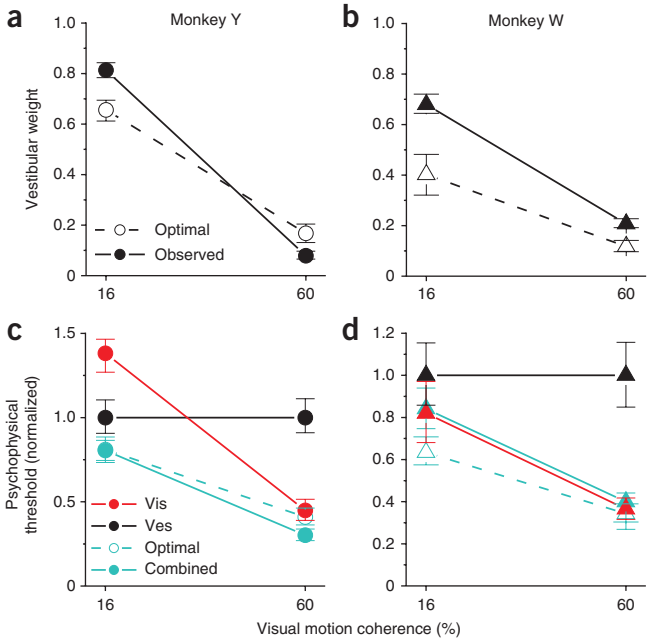


Figure 2 Average behavioral performance. (**a**,**b**) Optimal (equation (2), open symbols and dashed line) and observed (equation (4), filled symbols and solid line) vestibular weights as a function of visual motion coherence (cue reliability), shown separately for the two monkeys (**a**, monkey Y, $N = 40$ sessions; **b**, monkey W, $N = 26$). (**c**,**d**) Optimal (equation (3)) and observed (estimated from the psychometric fits) psychophysical thresholds, normalized separately by each monkey's vestibular threshold. Error bars represent 95% confidence intervals computed with a bootstrap procedure.

derived from the same framework (equation (1)) as the optimal weights (see Online Methods for derivation)

$$w_{\mathrm{ves-obs}} = \frac{PSE_{\mathrm{comb}} - PSE_{\mathrm{comb},\Delta=0} + \frac{\Delta}{2}}{\Delta} \qquad (4)$$

For the example dataset (**Fig. 1c**,**d**), the observed vestibular weights were 0.72 and 0.08 for 16% and 60% coherence, respectively.

We determined vestibular weights for two monkeys (**Fig. 2a**,**b**). Both monkeys showed robust changes in observed weights as a function of coherence (session-wise paired t tests: monkey Y, $N = 40$ sessions, $P < 10^{-28}$; monkey W, $N = 26$, $P < 10^{-13}$). Observed weights were reasonably close to the optimal predictions, especially for monkey Y (**Fig. 2a**). However, there were significant deviations from optimality in both animals (as reported previously for a larger sample of monkey and human subjects[13]), with the most consistent effect being an over-weighting of the vestibular cue at low coherence (paired t tests, observed > optimal at 16% coh: monkey W, $P < 10^{-8}$; monkey Y, $P < 10^{-6}$). Monkey W also over-weighted the vestibular cue at high coherence (60% coh, $P < 10^{-5}$), but the opposite trend was observed for monkey Y (observed < optimal at 60% coh, $P = 0.003$). These types of deviations from optimality, apparent over- or under-weighting of a particular sensory cue, are not unprecedented in the human psychophysics literature[5,18,19] and present an opportunity to look for neural signatures of the particular deviations that we observed.

We also determined the corresponding psychophysical thresholds for the two monkeys (**Fig. 2c**,**d**). Note that, unlike our previous study[11], these experiments were not designed to reveal optimal improvements in psychophysical thresholds under cue combination (equation (3)), as coherence was never chosen to equate the single-cue thresholds. Nevertheless, we generally observed near-optimal sensitivity in the combined modality (**Fig. 2c**,**d**). The most notable deviation was the case of 16% coherence for monkey W, which is also the condition of greatest vestibular over-weighting (**Fig. 2b**).

Single-neuron correlates of cue reweighting

To explore the neural basis of these behavioral effects, we recorded single-unit activity from cortical area MSTd, an extrastriate region that receives both visual and vestibular signals related to self-motion[12,17,20,21]. Our primary strategy was to decode a population of MSTd responses to predict behavioral choices under each set of experimental conditions, thereby constructing simulated psychometric functions from decoded neural responses. We then repeated the analyses of the previous section to test whether MSTd activity can account for cue integration behavior in this task. These analyses were performed on a sample of 108 neurons (60 from monkey Y, 48 from monkey W) that had significant tuning for at least one stimulus modality over a small range of headings near straight ahead (see Online Methods for detailed inclusion criteria).

Before describing the decoding results, we first illustrate the basic pattern of responses seen in individual neurons. Tuning curves (firing rate as a function of heading) for an example neuron (**Fig. 3a–c**), as for most MSTd neurons[11,12], were approximately linear over the small range of headings tested in the discrimination task (±10°). This neuron showed significant tuning ($P < 0.05$, one-way ANOVA) and a preference for leftward headings in all conditions and modalities. Note that the tuning curves from cue-conflict trials (**Fig. 3b**,**c**) shifted up or down depending on both the direction of the conflict (the sign of Δ) and the relative cue reliability (motion coherence). For most headings, the example neuron responded to the cue-conflict with an increase or decrease in firing rate depending on which cue was more 'reliable' in terms of neuronal sensitivity. Take, for example, the combined tuning at 60% coherence when Δ = −4° (visual heading to the left of vestibular; **Fig. 3c**). Because the cell preferred leftward headings, the dominance of the more reliable visual cue resulted in a greater firing rate relative to the Δ = 0° curve (and vice versa for Δ = +4°). The direction of the shifts was largely reversed at 16% coherence (**Fig. 3b**), for which the vestibular cue was more reliable.

These effects of cue-conflict and coherence on tuning curves suggest that MSTd neurons may be performing a weighted summation of inputs, as discussed below. We then considered whether these relatively simple changes in firing rates can account for the perceptual effects of cue reliability. Because perception arises from

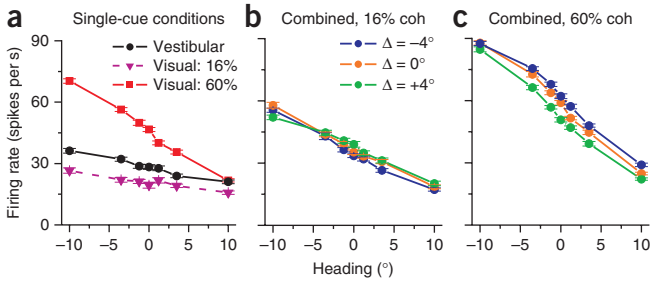


Figure 3 Example MSTd neuron showing a correlate of trial-by-trial cue reweighting. (**a**–**c**) Mean firing rate (spikes per s) ± s.e.m. is plotted as a function of heading for the single-cue trials (**a**), and combined trials at low (**b**) and high (**c**) coherence. The shift in combined tuning curves with cue conflict, in opposite directions for the two levels of reliability, forms the basis for the reweighting effects in the population decoding analysis depicted in **Figures 4** and **6** (see **Supplementary Figs. 1** and **2** for single-cell neurometric analyses).

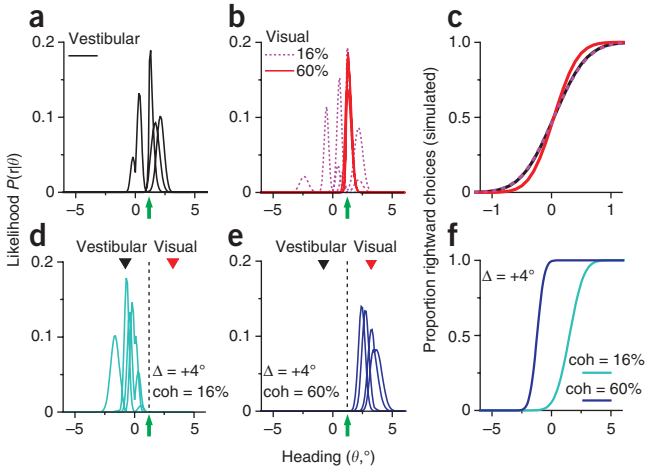


Figure 4 Likelihood-based decoding approach used to simulate behavioral performance based on MSTd activity. (**a**,**b**) Example likelihood functions ($P(\mathbf{r}|\theta)$) for the single-cue modalities. Four individual trials of the same heading ($\theta = 1.2°$, green arrow) are superimposed for each condition. Likelihoods were computed from equation (14) using simulated population responses (**r**) comprised of random draws of single-neuron activity. (**c**) Simulated psychometric functions for a decoded population that included all 108 MSTd neurons in our sample. (**d**,**e**) Combined modality likelihood functions for $\theta = 1.2°$ (green arrow and dashed line) and $\Delta = +4°$, for low (cyan) and high (blue) coherence. Black and red inverted triangles indicate the headings specified by vestibular and visual cues, respectively, in this stimulus configuration. (**f**) Psychometric functions for the simulated combined modality, showing the shift in the PSE resulting from the change in coherence (that is, reweighting).

the concurrent activity of many neurons, we focused on making behavioral predictions from population activity using a decoding approach. We also performed cell-by-cell neurometric analyses (see **Supplementary Analysis**), which gave similar results (**Supplementary Figs. 1** and **2**).

Likelihood-based decoding of MSTd neuronal populations

We used a well-established method of maximum-likelihood decoding[22–24] to convert MSTd population responses into perceptual choices made by an ideal observer performing the heading discrimination task. Our approach was to simulate population activity on individual trials by pooling responses from our sample of MSTd neurons, a strategy made possible by having a fixed set of stimuli and conditions across recording sessions. On each simulated trial, the decoder estimated the most likely heading on the basis of the population response. Responses (**r**) were generated by taking random draws of single-trial firing rates from each recorded neuron under the particular set of conditions (stimulus modality, heading (θ) and conflict angle (Δ)) being simulated on that trial. We then computed the full likelihood function $P(\mathbf{r}|\theta)$ using an expression (equation (14), Online Methods) that assumes independent Poisson variability (see **Supplementary Analysis**; relaxing the assumptions of independence and Poisson noise did not substantially affect the results). Assuming a flat prior, we normalized the likelihood to obtain the posterior distribution, $p(\theta|\mathbf{r})$, then computed the area under the posterior favoring leftward versus rightward headings. When the integrated posterior for rightward headings exceeded that for leftward headings, the ideal observer registered a rightward choice, and vice versa.

Notably, cue-conflict trials were decoded with respect to the non-conflict ($\Delta = 0°$) tuning curves (for example, see **Fig. 3b**,**c**). The implicit assumption here is that the decoder, or downstream area reading out MSTd activity, does not alter how it interprets the population response on the basis of the (unpredictable) presence or absence of a cue-conflict. Given that animals typically experience self-motion without a consistent conflict between visual and vestibular cues, it is reasonable to assume that the brain interprets neuronal responses as though there was no cue-conflict. This assumption allows the shift in tuning curves resulting from the cue-conflict (**Fig. 3b**,**c**) to manifest as a shift in the likelihoods and, thus, the PSE of the simulated psychometric functions, as described below.

We first computed likelihood functions from our full sample of MSTd neurons ($N = 108$) in the single-cue conditions (**Fig. 4a**,**b**). Results from four representative trials are overlaid for a constant simulated heading of 1.2°. Note that the vestibular (**Fig. 4a**) and low-coherence visual (**Fig. 4b**) likelihoods are more variable in their position on the heading axis compared with the high-coherence visual condition (**Fig. 4b**). This differential variability is reflected in the slopes of the simulated psychometric functions produced by the decoder (**Fig. 4c**), which yielded thresholds of 0.41°, 0.38° and 0.27°, respectively. From equation (2), these threshold values yield optimal vestibular weights of 0.46 and 0.30 for low and high coherence, respectively.

We next computed the combined likelihoods for the same simulated heading of 1.2° when a positive cue-conflict ($\Delta = +4°$) was introduced (**Fig. 4d**,**e**). In this stimulus condition, the visual heading is +3.2°

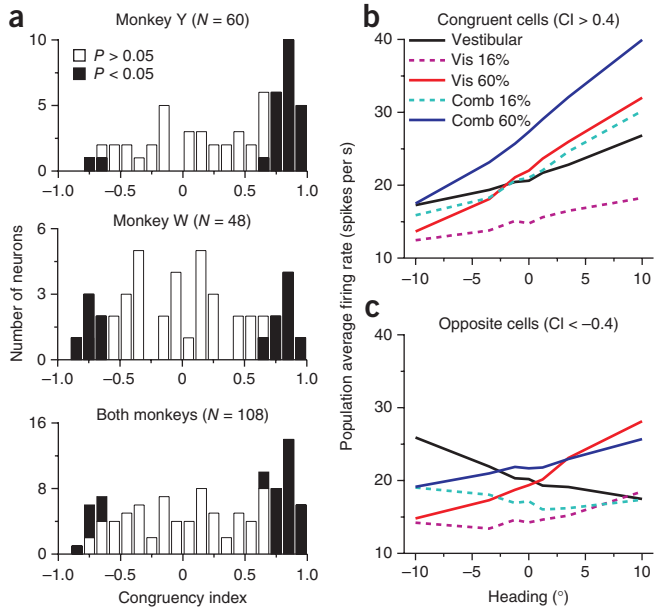


Figure 5 Visual-vestibular congruency and average MSTd tuning curves. (**a**) Histogram of congruency index (CI) values for monkey Y (top), monkey W (middle) and both animals together (bottom). Positive congruency index values indicate consistent tuning slope across visual (60% coh) and vestibular single-cue conditions, whereas negative values indicate opposite tuning slopes. Filled bars indicate congruency index values whose constituent correlation coefficients were both statistically significant[11]; however, here we defined congruent and opposite cells by an arbitrary criterion of congruency index > 0.4 and congruency index < –0.4, respectively. (**b**,**c**) Population average of MSTd tuning curves for the five stimulus conditions, vestibular (black), low-coherence visual (magenta, dashed), high-coherence visual (red), low-coherence combined (cyan, dashed) and high-coherence combined (blue), separated into congruent (**b**) and opposite (**c**) classes. Prior to averaging, some neurons' tuning preferences were mirrored such that all cells preferred rightward heading in the high-coherence visual modality.

and the vestibular heading is −0.8°. At 16% coherence (**Fig. 4d**), the likelihood functions tended to peak closer to the vestibular heading, whereas they peaked closer to the visual heading at 60% coherence (**Fig. 4e**). This generated more leftward choices at low coherence and more rightward choices at high coherence, and thus a shift of the combined psychometric function in opposite directions relative to zero (**Fig. 4f**). The observed weights corresponding to these PSE shifts were 0.80 and 0.14 for low and high coherence, respectively ($P << 0.05$, bootstrap). Thus, the population decoding approach reproduced the robust cue reweighting effect observed in the behavior (**Fig. 2a**,**b**). It is important to emphasize that, because coherence was varied randomly from trial to trial, this neural correlate of cue reweighting in MSTd must occur fairly rapidly (that is, on the timescale of each 2-s trial).

Decoding summary and the effect of tuning congruency

Previously[11], we found that the relative tuning for visual and vestibular heading cues in MSTd varies along a continuum from congruent to opposite, where congruency in the present context refers to the similarity of tuning slopes for the two single-cue modalities. We quantified this property for each neuron using a congruency index, defined as the product of Pearson correlation coefficients comparing firing rate versus heading for the two modalities[11] (**Fig. 5a**). Positive congruency index values indicate visual and vestibular tuning curves with consistent slopes, negative values indicate opposite tuning slopes, and values near 0 occur when the tuning curve for either modality is flat or even symmetric. Our previous study[11] used a statistical criterion on the congruency index to classify neurons as either congruent or opposite; however, we found this to be too restrictive for the present dataset, as it did not permit sufficient sample sizes to be analyzed for each animal and congruency class. Thus, we defined congruent and opposite cells by an arbitrary criterion of congruency index > 0.4 (N = 46 of 108 cells, 43%) or congruency index < −0.4 (N = 23 of 108 cells, 21%), respectively. We obtained similar results for different congruency index criteria (±0.5, ±0.2 or simply congruency index > 0 versus < 0).

Congruency is an important attribute for decoding MSTd responses because only congruent cells show increased heading sensitivity during cue combination that parallels behavior[11]. This can be seen in our dataset by examining the relative slopes of average tuning curves for each stimulus modality (**Fig. 5b**; aligned so that all cells prefer rightward in the visual modality). The combined tuning curve of congruent cells at 16% coherence was steeper than either of the single-cue tuning curves, and likewise at 60% coherence, resulting in improved neuronal sensitivity (**Supplementary Fig. 3**). In contrast, opposite cells are generally less sensitive to heading under combined stimulation because the visual and vestibular signals counteract each other and the tuning becomes flatter (**Fig. 5c**). Thus, we examined how decoding subpopulations of congruent and opposite cells compares to decoding all neurons.

Decoding all cells without regard to congruency (N = 108), as mentioned above, yielded a strong reweighting effect (**Fig. 6a**), but did not reproduce the improvement in psychophysical threshold that is characteristic of behavioral cue integration (**Fig. 6b**). This outcome is not surprising given the contribution of opposite cells (**Fig. 6c**,**d**), for which combined thresholds were substantially worse than thresholds for the best single cue. In contrast, decoding only congruent cells yielded both robust cue reweighting (**Fig. 6e**) and a significant reduction in threshold ($P < 0.05$; **Fig. 6f**), matching the behavioral results (**Fig. 6g**,**h**) quite well. Notably, this subpopulation of congruent neurons also reproduced the over-weighting of the vestibular cue at low coherence and the slight under-weighting at high coherence (**Fig. 6e**,**g**). These findings support the hypothesis that congruent cells in MSTd are selectively read out to support cue integration behavior[11] and suggest that the neural representation of heading in MSTd may contribute to deviations from optimality in this task.

We also compared behavior with decoder performance for each animal separately. Despite the small samples of congruent cells (N = 32 for monkey Y and 14 for monkey W), the decoder captured a number of aspects of the individual differences between animals in

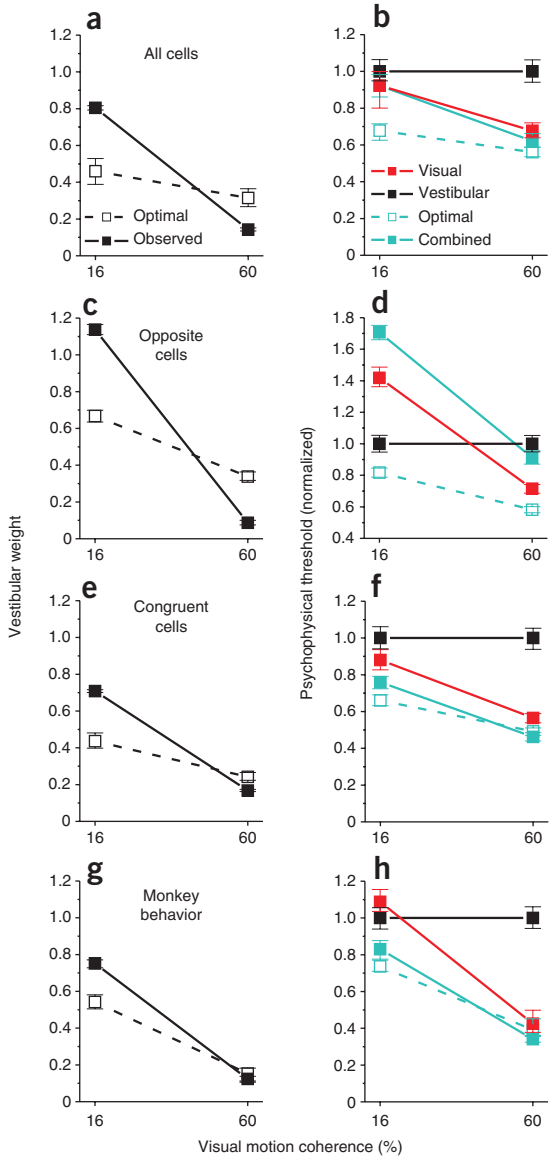


Figure 6 Population decoding results and comparison with monkey behavior. (**a**–**f**) Weights (left column, data are presented as in **Fig. 2a**,**b**; from equation (2) and (4)) and thresholds (right column, data presented as in **Fig. 2c**,**d**; from equation (3) and psychometric fits to real or simulated choice data) quantifying the performance of an optimal observer reading out MSTd population activity. Thresholds were normalized by the value of the vestibular threshold. The population of neurons included in the decoder was varied to examine the readout of all cells (**a**,**b**), opposite cells only (**c**,**d**; note the different ordinate scale in **d**) or congruent cells only (**e**,**f**). (**g**,**h**) Monkey behavioral performance (pooled across the two animals) is summarized. Error bars indicate 95% confidence intervals.

weights and thresholds (**Supplementary Fig. 4**). This suggests that individual differences in cue integration behavior may at least partly reflect differences in the neural representation of heading in MSTd.

Multisensory neural combination rule: optimal versus observed

Thus far we have shown that decoded MSTd responses can account for behavior, but we have not considered the computations taking place at the level of single neurons. Returning to equation (1), consider that, instead of dealing with abstract signals S_{ves} and S_{vis}, the brain must combine signals in the form of neural firing rates. We now describe, at a mechanistic level, how multisensory neurons should combine their inputs to achieve optimal cue integration, and then we test whether MSTd neurons indeed follow these predictions. Our previous work[25] established that a linear combination rule is sufficient to describe the combined firing rates (tuning curves) of most MSTd neurons

$$f_{comb}(\theta,c) = A_{ves}(c)f_{ves}(\theta) + A_{vis}(c)f_{vis}(\theta,c) \quad (5)$$

where f_{ves}, f_{vis} and f_{comb} are the tuning curves of a particular MSTd neuron for the vestibular, visual and combined modalities, θ represents heading, c denotes coherence, and A_{ves} and A_{vis} are the neural weights (we use the term neural weights to distinguish them from the perceptual weights of equations (1) and (2)).

The key issue is to understand the relationship between the perceptual weights (w_{ves} and w_{vis} in equation (1)) and the neural weights (A_{ves} and A_{vis} in equation (5)). One might expect that neural and perceptual weights should exhibit the same dependence on cue reliability (coherence); that is, that the neural vestibular weight A_{ves} should decrease with coherence as do the perceptual weights (**Fig. 2a**,**b**). Notably, however, this needs not be the case: the relationship between perceptual and neural weights depends on the statistics of the neuronal spike counts and on how the tuning curves are modulated by coherence. If neurons fire with Poisson statistics and tuning curves are multiplicatively scaled by coherence, then the optimal neural weights will be equal to 1 and independent of coherence[14], whereas the perceptual weights still clearly depend on coherence. For other dependencies of tuning curves on cue reliability, neural and perceptual weights may share a similar dependence on coherence, but this remains to be determined.

It is therefore critical that we derive the predicted optimal neural weights for the MSTd neurons that we recorded. In our data, coherence does not have a simple multiplicative effect on tuning curves. Instead, as expected from the properties of middle temporal (MT/V5)[26] and MST[27] neurons when presented with conventional random-dot stimuli, the effect of coherence in MSTd is better described as a scaling plus a baseline that decreases with coherence

$$f_{vis} = \alpha c f^{*} + \beta(1 - c) \quad (6)$$

where f^* is a generic linear tuning function, α and β are constants, and c denotes coherence. We did not quantify this effect here, but it can be visualized in **Figures 3a** and **5b**,**c** (see also **Supplementary Figure 3** in ref. 25 for an illustration with the full heading tuning curve).

Using equation (5) and a few well-justified assumptions, we can derive the optimal neural weights (**Supplementary Analysis**). This derivation yields a simple expression for the optimal ratio of neural weights, ρ_{opt}, where $\rho_{opt} = A_{ves-opt} / A_{vis-opt}$

$$\rho_{opt}(c) = \frac{f'_{ves}(0)f_{vis}(0,c)}{f_{ves}(0)f'_{vis}(0,c)} \quad (7)$$

In this equation, f'_{ves} and f'_{vis} denote the derivatives of the tuning curves with respect to θ, and all terms are evaluated at the reference heading ($\theta = 0°$). Substituting equation (6) into equation (7), we obtain

$$\rho_{opt} = \frac{f'_{ves}}{f_{ves}}\left(\frac{f^{*}}{f^{*\prime}} - \frac{\beta}{\alpha f^{*\prime}} + \frac{\beta}{\alpha f^{*\prime} c}\right) \quad (8)$$

which predicts that the optimal weight ratio is inversely proportional to coherence.

With an expression in hand for the optimal neural weight ratio (equation (7)), it is possible to test whether the neural weights measured in MSTd are consistent with optimal predictions. We measured neural weights for MSTd cells previously[25], but the monkeys in the previous study were passively fixating rather than performing a psychophysical task, and coherence was varied in blocks of trials. To the best of our knowledge, this is the first study to measure both behavioral and neural weights with multiple interleaved levels of cue reliability, and the first to test whether these weights are optimal (as defined by equation (7)). We fit equation (5) to the responses of all neurons (N = 108) separately for the two coherence levels and plotted the distribution of R^2 values for the fits (**Fig. 7a**,**b**). Despite having only two free parameters (A_{ves} and A_{vis}) for each coherence, the model explained the data reasonably well for most neurons (low coherence, 88 of 108 cells, 81%, with significant correlation between responses and model fits, $P < 0.05$; high coherence, 94 of 108 cells, 87%; **Fig. 7a**,**b**). Some poor fits are expected because several neurons had vestibular or low-coherence visual tuning curves that were essentially flat. Consistent with previous findings[25], the neural weights changed as a function of coherence, with A_{ves} decreasing and A_{vis} increasing as coherence increased

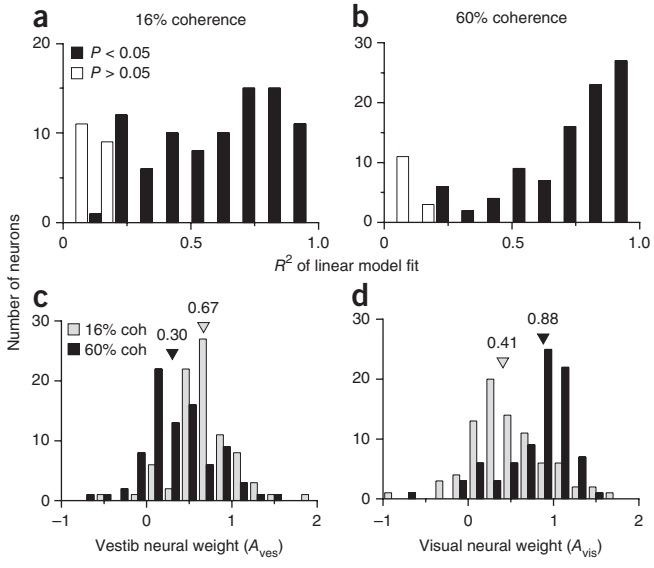


Figure 7 Goodness-of-fit of linear weighted sum model and distribution of vestibular and visual neural weights. Combined responses during the discrimination task (N = 108) were modeled as a weighted sum of visual and vestibular responses, separately for each coherence level (equation (5)). (**a**,**b**) Histograms of a goodness-of-fit metric (R^2), taken as the square of the correlation coefficient between the modeled response and the real response. The statistical significance of this correlation was used to code the R^2 histograms. (**c**,**d**) Histograms of vestibular (**c**) and visual (**d**) neural weights, separated by coherence (gray bars = 16%, black bars = 60%). Color-matched arrowheads indicate medians of the distributions. Only neurons with significant R^2 values for both coherences were included (N = 83).

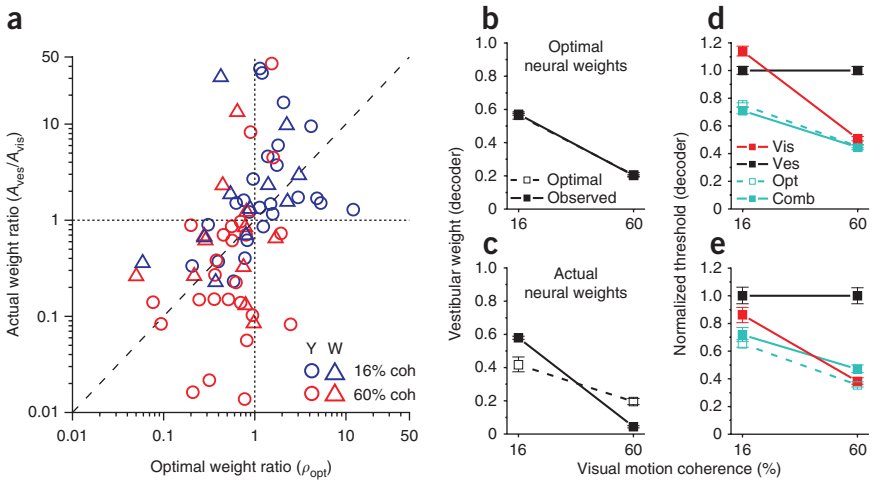


Figure 8 Comparison of optimal and actual (fitted) neural weights. (**a**) Actual weight ratios (A_{ves}/A_{vis}) for each cell were derived from the best-fitting linear model (equation (5), as in **Fig. 7**), and optimal weight ratios (ρ_{opt}) for the corresponding cells were computed according to equation (7). Symbol color indicates coherence (16%, blue; 60%, red) and shape indicates monkey identity. Note that the derivation of equation (7) assumes congruent tuning (see **Supplementary Analysis**), and ρ_{opt} is therefore constrained to be positive (because the sign of the tuning slopes will be equal). Thus, only congruent cells with positive weight ratios were included in this comparison ($N = 36$ for low coherence, 37 for high coherence). (**b**,**d**) Decoder performance (data are presented as in **Figure 6**, using equation (2)–(4) and fits to simulated choice data) based on congruent neurons, after replacing combined modality responses with weighted sums of single-cue responses, using the optimal weights from equation (7) (abscissa in **a**). (**c**,**e**) Data are presented as in **b** and **c**, using the actual (fitted) weights (ordinate in **a**) to generate the artificial combined responses.

(paired t test comparing 16% versus 60% coherence, $P < 10^{-7}$ for both A_{ves} and A_{vis}; **Fig. 7c**,**d**). For the majority of neurons (51% of all cells, 68% of congruent cells), a version of the model in which weights were allowed to vary with coherence (independent-weights model) provided significantly better fits (sequential F test, $P < 0.05$) than a model with a single set of weights for both coherences (yoked-weights model; see Online Methods and ref. 25). Because coherence was randomized from trial to trial, these changes in neural weights must have occurred on a rapid time scale and are therefore unlikely to reflect changes in synaptic weights (see Discussion and ref. 28). Indeed, examining the time course of neural weights across the 2-s trial duration (**Supplementary Fig. 5**) revealed that the effect was established shortly after response onset and persisted through most of the stimulus duration.

How do the optimal neural weights defined by equation (7) compare with the actual neural weights (**Fig. 7c**,**d**) determined by fitting MSTd responses? The actual weight ratios (A_{ves}/A_{vis}) were significantly correlated with the corresponding optimal weight ratios, ρ_{opt} (Spearman's rank correlation, $\rho = 0.57$, $P < 0.0001$ for all data points; $\rho = 0.31$, $P = 0.06$ for 16% coh; $\rho = 0.40$, $P = 0.02$ for 60% coh; **Fig. 8a**). Note that the trend of $A_{ves} > A_{vis}$ at low coherence and $A_{ves} < A_{vis}$ at high coherence held for both the actual weights (**Fig. 7c**,**d**) and the optimal weights ($\rho_{opt} > 1$ for 16% coh, one-tailed Wilcoxon signed rank test versus hypothetical median of 1, $P = 0.04$; $\rho_{opt} < 1$ for 60% coh, $P = 0.0001$; **Fig. 8a**).

Notably, the actual neural weight ratios changed with coherence to a greater extent than the optimal weight ratios, in a manner that favored the vestibular cue at low coherence (Wilcoxon matched pairs test, $A_{ves}/A_{vis} > \rho_{opt}$, $P = 0.007$ for 16% coh; **Fig. 8a**). There was a trend in the opposite direction for 60% coherence when considering ratios < 1, but, overall, the difference for 60% was not significant ($P = 0.78$). However, the slope of a linear fit to all data points (type II regression on log-transformed data) was significantly greater than 1 ($P < 0.05$, bootstrap) (**Fig. 8a**). This pattern of results suggests that the multisensory combination rule in MSTd may predict systematic deviations from optimality similar to those seen in the behavior (**Fig. 2a**,**b**).

To solidify this prediction, we generated artificial combined responses of model neurons using each of two weighting schemes: optimal neural weights (from equation (7)) and actual neural weights (**Fig. 7c**,**d**). These responses were generated from equation (5) using either the best-fitting neural weights or by setting $A_{vis} = 1$ and $A_{ves} = \rho_{opt}$ (equation (7)). We then decoded these artificial responses to see how the neural weights translate into predicted behavioral weighting of cues. Using the optimal neural weights (**Fig. 8b**) resulted in optimal cue weighting at the level of simulated behavior (observed decoder weights nearly equal to optimal predictions, $P > 0.05$, bootstrap), as expected from the theory. In contrast, decoding artificial responses generated with the actual neural weights (**Fig. 8c**) yielded results similar to those obtained when decoding measured MSTd responses to combined stimuli (**Fig. 6e**), including a steeper dependence on coherence of observed versus optimal weights, and vestibular overweighting at low coherence. Decoder thresholds showed a similar pattern: matching optimal performance when derived from optimal neural weights (**Fig. 8d**), but suboptimal when derived from actual neural weights (**Fig. 8e**).

In summary, these findings provide a theoretical basis for the changes in neural weights with coherence that we observed and suggest that departures from optimality in behavior (**Fig. 2**) can be at least partially explained by departures from optimality in the cue-combination rule employed by MSTd neurons. This establishes a vital link between descriptions of cue integration at the level of single neurons and the level of behavior.

DISCUSSION

We investigated the neural basis of multisensory cue integration using the classic perturbation analysis method of cue-conflict psychophysics[1–3,5,6,8], combined with single-unit recordings in monkeys. We found that the activity of multisensory neurons in area MSTd is modulated by changes in cue reliability across trials, analogous to the dynamic adjustment of psychophysical weights that is a hallmark of optimal integration schemes. Robust correlates of cue integration behavior were observed in individual neurons (**Fig. 3** and **Supplementary Figs. 1** and **2**) and with a population-level decoding approach (**Figs. 4** and **6** and **Supplementary Fig. 4**). We also found that MSTd neurons combine their inputs in a manner that is broadly compatible with a theoretically derived optimal cue-combination rule and that measured departures from this rule could at least partially explain deviations from optimality at the level of behavior. Together with previous findings[11,12], our results strongly implicate area MSTd in the integration of visual and vestibular cues for self-motion perception. More generally, our findings provide new insight into the neural computations underlying reliability-based cue weighting, a type of

statistical inference that occurs within and across nearly all sensory modalities and tasks in which it has been tested.

Cue integration theory and probabilistic inference

It has long been appreciated that sensory information is probabilistic and cannot specify environmental variables with certainty. Perception, therefore, is often considered to be a problem of inference[16]. How does the nervous system infer the most likely configuration of the world from limited sensory data? Compounding this problem, the degree of uncertainty itself can vary unpredictably, as in most complex natural environments where multiple cues are available. One broad hypothesis[29] states that the brain computes conditional probability distributions over stimulus values, rather than only having access to estimates of these values. Because probability distributions are a natural way to represent reliability, the weighting of sensory cues by reliability[1–7] has been taken as indirect evidence for such a probabilistic coding scheme. However, despite numerous insights from human psychophysics, the neural substrates underlying reliability-weighted cue integration have remained obscure.

Although we cannot rule out alternative ways to perform reliability-weighted cue integration (for example, using so-called ancillary cues[6] that are unrelated to the sensory estimate, but convey information about its reliability), our results are consistent with a neural theory known as probabilistic population coding (PPC)[14]. According to this theory, the brain performs inference by making use of the probability distributions encoded in the sensory population activity itself. This strategy has particular advantages in a dynamic cue-integration context, as the required representations of cue uncertainty arise quickly and automatically, without the need for learning.

Notably, the specific operations required to implement optimal cue integration depend on the nature of the neural code. The key assertion of the PPC framework is that the code is specified by the likelihood function, $P(\mathbf{r}|\theta,c)$; in our case, θ denotes heading and c denotes coherence. In the original formulation of PPC[14], it was suggested that a linear combination of sensory inputs with weights that are independent of coherence (or any factor controlling the reliability of the stimulus) should be sufficient for optimal integration, whereas we found neural weights that depend on coherence (**Fig. 7c**,**d** and ref. 25). This prediction[14], however, assumed that tuning curves are multiplicatively scaled by coherence, which is not the case for MSTd neurons[25,27]; the amplitude of visual tuning is proportional to coherence, but the baseline is inversely proportional to coherence (equation (6)). Once this observation is used to incorporate the proper likelihood function (note that measuring a tuning curve is equivalent to sampling a set of probability distributions $P(\mathrm{r}|\theta_i,c)$), the PPC theory predicts coherence-dependent neural weights that correlated with those found experimentally (**Fig. 8a**). Systematic differences between actual and optimal neural weights (**Fig. 8a**) are sufficient to explain some of the observed deviations from optimal performance (**Fig. 6e**,**g**), as revealed by decoding artificially generated combined responses (**Fig. 8b**–**e**). These results suggest that MSTd neurons, to a first approximation, weight their inputs in a manner that is predicted by probabilistic models such as PPC, and that quantitative details of this neural weighting scheme can place important constraints on behavior.

The cellular and/or circuit mechanism(s) that give rise to the reliability-dependent neural combination rule remain unclear. Recent modeling efforts[28] suggest that a form of divisive normalization, acting in the multisensory representation, may account for changes in neural weights with coherence, as well as other classic observations[30] in the multisensory integration literature. Our finding that the neural weights change rapidly, both across (**Fig. 7c**,**d**) and within trials (**Supplementary Fig. 6**), is consistent with a fast network mechanism such as normalization, rather than a slower mechanism involving synaptic weight changes. It will be worthwhile for future studies in a range of sensory systems to quantify the neural combination rule and to test the predictions of the normalization model[28] against alternatives.

Neural substrates of self-motion perception

Perception of self-motion is a multifaceted cognitive process, both at the input stage (integrating visual, vestibular, somatosensory, proprioceptive and, perhaps, auditory cues) and the output stage (informing motor control and planning, spatial constancy, navigation and memory). Thus, although MSTd appears to be involved in heading perception, it is highly likely that other regions participate as well. Several cortical areas believed to receive visual and vestibular signals related to self-motion, such as the ventral intraparietal area[31,32] and frontal eye fields[33,34], are also strongly interconnected with MSTd[35,36], suggesting the existence of multiple heading representations that could work in parallel. The relatively long (2 s) trials and gradual stimulus onset that we used make it plausible that the multisensory responses of MSTd neurons are shaped by activity across this network of regions.

Although the source of visual motion signals[37,38] and the computations underlying optic flow selectivity[39–41] are fairly well understood, the origin and properties of vestibular signals projecting to these areas remain less clear. A wide-ranging effort to record and manipulate neural activity across a variety of regions will be necessary to tease apart the circuitry underlying this complex and important perceptual ability. Such efforts may eventually help in targeting new therapies for individuals with debilitating deficits in spatial orientation and navigation.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/natureneuroscience/.

Note: Supplementary information is available on the Nature Neuroscience website.

ACKNOWLEDGMENTS

We thank A. Turner, J. Arand, H. Schoknecht, B. Adeyemo and J. Lin for animal care and technical help, and V. Rao, S. Nelson and M. Jazayeri for discussions and comments on the manuscript. We are indebted to Y. Guan and S. Liu for generously contributing data to this project. This work was supported by grants from the US National Institutes of Health (EY019087 to D.E.A. and EY016178 to G.C.D.) and a US National Institutes of Health Institutional National Research Service Award (5-T32-EY13360-07 supporting C.R.F.). A.P. was supported by grants from the National Science Foundation (BCS0446730), Multidisciplinary University Research Initiative (N00014-07-1-0937) and the James McDonnell Foundation.

AUTHOR CONTRIBUTIONS

C.R.F., G.C.D. and D.E.A. conceived the study and designed the analyses. C.R.F. performed the experiments and analyzed the data. A.P. derived equations 6–8 and consulted on all theoretical aspects of the work. All of the authors wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/natureneuroscience/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Alais, D. & Burr, D. The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* **14**, 257–262 (2004).
2. Ernst, M.O. & Banks, M.S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429–433 (2002).
3. Hillis, J.M., Watt, S.J., Landy, M.S. & Banks, M.S. Slant from texture and disparity cues: optimal cue combination. *J. Vis.* **4**, 967–992 (2004).
4. Jacobs, R.A. Optimal integration of texture and motion cues to depth. *Vision Res.* **39**, 3621–3629 (1999).

5. Knill, D.C. & Saunders, J.A. Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Res.* **43**, 2539–2558 (2003).
6. Landy, M.S., Maloney, L.T., Johnston, E.B. & Young, M. Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Res.* **35**, 389–412 (1995).
7. van Beers, R.J., Wolpert, D.M. & Haggard, P. When feeling is more important than seeing in sensorimotor adaptation. *Curr. Biol.* **12**, 834–837 (2002).
8. Young, M.J., Landy, M.S. & Maloney, L.T. A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision Res.* **33**, 2685–2696 (1993).
9. Clark, J.J. & Yuille, A.L. *Data Fusion for Sensory Information Processing Systems* (Kluwer Academic, Boston, 1990).
10. Cochran, W.G. Problems arising in the analysis of a series of similar experiments. *J. R. Stat. Soc.* **4** (suppl.) 102–118 (1937).
11. Gu, Y., Angelaki, D.E. & DeAngelis, G.C. Neural correlates of multisensory cue integration in macaque MSTd. *Nat. Neurosci.* **11**, 1201–1210 (2008).
12. Gu, Y., DeAngelis, G.C. & Angelaki, D.E. A functional link between area MSTd and heading perception based on vestibular signals. *Nat. Neurosci.* **10**, 1038–1047 (2007).
13. Fetsch, C.R., Turner, A.H., DeAngelis, G.C. & Angelaki, D.E. Dynamic reweighting of visual and vestibular cues during self-motion perception. *J. Neurosci.* **29**, 15601–15612 (2009).
14. Ma, W.J., Beck, J.M., Latham, P.E. & Pouget, A. Bayesian inference with probabilistic population codes. *Nat. Neurosci.* **9**, 1432–1438 (2006).
15. Landy, M.S., Banks, M.S. & Knill, D.C. Ideal-observer models of cue integration. in *Sensory Cue Integration* (eds. Trommershäuser, J., Kording, K.P. & Landy, M.S.) 5–29 (Oxford University Press, New York, 2011).
16. Knill, D.C. & Richards, W. *Perception as Bayesian Inference* (Cambridge University Press, New York, 1996).
17. Gu, Y., Watkins, P.V., Angelaki, D.E. & DeAngelis, G.C. Visual and nonvisual contributions to three-dimensional heading selectivity in the medial superior temporal area. *J. Neurosci.* **26**, 73–85 (2006).
18. Battaglia, P.W., Jacobs, R.A. & Aslin, R.N. Bayesian integration of visual and auditory signals for spatial localization. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **20**, 1391–1397 (2003).
19. Rosas, P., Wagemans, J., Ernst, M.O. & Wichmann, F.A. Texture and haptic cues in slant discrimination: reliability-based cue weighting without statistically optimal cue combination. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **22**, 801–809 (2005).
20. Duffy, C.J. MST neurons respond to optic flow and translational movement. *J. Neurophysiol.* **80**, 1816–1827 (1998).
21. Takahashi, K. *et al.* Multimodal coding of three-dimensional rotation and translation in area MSTd: comparison of visual and vestibular selectivity. *J. Neurosci.* **27**, 9742–9756 (2007).
22. Dayan, P. & Abbott, L.F. *Theoretical Neuroscience* (MIT press, Cambridge, Massachusetts, 2001).
23. Foldiak, P. The 'ideal homunculus': statistical inference from neural population responses. in *Computation and Neural Systems* (eds. F.H. Eeckman & J.M. Bower) 55–60 (Kluwer Academic Publishers, Norwell, Massachusetts, 1993).
24. Sanger, T.D. Probability density estimation for the interpretation of neural population codes. *J. Neurophysiol.* **76**, 2790–2793 (1996).
25. Morgan, M.L., DeAngelis, G.C. & Angelaki, D.E. Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron* **59**, 662–673 (2008).
26. Britten, K.H., Shadlen, M.N., Newsome, W.T. & Movshon, J.A. Responses of neurons in macaque MT to stochastic motion signals. *Vis. Neurosci.* **10**, 1157–1169 (1993).
27. Heuer, H.W. & Britten, K.H. Linear responses to stochastic motion signals in area MST. *J. Neurophysiol.* **98**, 1115–1124 (2007).
28. Ohshiro, T., Angelaki, D.E. & DeAngelis, G.C. A normalization model of multisensory integration. *Nat. Neurosci.* **14**, 775–782 (2011).
29. Knill, D.C. & Pouget, A. The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* **27**, 712–719 (2004).
30. Stein, B.E. & Meredith, M.A. *The Merging of the Senses* (MIT Press, Cambridge, Massachusetts, 1993).
31. Bremmer, F., Klam, F., Duhamel, J.R., Ben Hamed, S. & Graf, W. Visual-vestibular interactive responses in the macaque ventral intraparietal area (VIP). *Eur. J. Neurosci.* **16**, 1569–1586 (2002).
32. Chen, A., DeAngelis, G.C. & Angelaki, D.E. Representation of vestibular and visual cues to self-motion in ventral intraparietal cortex. *J. Neurosci.* **31**, 12036–12052 (2011).
33. Fukushima, K. Corticovestibular interactions: anatomy, electrophysiology and functional considerations. *Exp. Brain Res.* **117**, 1–16 (1997).
34. Xiao, Q., Barborica, A. & Ferrera, V.P. Radial motion bias in macaque frontal eye field. *Vis. Neurosci.* **23**, 49–60 (2006).
35. Lewis, J.W. & Van Essen, D.C. Corticocortical connections of visual, sensorimotor and multimodal processing areas in the parietal lobe of the macaque monkey. *J. Comp. Neurol.* **428**, 112–137 (2000).
36. Stanton, G.B., Bruce, C.J. & Goldberg, M.E. Topography of projections to posterior cortical areas from the macaque frontal eye fields. *J. Comp. Neurol.* **353**, 291–305 (1995).
37. Felleman, D.J. & Van Essen, D.C. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1**, 1–47 (1991).
38. Maunsell, J.H. & van Essen, D.C. The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *J. Neurosci.* **3**, 2563–2586 (1983).
39. Lappe, M., Bremmer, F., Pekel, M., Thiele, A. & Hoffmann, K.P. Optic flow processing in monkey STS: a theoretical and experimental approach. *J. Neurosci.* **16**, 6265–6285 (1996).
40. Fukushima, K. Extraction of visual motion and optic flow. *Neural Netw.* **21**, 774–785 (2008).
41. Perrone, J.A. & Stone, L.S. Emulating the visual receptive-field properties of MST neurons with a template model of heading estimation. *J. Neurosci.* **18**, 5958–5975 (1998).

ONLINE METHODS

Subjects, stimuli and behavioral task. Experimental procedures were in accordance with US National Institutes of Health guidelines and approved by the Animal Studies Committee at Washington University. Two rhesus monkeys (*Macaca mulatta*) were trained to perform a fine heading discrimination task in the horizontal plane, as described previously[11–13]. Monkeys were seated in a virtual-reality setup consisting of a motion platform (MOOG 6DOF2000E, Moog), eye-coil frame, projector (Mirage 2000, Christie) and rear-projection screen. In each trial, a 2-s translational motion stimulus with a Gaussian velocity profile (peak velocity = 0.45 m s^{-1}, peak acceleration = 0.98 m s^{-2}) was delivered in one of three randomly interleaved stimulus modalities: vestibular (inertial motion without optic flow), visual (optic flow without inertial motion) and combined (synchronous inertial motion and optic flow). Although non-vestibular cues (for example, somatosensation and proprioception) were available during inertial motion, we refer to this condition as vestibular because both behavioral performance and MSTd responses strongly depend on intact vestibular labyrinths[12,21]. Optic flow stimuli accurately simulated movement of the observer through a three-dimensional cloud of random dots, providing multiple depth cues, such as relative dot size, motion parallax and binocular disparity via red-green glasses.

Across trials, the direction of translation (heading) was varied logarithmically in small steps around straight ahead (0°, ±1.23°, ±3.5° and ±10°; positive = rightward, negative = leftward). The monkey was required to fixate a central target (2° × 2° window) for the full 2-s stimulus presentation. After stimulus offset, two choice targets appeared and the monkey indicated whether his perceived heading was to the right or left of straight ahead via a saccade to the rightward or leftward target. A correct choice resulted in a juice reward and incorrect choices were not penalized.

The cue-conflict variant of this task was described previously[13]. Briefly, visual and vestibular heading trajectories were separated by a small conflict angle (Δ) on two-thirds of combined trials; the remaining third were nonconflict (Δ = 0°). The value Δ = +4° indicates that the visual trajectory was displaced 2° to the right and the vestibular trajectory 2° to the left of the assigned heading for a given trial, and vice versa for Δ = −4° (**Fig. 1a**). As in previous human psychophysical studies (for example, refs. 1,2,5), we designed the conflict angle to be large enough to probe the perceptual weights assigned to the two cues, but small enough to prevent subjects from completely discounting one cue (that is, engaging in robust estimation[6,42] or inferring separate causes[43,44]). On cue-conflict trials in which the heading was less than half the magnitude of Δ (that is, headings of 0° or ±1.23° when Δ = ±4°), the visual and vestibular cues specified a different sign of heading and thus the correct choice was undefined. These trials were rewarded irrespective of choice with a probability of 60–65%. This random reward schedule, present on an unpredictable 19% of trials, was unlikely to be detected by the monkeys, and did not appear to affect behavioral choices on those trials[13].

We varied the reliability of the visual cue by manipulating the motion coherence of the optic flow pattern. Motion coherence refers to the percentage of dots that moved coherently to simulate the intended heading direction, whereas the remainder were randomly relocated in the three-dimensional dot cloud on every video frame. Trials in the visual and combined modalities were assigned a motion coherence of either 16% (low) or 60% (high) while vestibular reliability was held constant. These two coherence levels were chosen to straddle the fixed reliability of the vestibular cue while avoiding extremes (for example, 0% and 100%) so that the less reliable cue would retain some influence on perception.

Neurophysiological recordings. We recorded extracellular single-unit activity in area MSTd using standard techniques as described previously[17]. MSTd was initially targeted using structural magnetic resonance imaging and verified according to known physiological response properties[45,46]. Once a neuron was isolated, we measured visual and vestibular heading selectivity during passive fixation by testing ten coarsely sampled directions in the horizontal plane (0°, ±22.5°, ±45°, ±90°, ±135° and 180° relative to straight ahead; 3–5 repetitions each, for a total of 60–100 trials). Only neurons with significant heading tuning (1-way ANOVA, $P < 0.05$) for both modalities were tested further, which constitutes 50–60% of neurons in MSTd[11]. Because we were interested in the effects of subtle angular displacements of heading (cue-conflicts) during the discrimination task, we also required at least one modality to have significant nonzero tuning slope over the three headings nearest to straight ahead (0° and ±22.5°). We performed linear regression on the firing rates for these three heading values, and if the 95% confidence interval of the regression slope did not include zero, the neuron was accepted and we began the discrimination task. This sample included 151 out of 362 well-isolated neurons (42%); however, an additional 43 neurons were rejected *post hoc*, either because isolation was lost before collecting enough stimulus repetitions (minimum of 5; nine neurons rejected), or because they lacked significant tuning in any stimulus modality over the narrower range of headings (±10°) tested in the discrimination task (34 neurons).

Note that one repetition of each heading and stimulus type in the discrimination task consisted of 63 trials: seven headings, three stimulus modalities, two coherence levels (visual and combined modalities only) and three conflict angles (combined modality only). We obtained as many repetitions as permitted by the stability of neuronal isolation and motivation of the animal. The average number of repetitions for the final sample of 108 neurons was 11.2, with 62% having at least ten repetitions (630 trials). Including fixation trials during the screening procedure, the number of trials required for an average dataset was 800–1,000, typically lasting 2–2.5 h. Neuronal responses for all analyses were defined as the firing rate over the middle 1 s of the stimulus duration, which contains most of the variation in both stimulus velocity and MSTd firing rates[11,47] (**Supplementary Fig. 5**).

Behavioral data analysis. For each trial type (combination of stimulus modality, coherence and conflict angle), we plotted the proportion of rightward choices by the monkey as a function of signed heading (**Fig. 1b–d**). We fit these data with cumulative Gaussian functions using a maximum-likelihood method (psignifit version 2.5.6)[48]. The psychophysical threshold and PSE were defined as the s.d. (σ) and mean (μ), respectively, of the best-fitting function. From the single-cue thresholds, we computed predicted weights for an optimal observer according to equation (2), and optimal thresholds according to equation (3).

To compare monkey behavior to optimal predictions, we computed observed weights from the PSEs in the cue-conflict conditions (Δ = ±4°) as follows. We start by rewriting equation (1), defining an internal signal S_{comb} that is a weighted sum of vestibular and visual signals S_{ves} and S_{vis} ($w_{vis} = 1 - w_{ves}$)

$$S_{comb} = w_{ves}S_{ves} + (1 - w_{ves})S_{vis}$$

Taking the mean of both sides yields

$$\mu_{comb} = w_{ves}\mu_{ves} + (1 - w_{ves})\mu_{vis} \quad (9)$$

such that

$$w_{ves} = \frac{\mu_{vis} - \mu_{comb}}{\mu_{vis} - \mu_{ves}} \quad (10)$$

Under normal conditions (that is, congruent stimulation), $\mu_{ves} = \mu_{vis} = \mu_{comb}$, and w_{ves} is undefined. However, in our cue-conflict conditions

$$\mu_{ves} = \theta - \frac{\Delta}{2} + b_{ves}$$
$$\mu_{vis} = \theta + \frac{\Delta}{2} + b_{vis} \quad (11)$$

where b_{ves} and b_{vis} are single-cue bias terms (assumed to be independent of θ). These biases, when estimated from behavioral data, are equal to $-PSE_{ves}$ and $-PSE_{vis}$, respectively. Thus, substituting equation (11) into equation (10), we have

$$w_{ves} = \frac{\theta + \frac{\Delta}{2} - PSE_{vis} + \mu_{comb}}{PSE_{ves} - PSE_{vis} + \Delta} \quad (12)$$

doi:10.1038/nn.2983

The combined PSE is the value of θ at which the observer has equal probability of a leftward or rightward choice, which occurs when $\mu_{comb} = 0$. Thus, we make the substitutions $\mu_{comb} = 0$ and $\theta = PSE_{comb}$ to obtain

$$w_{ves} = \frac{PSE_{comb} + PSE_{vis} + \frac{\Delta}{2}}{PSE_{ves} - PSE_{vis} + \Delta} \quad (13)$$

This derivation assumes that the observed bias in the combined condition ($-PSE_{comb}$ for $\Delta = 0$) is a weighted sum of biases in the single-cue conditions (equation (11)). Indeed, we did find a significant correlation (Pearson's $r = 0.64$, $P < 0.0001$) between $PSE_{comb,\Delta = 0}$ and the quantity $w_{ves\text{-}opt}{}^{*}PSE_{ves} + w_{vis\text{-}opt}{}^{*}PSE_{vis}$, where $w_{ves\text{-}opt}$ and $w_{vis\text{-}opt}$ come from equation (2). However, owing to unexplained behavioral variability, a substantial fraction of the variation in $PSE_{comb,\Delta = 0}$ across sessions was unrelated to PSE_{ves} and PSE_{vis} (see also ref. 13). Furthermore, we desired an expression that would apply equally well to our neurometric analyses (**Supplementary Figs. 1** and **2**), for which PSE_{ves} and PSE_{vis} are zero by construction. For these reasons, and for consistency with previous work[2,13], we made the simplifying assumption that the relevant bias is adequately captured by the measured $PSE_{comb,\Delta = 0}$. This amounts to setting $PSE_{ves} = PSE_{vis} = PSE_{comb,\Delta = 0}$ in equation (13), which yields our equation (4)

$$w_{ves-obs} = \frac{PSE_{comb} - PSE_{comb,\Delta = 0} + \frac{\Delta}{2}}{\Delta}$$

Reanalyzing the data using equation (13) instead of equation (4) did not significantly change the observed weights, either for behavior or neurons (Wilcoxon matched pairs test, $P > 0.05$ for both).

Observed weights were computed separately for $\Delta = -4°$ and $\Delta = +4°$ and then averaged for a given coherence level. All sessions having at least eight (mean = 12.6) stimulus repetitions ($N = 40$ sessions for monkey Y, 26 for monkey W) were included in behavioral analyses. For most analyses (such as those shown in **Fig. 2**), we pooled the choice data across sessions (separately for the two animals, except for **Fig. 6g,h** in which data were pooled across animals) and a single set of psychometric functions was fit to the data, yielding a single set of optimal and observed weights and thresholds. We computed 95% confidence intervals by resampling the choice data with replacement, refitting the psychometric functions, and recomputing weights and thresholds (bootstrap percentile method). Quantities for which 95% confidence intervals did not overlap were considered to be significantly different at $P < 0.05$.

Likelihood-based population decoding. We simulated individual trials of the behavioral task by decoding population activity patterns from our complete sample (or a desired subset) of MSTd neurons. The population response $\mathbf{r}$ for a given heading (θ) and set of conditions was used to compute the likelihood function $P(\mathbf{r}|\theta)$, assuming independent Poisson variability[14,22,24,49]

$$P(\mathbf{r}\,|\,\theta) = \prod_i \frac{e^{-f_i(\theta)} f_i(\theta)^{r_i}}{r_i!} \quad (14)$$

Here, f_i is the tuning function of the ith neuron in the population (linearly interpolated to 0.1° resolution) and r_i is the response (firing rate) of neuron i on that particular trial. Note that the likelihood is a function of θ, not $\mathbf{r}$; it specifies the relative likelihood of each possible θ (the parameter in conventional statistical usage) given the observed responses (the data or outcomes), and is not a probability distribution (that is, does not sum to 1).

To convert the likelihood function into a simulated choice, we assumed a flat prior distribution over θ, equating the normalized likelihood function to the posterior density $P(\theta|\mathbf{r})$. We then compared the summed posterior for negative headings to that for positive headings, and the ideal observer chose 'rightward' if the area under the curve was greater for positive headings. Other decision rules, such as taking the peak of the likelihood (maximum-likelihood estimate, MLE), or comparing the MLE to the peak of a 'reference' likelihood based on a simulated zero-heading trial[50], gave similar results. The integrated-posterior method produces identical results to MLE for symmetric (for example, Gaussian) likelihoods, but preserves optimality in the case of skewed or multi-peaked likelihoods, which occasionally occurred in our data. Each heading and trial type was repeated 100–200 times, and cumulative Gaussian functions were fit to the proportion of rightward choices versus heading. We used these simulated psychometric functions to compute optimal and observed weights and thresholds (and their confidence intervals), as described above (equations (2)–(4)). Further details and assumptions of the decoder are discussed in the **Supplementary Analysis**.

Linear model fitting of combined responses. Following reference 25, we computed the best-fitting weights of the linear model shown in equation (5). The combined responses were fit simultaneously for all three values of Δ, which required some linear interpolation and extrapolation of single-cue tuning curves in order to estimate responses at $\theta + \Delta/2$ and $\theta - \Delta/2$ for all necessary values of θ (only seven log-spaced headings were presented during the experiments). The fit was performed by simple linear regression. We did not include a constant term, unlike reference 25 (but similar to ref. 11), because the model was less well constrained by the narrower heading range in the present study, resulting in some large outliers when given the additional degree of freedom. However, the main effect of coherence on neural weights did not depend on the presence or absence of a constant term. We defined R^2 (**Fig. 7a,b**) as the square of the Pearson correlation coefficient (r) between the model and the data, and significant fits were taken as those with $P < 0.05$ (via transformation of the r value to a t statistic with $n - 2$ degrees of freedom). Only cells with significant fits for both coherence levels were included in the comparison of **Figure 7c,d** ($N = 83$ of 108 cells, 77%).

In the yoked-weights version of the model, A_{ves} and A_{vis} were constrained to have the same value across coherence levels (two free parameters), whereas the independent-weights version allowed different weights for each coherence (four free parameters). To test whether the fit was significantly better for the independent weights model, accounting for the difference in number of free parameters, we used a sequential F test for comparison of nested models.

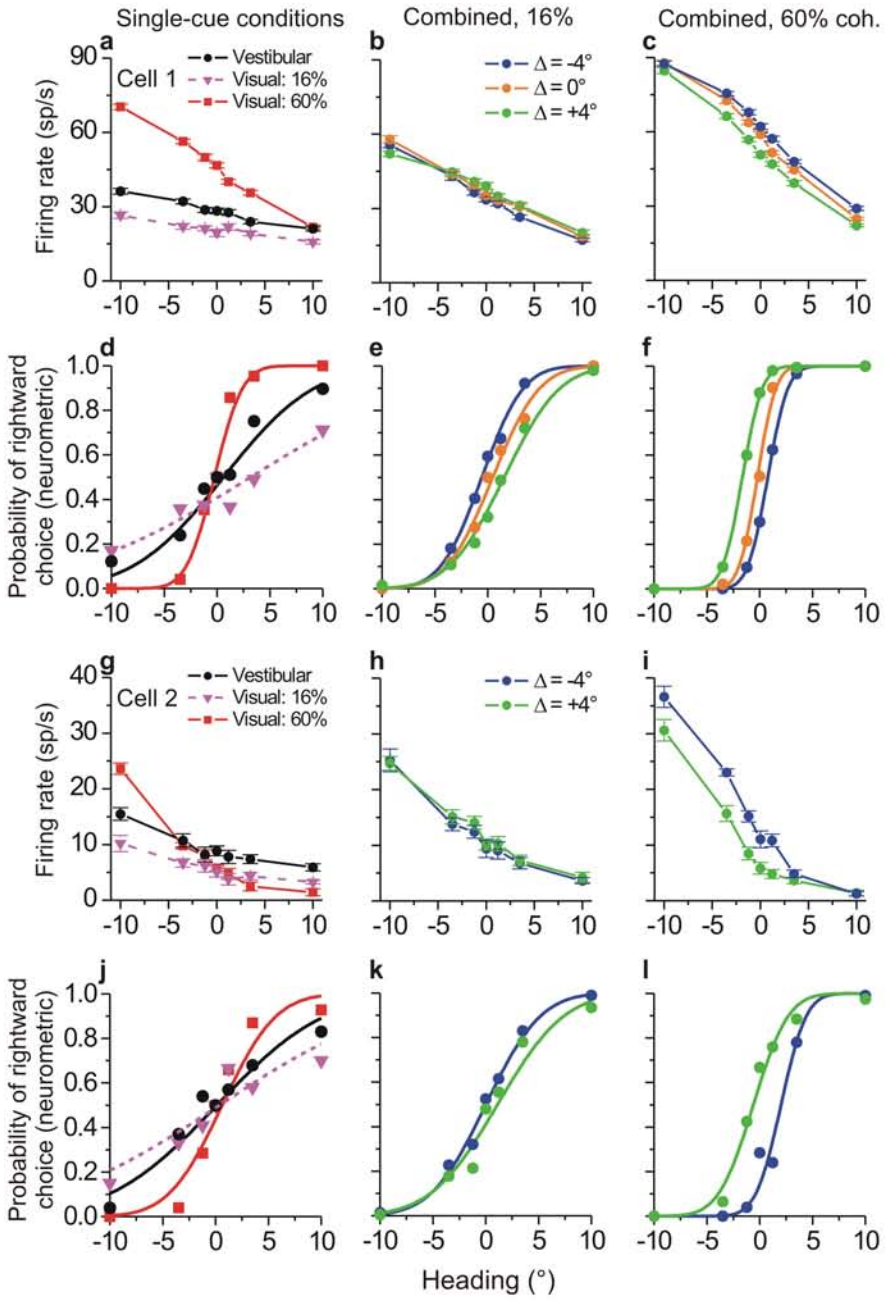
42. Knill, D.C. Robust cue integration: a Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. *J. Vis.* **7**, 1–24 (2007).
43. Cheng, K., Shettleworth, S.J., Huttenlocher, J. & Rieser, J.J. Bayesian integration of spatial information. *Psychol. Bull.* **133**, 625–637 (2007).
44. Körding, K.P. *et al.* Causal inference in multisensory perception. *PLoS ONE* **2**, e943 (2007).
45. Komatsu, H. & Wurtz, R.H. Relation of cortical areas MT and MST to pursuit eye movements. I. Localization and visual properties of neurons. *J. Neurophysiol.* **60**, 580–603 (1988).
46. Tanaka, K. & Saito, H. Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey. *J. Neurophysiol.* **62**, 626–641 (1989).
47. Fetsch, C.R. *et al.* Spatiotemporal properties of vestibular responses in area MSTd. *J. Neurophysiol.* **104**, 1506–1522 (2010).
48. Wichmann, F.A. & Hill, N.J. The psychometric function. I. Fitting, sampling, and goodness of fit. *Percept. Psychophys.* **63**, 1293–1313 (2001).
49. Jazayeri, M. & Movshon, J.A. Optimal representation of sensory information by neural populations. *Nat. Neurosci.* **9**, 690–696 (2006).
50. Gu, Y., Fetsch, C.R., Adeyemo, B., DeAngelis, G.C. & Angelaki, D.E. Decoding of MSTd population activity accounts for variations in the precision of heading perception. *Neuron* **66**, 596–609 (2010).

Single-cell neurometric (ROC) analysis

To compliment the population decoding approach (**Figs. 4, 6**, main text), we also tested the ability of individual MSTd neurons to account for cue reweighting behavior. Specifically, we used receiver operating characteristic (ROC) analysis to quantify the behavior of an ideal observer performing the same task as the animal but using only the firing rate responses of a given neuron. Response distributions for each nonzero heading were compared to the distribution for heading = 0° (taken as a proxy for the internal reference) to compute the proportion of rightward choices made by the ideal observer. As with the psychophysical analysis (**Fig. 1b-d**), these 'neurometric' choice data were fit with cumulative Gaussian functions, and the neuronal PSE and neuronal threshold were taken as the mean and standard deviation of the fitted function, respectively.

For the example neuron shown in **Fig. 3** (replotted in **Fig. S1a-c**), neuronal thresholds in the single-cue trials were 6.7°, 13.6°, and 2.1° for the vestibular, low-coherence visual, and high-coherence visual conditions, respectively (**Fig. S1d**). These thresholds quantify the neuron's sensitivity to small variations in heading for each modality, which is primarily dictated by the slopes of the respective tuning functions. Using Eq. 2, single-cue neuronal thresholds were used to generate optimal vestibular weights for the ideal observer (here, 0.80 and 0.08 for low and high coherence, respectively). These were compared to observed weights computed with Eq. 4, using the neurometric PSEs from the combined, cue-conflict trials (**Fig. S1e,f**). Note that the reference distribution (responses at heading = 0°) was taken from the $\Delta = 0°$ trials, regardless of the value of Δ. The use of non-conflict responses for the reference is a sensible strategy because the monkey cannot predict whether a trial will be a cue-conflict trial, and may not even detect the conflict when present[1]. Because the firing rates during cue-conflict (green and blue traces in **Fig. 3b,c** and **Fig. S1b,c**) were compared with the non-conflict trials (orange traces), the shift of the tuning curves led to shifts of the neurometric functions, in a manner

Nature Neuroscience: doi:10.1038/nn.2983



Supplementary Figure S1. Example neurons and neurometric functions via ROC analysis

a-c (cell 1, replotted from **Fig. 3**) and **g-i** (cell 2) depict tuning curves from two example MSTd neurons. Mean firing rate (spikes/s) ± s.e.m. is plotted as a function of heading for the single-cue trials (**a,g**), and combined trials at low (**b,h**) and high (**c,i**) coherence. Non-conflict (Δ=0°) data for example cell 2 are omitted for clarity. **d-f** and **j-l** show neurometric functions for the corresponding firing rate data in **a-c** and **g-i**, respectively (compare with behavioral effects shown in **Fig. 1b-d**). The analysis is described in detail on p. 1, 3 of this document. Briefly, the probability of a rightward choice by an ideal observer was computed using ROC analysis by comparing the firing rate distribution at each heading value with the distribution at heading = 0° (and Δ = 0°, for the combined data). As with the behavioral analysis, these neurometric functions were used to compute optimal and observed vestibular weights (Eqs. 2 and 4) for each coherence level. For the two example neurons, these weights were as follows:

	Cell 1		Cell 2	
	optimal	observed	opt.	obs.
16% coh.	0.80	0.76	0.72	0.64
60% coh.	0.08	0.16	0.20	0.17

analogous to the psychometric functions (**Fig. 1c,d**). For this neuron, the observed weights corresponding to these PSE shifts were 0.76 and 0.16 for low and high coherences, respectively. A second example cell is shown in **Fig. S1g-l**.

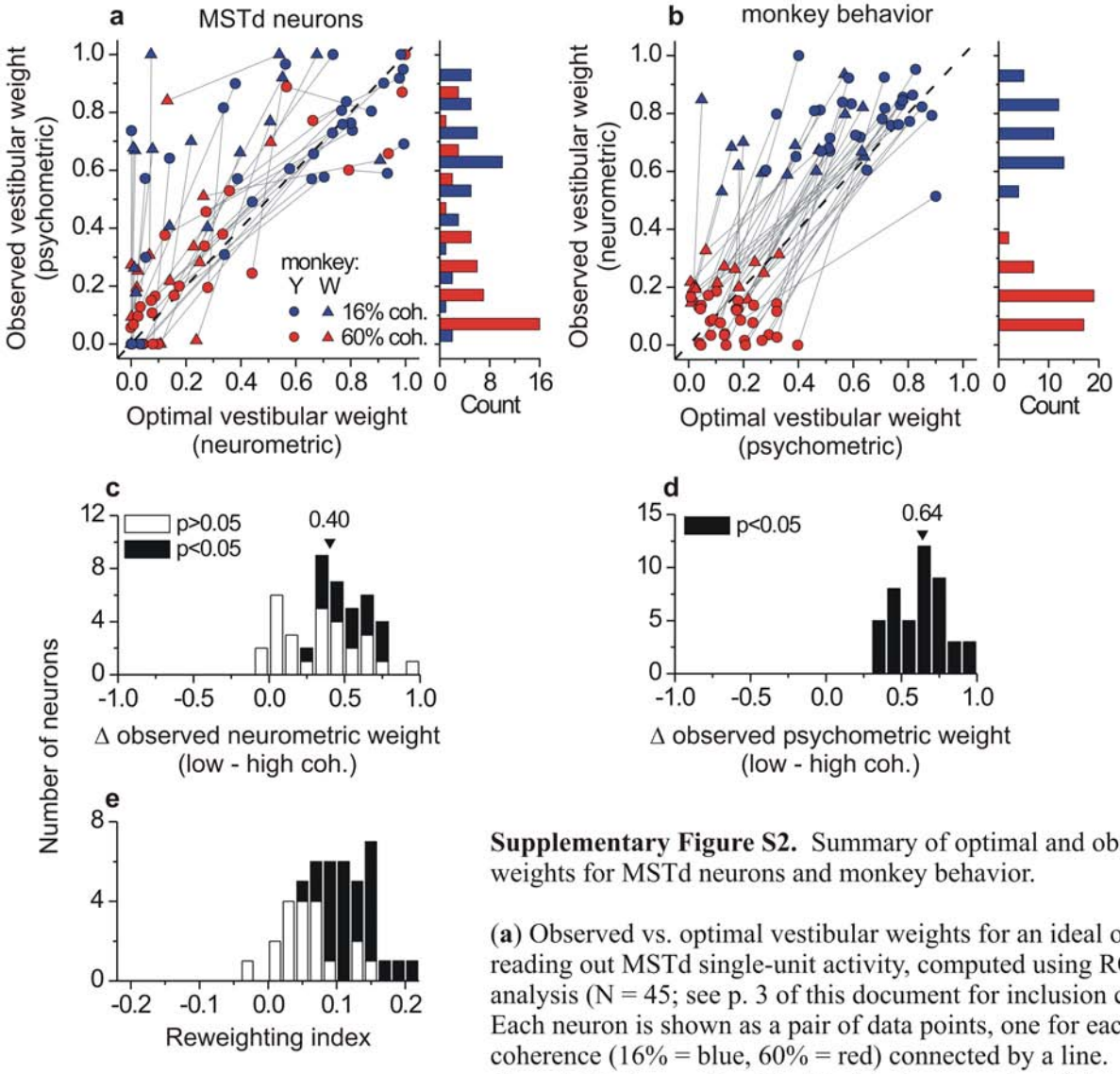
Because the ROC analysis approach depended on reliable PSE estimates from neurometric functions, it could only be performed on neurons with clear, monotonic tuning around straight ahead.

Nature Neuroscience: doi:10.1038/nn.2983

Otherwise, the neurometric functions were too flat to provide reliable estimates of the neuronal thresholds and PSEs. Thus, we required the slope of the linear regression of firing rate vs. heading during the discrimination task to be significantly greater or less than zero for at least one single-cue modality, as well as the combined modality at both coherence levels. These criteria resulted in a sample of 45 neurons (30 from monkey Y, 15 from monkey W). This sample excluded most neurons with visual and vestibular tuning that was clearly opposite in slope ('opposite cells'[2], see **Fig. 5**), as the combined tuning for these neurons was often too flat or non-monotonic. Of the 23 neurons classified as opposite (congruency index (CI) < -0.4), all but one failed to meet the criteria for ROC analysis, whereas 34 of the 46 congruent cells (CI > 0.4) passed these criteria.

Weights for this subpopulation of MSTd neurons are summarized in **Fig. S2a**, with each neuron represented by two data points (one for each coherence, connected by a thin line) and plotted as observed vs. optimal vestibular weight. The main result to take away from this figure is that most cells showed large changes in observed weights during combined trials (ordinate) as a function of coherence. Median values for low (0.691) and high (0.195) coherence were highly significantly different (Wilcoxon matched-pairs signed rank test, $p < 10^{-7}$). The paired difference in observed weights (low minus high coherence) was significantly greater than zero ($p < 0.05$, two-tailed bootstrap test) for 17 of 45 neurons (38%) and not significantly less than zero for any neurons (**Fig. S2c**). These results suggest that MSTd neurons respond robustly to trial-by-trial changes in cue reliability. Notably, these effects occurred for changes in reliability that were behaviorally relevant: the two coherence levels were chosen to be effective at driving behavioral reweighting (i.e., **Fig. 2a,b**), and were not tailored to drive changes in neurometric weights. Psychophysical weights from these 45 sessions are shown in the same formats for comparison (**Fig. S2b,d**).

Nature Neuroscience: doi:10.1038/nn.2983



Supplementary Figure S2. Summary of optimal and observed weights for MSTd neurons and monkey behavior.

(**a**) Observed vs. optimal vestibular weights for an ideal observer reading out MSTd single-unit activity, computed using ROC analysis (N = 45; see p. 3 of this document for inclusion criteria). Each neuron is shown as a pair of data points, one for each coherence (16% = blue, 60% = red) connected by a line. Symbol shape indicates monkey identity (circles = monkey Y, triangles = monkey W). Marginal histogram shows the distribution of observed weights for the two coherence levels. A handful of cells had observed weights greater than 1 or less than 0 (i.e., a PSE shift greater than $|\Delta|/2$), mainly because of uncertainty in the PSEs estimated from the cumulative Gaussian fits. These were plotted at 0 or 1 for display purposes. (**b**) Same analysis and format as **a**, but for the psychophysical data from the corresponding 45 sessions. (**c**) Histogram of paired differences in observed weights (low minus high coherence) derived from ROC analysis (i.e., the vertical component of the connecting lines in **a**). Black bars indicate individually significant differences ($p < 0.05$, bootstrap). (**d**) Same format as **c**, but for the psychophysical data shown in **b**. (**e**) Histogram of a model-free 'reweighting index' used to confirm the result in **c** without relying on ROC analysis (see p. 8-9 of this document).

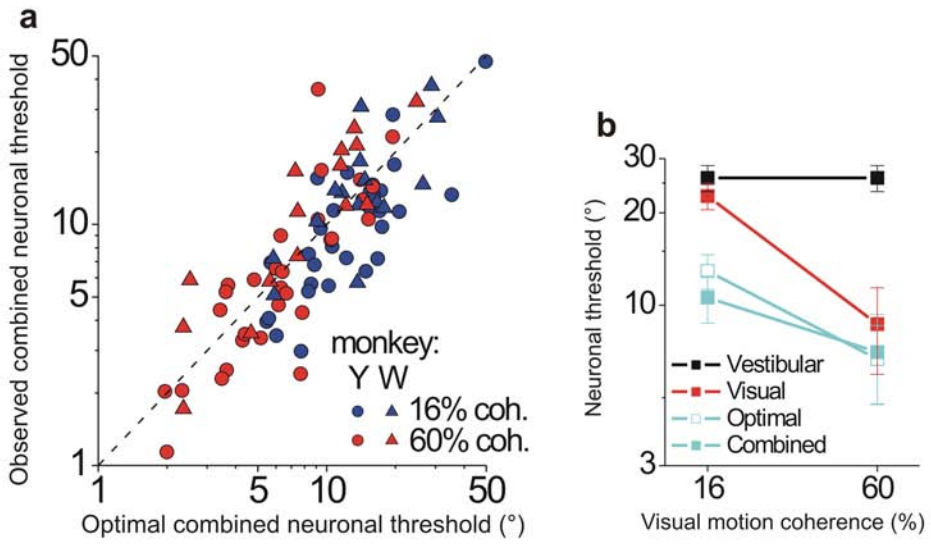
Nature Neuroscience: doi:10.1038/nn.2983

The second main result from **Fig. S2a** involves the comparison between optimal and observed weights. Many of the data points and connecting lines trend along the unity-slope diagonal, indicating weights that were close to optimal predictions. Optimal and observed weights were significantly correlated for both coherences together (Spearman's rank correlation, rho = 0.73, $p < 0.0001$) and for each coherence separately (16%: rho = 0.49, $p = 0.0006$; 60%: rho = 0.77, $p < 10^{-9}$). However, many data points also lie well above the unity-slope diagonal (observed weight > optimal weight; Wilcoxon matched-pairs signed rank test: $p < 0.0001$ for both monkeys together and monkey W individually; $p = 0.21$ for monkey Y individually), some with nearly vertical connecting lines. These cells tended to have relatively flat vestibular tuning and therefore low optimal vestibular weights. Yet, the vestibular cue nonetheless had a sizeable influence on the combined tuning, resulting in observed vestibular weights that were greater than the prediction. This apparent deviation from optimality (as defined by Eq. 2) is intriguing because it parallels the over-weighting of the vestibular cue at low coherence, observed in behavior (**Fig. 2a,b** and **Fig. S2b**). Statistical comparisons from **Fig. S2a** also held for **Fig. S2b** (observed weight > optimal weight; $p < 0.001$ for both monkeys together and monkey W individually; $p = 0.28$ for monkey Y individually), although a notable difference is the vestibular under-weighting (visual over-weighting) for monkey Y at 60% coherence (**Fig. S2b**, red circles below the diagonal; also evident in **Fig. 2a**).

In summary, individual MSTd neurons reweight visual and vestibular cues on a trial-by-trial basis, supporting a possible role in perceptual reweighting and consistent with the population decoding results (**Figs. 4, 6**). Many cells perform nearly optimally, while a subset appears to over-weight vestibular cues, similar to the monkeys' behavior at low coherence. This unexpected link between neuronal activity and a modality-specific perceptual bias adds support for the hypothesis that MSTd neurons contribute to visual-vestibular cue integration for heading perception. Lastly, neuronal

Nature Neuroscience: doi:10.1038/nn.2983

thresholds for this sample (**Fig. S3**) decreased under combined stimulation, in accordance with optimal predictions (Eq. 3) and consistent with our previous work[2].



Supplementary Figure S3. Summary of neuronal thresholds based on ROC analysis.

(**a**) Optimal vs. observed neuronal thresholds in the combined modality, including only the sample of 45 neurons deemed suitable for ROC analysis (as in **Supplementary Fig. S2**). Optimal thresholds for each neuron were computed from single-cue neuronal thresholds using Eq. 5, and observed thresholds were taken from the combined modality at $\Delta = 0°$. Colors/symbols as in **Supplementary Fig. S2**. The Spearman rank correlation between optimal and observed thresholds was highly significant for both coherence levels and both monkeys individually (16% coh., monkey Y: rho=0.75, p<0.0001; 16% coh., monkey W: rho=0.68, p=0.005; 60% coh., monkey Y: rho=0.82, p<0.0001; 60% coh., monkey W: rho=0.87, p<0.0001).

(**b**) Summary of neuronal threshold data in **a**, as well as the single-cue conditions for comparison. Points indicate geometric means and error bars are geometric standard deviations.

Similar to previous findings (Gu et al., 2008), when single-cue thresholds were roughly matched (our 16% coherence data), combined thresholds were significantly lower than the best single-cue condition (Wilcoxon matched pairs test, $p < 0.0001$). In fact, they were slightly but significantly less than the 'optimal' prediction ($p = 0.007$). However, note that our definition of optimality (Eq. 3) assumes a particular linear cue integration scheme (primarily used to describe behavioral performance), and does not necessarily represent an upper limit on neuronal sensitivity. At 60% coherence, thresholds were not significantly different from the prediction ($p = 0.28$), and slightly but not significantly ($p = 0.34$) lower than the best single-cue (visual).

Nature Neuroscience: doi:10.1038/nn.2983

ROC analysis approach: assumptions and caveats

The neurometric analysis described above assumes that (a) the ideal observer knows the preferred heading (sign of tuning slope) for each neuron and modality, and (b) the response variability at zero heading is a reasonable approximation of the uncertainty in the internal reference against which the monkey is discriminating. The first assumption is common to all similar approaches and is necessary for the observer to interpret firing rates as signaling a particular stimulus in the world. The second assumption is an alternative to the 'neuron-antineuron' framework used in some studies for measuring neuronal sensitivity with ROC analysis, including our own previous work[2]. In that procedure, the response distribution for a particular heading would be compared with the distribution for the corresponding heading of opposite sign, thereby invoking an antineuron with mirror-symmetric tuning. Instead, we found that using the zero-heading responses as the reference distribution was more intuitive when the goal was to use neuronal PSEs during cue-conflict as a measure of cue reweighting.

The main effect of changing our analysis to a neuron-antineuron approach would be to increase neuronal sensitivity (decrease neuronal thresholds) across the board; i.e., for monotonic tuning functions, there will generally be greater separation between the firing rate distributions for $+\theta$ vs. $-\theta$ for a given θ, as opposed to each θ vs. zero heading. However, this change would not alter the *relative* neuronal thresholds across modalities (i.e., optimal weights), nor the PSEs during cue-conflict (observed weights), provided that the non-conflict ($\Delta = 0^{\circ}$) responses were still used as the reference.

Reweighting index

To ensure that the results of **Fig. S2a,c** were not dependent on any assumptions of ROC analysis or the fitting of neurometric functions, we computed a model-free 'reweighting index' (**Fig. S2e**) for the same sample of neurons, as follows. First we computed the mean normalized change in firing rate

Nature Neuroscience: doi:10.1038/nn.2983

('delta-FR') caused by the cue-conflict (*see note below), separately for each coherence level, and collapsed across positive and negative Δ:

$$\textit{delta-FR} = \frac{1}{2}\sum_{\Delta\in-4,+4}\frac{1}{7}\sum_{h=1}^{7}\frac{1}{N}\sum_{t=1}^{N}\frac{r_t(h,\Delta)-\bar{r}(h)}{\bar{r}_{\max}^{\Delta=0}-\bar{r}_{\min}^{\Delta=0}} * \mathrm{sgn}(\Delta) * \mathrm{sgn}(\textit{slope}) \qquad \text{(S1)}$$

$$\textit{Reweighting Index} = \textit{delta-FR}_{\text{coh=60\%}} - \textit{delta-FR}_{\text{coh=16\%}} \qquad \text{(S2)}$$

In this equation, the index *h* refers to heading, *t* refers to each of N trials, $r_t(h,\Delta)$ is the single-trial firing rate in the combined modality for Δ = ±4º, and *r*-bar is the mean combined firing rate for Δ = 0º. The normalization factor in the denominator is the total firing rate modulation (max-min) for the Δ = 0º tuning curve. Sgn is the sign function (signum), and *slope* refers to the slope of the combined tuning curve based on linear regression. These sign functions preserve the expected relationship between cue reliability and firing rate as a function of Δ and tuning slope – i.e., for the examples in **Fig. S1c** and **Fig. S1i** the tuning curve for Δ = +4º (green) shifts downward, but would have shifted upward for that same Δ and coherence if the tuning slope was positive instead of negative, and we wanted the delta-FR metric to have the same sign in both cases. The reweighting index itself was then computed as the difference between the delta-FR values for high versus low coherence, giving positive values for shifts of the tuning curve in the predicted direction. Statistical significance of the reweighting index was determined by bootstrapping the delta-FR metrics 1000 times (resampling across trials and heading values with replacement), recomputing the reweighting index each time, and determining whether more than 95% of the bootstrapped index values were greater or less than zero (percentile method at $p < 0.05$). The resulting distribution of reweighting index values was strongly shifted toward positive values (t test, $p < 0.0001$), with 26/45 (58%) of neurons having values significantly greater than zero ($p < 0.05$, bootstrap) and none significantly less than zero (**Fig. S2e**).

Nature Neuroscience: doi:10.1038/nn.2983

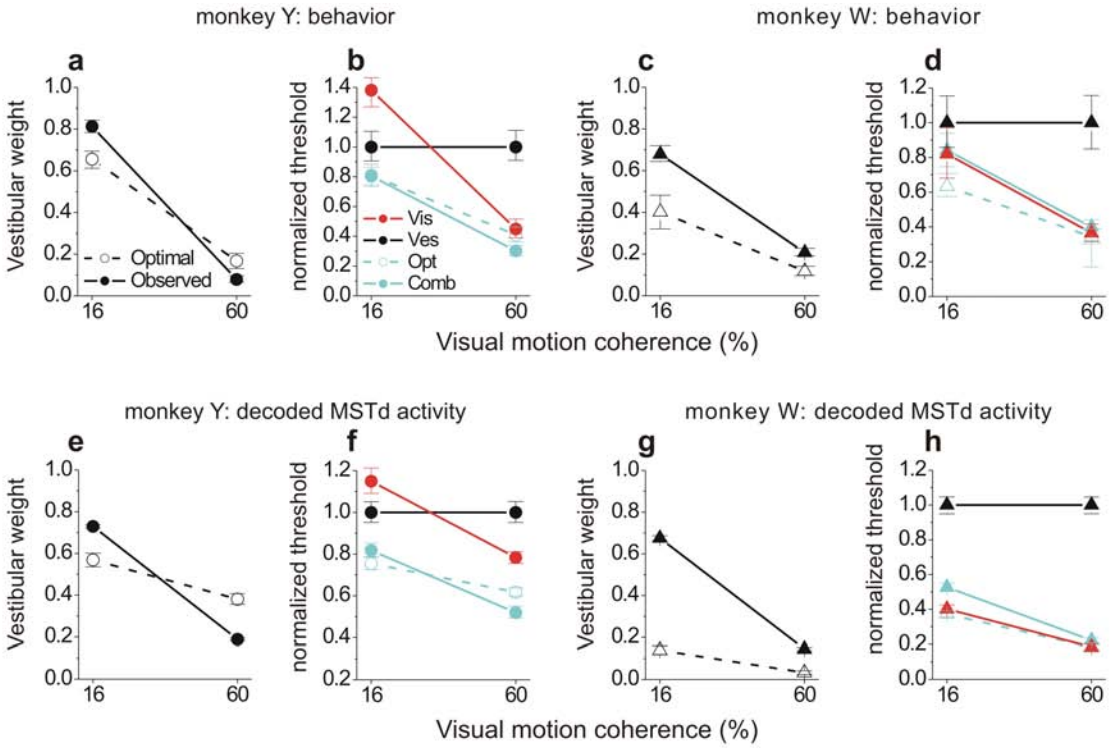
[*Note: the effect of cue-conflict on firing rates may be more aptly described as a lateral tuning shift, since the conflict manipulation itself is a 'lateral' one (change in heading specified by the constituent cues). However, for the monotonic tuning curves in our sample, this was largely indistinguishable from a vertical 'delta-FR', and we found the latter to be more reliable than common methods for estimating lateral shifts, such as curve fitting or cross-correlation (particularly for cells with relatively shallow tuning).]

Likelihood-based decoder: assumptions and caveats

Similar to the ROC analysis method for single neurons (see above), the decoder assumes knowledge of the neuronal tuning curves (f_i in Eq. 14) at the readout stage. In previous theoretical treatments[3-6], the choice of what to use for f_i was clear, mainly because only single-modality likelihood functions (or log-likelihoods) were being computed. In contrast, we are decoding from real neurons in 9 distinct trial types (3 stimulus modalities, 2 coherence levels, and 3 conflict angles), and therefore must make a decision about which tuning curves to substitute for f_i when computing the likelihood functions. We chose to allow the decoder full knowledge of the neuronal tuning curves in every condition – i.e., $f_i(\theta)$ was taken as the tuning curve in the particular condition (stimulus modality and coherence level) being simulated on that trial. An important exception to this was the combined, cue-conflict ($\Delta = \pm 4^\circ$) trials. These trials were decoded with respect to the non-conflict ($\Delta = 0^\circ$) tuning curve, to allow the simulation to produce a shift in the PSEs similar to the monkeys' behavior (e.g., see **Fig. S4**).

While we cannot be certain that the brain is capable of such a dynamic readout strategy, it may be feasible in our task because the subject knows the modality (and possibly the coherence level) fairly early in the trial. The presence or absence of platform motion is easily detected from vibration and auditory cues, and the approximate coherence ('low' or 'high', given that only 2 coherence levels are

Nature Neuroscience: doi:10.1038/nn.2983



Supplementary Figure S4. Comparison of behavior and decoded MSTd activity for each monkey separately.

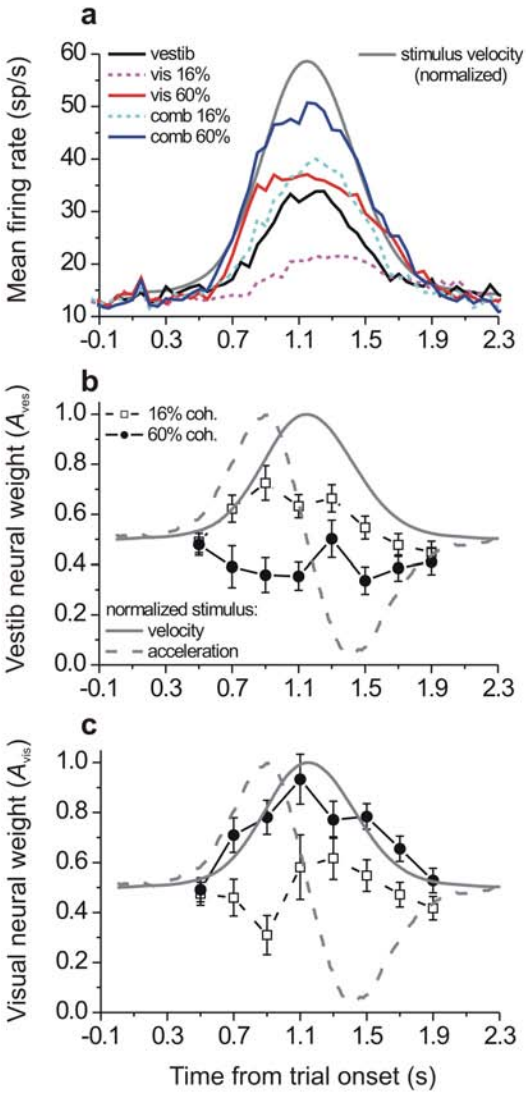
Weights and normalized thresholds are plotted separately for the two monkeys. The top row (**a-d**) shows the behavioral data (replotted from **Fig.2** for ease of comparison), while the bottom row (**e-h**) summarizes the performance of the population decoder using only the neurons recorded in the corresponding animals. Only congruent cells (congruency index > 0.4; see **Fig. 5**) were included. Error bars indicate 95% confidence intervals.

Despite the small sample sizes (N = 32 for monkey Y, 14 for monkey W), the pattern of results from the decoder was qualitatively similar in several respects to the behavior of each animal individually. In particular, the presence or absence of an improvement in combined thresholds (present for monkey Y, absent for monkey W) and the pattern of optimal vs. observed weights (monkey Y: observed > optimal for 16% coherence, vice versa for 60% coh.; monkey W: observed >> optimal for both coherences; $p < 0.05$ by bootstrap for each comparison) were apparent in both behavior and decoded MSTd activity.

possible) is apparent from the dynamics of the random dots before they begin to clearly simulate a particular heading trajectory (see **Fig. S5** for stimulus temporal profile).

The alternative strategy of using a fixed tuning curve for all conditions (e.g., the single-cue vestibular tuning) produced similar results with regard to cue reweighting. However, doing so hampered our attempts to analyze populations with substantial numbers of opposite cells. For example, when

Nature Neuroscience: doi:10.1038/nn.2983



Supplementary Figure S5. Time course of firing rate (PSTH) and neural weights.

(**a**) Mean firing rate as a function of time within trials (0 = onset of fixation), averaged across our sample of congruent MSTd neurons (N=46). Prior to averaging, responses were taken from the heading eliciting the largest firing rate during the discrimination task, determined separately for each condition/modality. Smooth gray curve shows Gaussian stimulus velocity profile (normalized to fit in window).

(**b,c**) Neural weights (see Eq. 5, main text) as a function of time within trials (A_{ves}, panel **b**; A_{vis}, panel **c**; 16% coherence, dashed lines and open squares; 60% coh., solid lines and filled circles), averaged across the congruent neuron sample (excluding 3 neurons with poor fits to the linear model; error bars = s.e.m.). Firing rates were computed in 200-ms non-overlapping windows, and the resulting tuning curves were used to re-fit the linear weighted sum model (Eq. 5) at each data point. Points are plotted at the window centers. The earliest data point shown represents the 400-600 ms window (plotted at 0.5 s) because responses did not begin to rise above baseline until then (see panel **a**). Normalized stimulus velocity (gray solid trace) and acceleration (gray dashed trace) are overlaid for comparison.

The results of **b,c** suggest that the separation of weights as a function of coherence (Fig. 7c,d; see also Morgan et al., *Neuron* 2008) occurs very soon after firing rates rise above baseline, and persists throughout most of the trial. This rapid time course is consistent with an underlying network mechanism such as divisive normalization (Ohshiro et al., *Nat. Nsci.* 2011).

decoding relative to the vestibular tuning curves, the simulation made too many errors in the visual modality, and the choice data could not be fit with a monotonically increasing function such as a cumulative Gaussian. Furthermore, even for congruent or mixed populations, using a tuning curve from one stimulus modality to decode another can produce large choice biases due to offsets between the two tuning curves. Thus, to ensure an unbiased and complete picture of the information contained in our MSTd population, we felt the 'full knowledge' approach was best suited for our purposes.

Nature Neuroscience: doi:10.1038/nn.2983

The response of each neuron in the population (r_i) can in principle be drawn from the actual single-trial responses for the desired heading and trial type. However, because population sensitivity (i.e., signal to noise ratio) increases with N, we found that for larger population sizes (N > ~40) the simulated psychometric functions (e.g., **Fig. 4c,f**; note abscissa scale) became too steep to be reliably fit using only the heading angles used in the experiments (0º, ±1.23º, ±3.5º, and ±10º). Thus, it became necessary to interpolate the tuning curves (linear, 0.1º resolution) and generate simulated firing rate distributions at this finer heading resolution. Responses were typically generated by sampling a Poisson process (*poissrnd* in Matlab) with the λ parameter (mean and variance) equal to the mean (interpolated) response. The results of the decoding model using this procedure were compared to the method of drawing real single-trial firing rates (when smaller population sizes permitted it), and no significant differences were found. As the Poisson method restricts the variance-to-mean ratio (Fano factor) to be 1, we also tested versions of the model using Gaussian noise and various Fano factors ranging from 1 to 2. We found no substantial differences between the two noise models with regard to the main conclusions of the study. Increasing the Fano factor only had a noticeable effect on the absolute thresholds for a given population size, not on the pattern of relative thresholds across modalities (optimal weights), nor the PSEs during cue-conflict (observed weights).

An additional assumption of this version of the decoder is that each neuron is independent, i.e. that noise correlations are zero. Employing the method of Shadlen et al.[7] for generating arbitrary correlation structure in simulated data, we tested other versions that included small positive correlations characteristic of MSTd neuron pairs recorded in the heading discrimination task[8]. As expected from the known effects of noise correlations on the fidelity of population codes[9, 10], positive correlations increased the simulated thresholds for a given population size. However, similar to the case of

Nature Neuroscience: doi:10.1038/nn.2983

increasing Fano factor, the relative single-cue thresholds and the combined PSEs were largely unaffected by correlations, and thus the conclusions with regard to cue reweighting did not change.

The one case where the basic Poisson decoding method seemed to fail was with artificial combined responses, generated by taking weighted sums of single-cue responses (**Fig. 8b-e**). Initially, we performed these simulations by generating combined tuning curves (mean firing rates) as weighted sums of single-cue tuning curves, then making draws from a Poisson process at the resulting means. This procedure failed to reproduce appropriate distributions of single-trial firing rates, as evidenced by the fact that even the decoding of optimal responses (i.e., generated with weights according to Eq. 7) did not result in optimal behavior according to Eqs. 2 and 3 (not shown). This problem was remedied by taking weighted sums of *single-trial* responses (rather than mean responses) to generate the artificial combined data. Since the weights in the summation were not fixed at 1, the resulting response distributions were typically not Poisson, requiring us to use an ad-hoc decoding method in place of Eq. 14. This ad-hoc method began with defining the distribution $P(\mathbf{r}_i|\theta)$ empirically, as a normalized histogram of single-trial firing rates for a given heading and cell. We then selected the probability value of the nearest bin corresponding to the particular r being simulated, and repeated this for all neurons and headings to essentially 'bootstrap' the full likelihood function $P(\mathbf{r}|\theta)$. This procedure was validated by the near-perfect match between optimal and observed weights and thresholds in **Fig. 8b,c**, and was also able to reproduce the basic pattern of results shown in **Fig. 6e,f**.

Variance as a signature of cue combination?

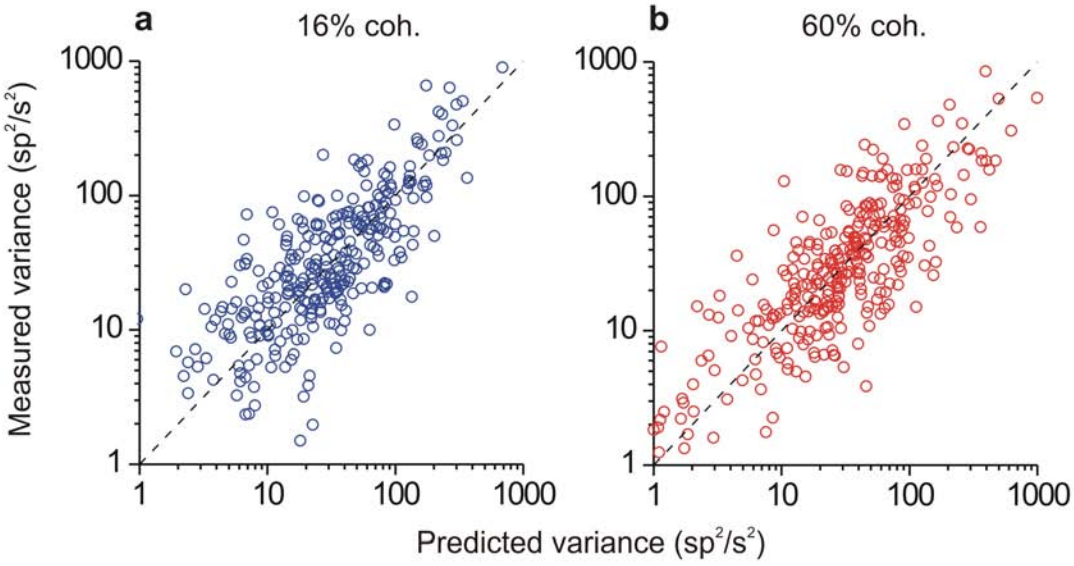
The fact that our decoder failed when taking weighted sums of mean single-cue responses and then drawing Poisson random deviates raises an interesting set of questions about the variance of multisensory responses. Assuming, as we have shown (**Fig. 7** and ref. [11]), that mean combined responses

Nature Neuroscience: doi:10.1038/nn.2983

are reasonably well modeled as weighted sums of single-cue responses, we can ask whether the variance of the combined response behaves as predicted from applying these weights. We tested this by applying the weights (derived from fits to the tuning curves; see main text) to the individual trial firing rates in the unimodal vestibular and visual conditions, then comparing the variance of these artificial weighted-sum 'trials' against the measured variance in the combined conditions. This comparison is plotted in **Fig. S6a,b** for low and high coherence, respectively (one data point per heading per neuron, $\Delta=0^{o}$ trials only; sample includes only congruent cells that were significantly well fit by the linear model, as in **Fig. 8b-e** (N=43)). The predicted (artificial) variance was well correlated with, and not significantly different from, the measured variance across both levels of reliability (Spearman's rho = 0.75 for 16% coh., rho = 0.77 for 60% coh., $p << 10^{-10}$ for both; Wilcoxon matched pairs test, p = 0.096 for 16% coh., p = 0.78 for 60% coh.).

This match is not trivial because the weights change rapidly with cue reliability (**Fig. 7c,d**) in a manner consistent with a network nonlinearity such as divisive normalization[12] (see also Discussion and **Fig. S5**). With such a nonlinearity at play, one might expect some degree of information loss during cue combination, which would imply greater response variance than predicted by a weighted sum of single-cue trials. The fact that combined variance is as predicted by the weighted sum model lends further support for the role of MSTd in contributing to optimal cue integration, as it suggests the combination can occur without substantial information loss. This apparent summation of variance during cue integration is roughly analogous to a recent result[13] showing an increase in variance in area LIP across time within a trial, due to the accumulation of momentary evidence during perceptual decision-making. Although a detailed examination of this phenomenon is beyond the scope of our study, it may be that variance becomes another useful tool for diagnosing whether and how candidate multisensory neuronal populations carry out the necessary computations for sensory cue integration.

Nature Neuroscience: doi:10.1038/nn.2983



Supplementary Figure S6. Combined response variance versus a prediction from the weighted sum of single-cue firing rates

For each neuron and heading, the variance of the combined responses for 16% coherence (**a**) and 60% coherence (**b**) trials is plotted against the variance of the weighted sum of single-cue trials having the corresponding heading and coherence ("predicted" variance). The weights in this sum were taken from the best-fitting linear model for each cell (**Fig. 7c,d**, main text), and the sample includes only congruent neurons with significant R^2 values (N=43).

The measured and predicted variance was strongly correlated and not significantly different from the unity line for both coherence levels (Spearman's rho = 0.75 for 16% coh., rho = 0.77 for 60% coh., $p << 10^{-10}$ for both; Wilcoxon matched pairs test, p = 0.096 for 16% coh., p = 0.78 for 60% coh.). This supports the plausibility of the linear weighted sum model, and suggests that cue combination by MSTd neurons may occur without a substantial increase in variance (and therefore loss of information).

Derivation of optimal neural weights

This section describes a derivation of the optimal weight ratio ρ_{opt} (Eq. 7 in main text), which we compared to the measured weight ratio in **Fig. 8a**. Let $f_{ves}(\theta)$, $f_{vis}(\theta,c)$, and $f_{comb}(\theta,c)$ represent the mean response (tuning function) of a particular cell for the three respective stimulus modalities as a function of heading (θ) and coherence (c). For mathematical convenience, we replace the neural weights A_{ves} and A_{vis} in Eq. 5 with a weight ratio $\rho = A_{ves} / A_{vis}$ and include a scaling constant a. We then have (dropping the θ and c dependence for clarity):

Nature Neuroscience: doi:10.1038/nn.2983

$$f_{comb} = a(\rho f_{ves} + f_{vis})$$
$$f'_{comb} = a(\rho f'_{ves} + f'_{vis}) \quad \text{(S3)}$$

where f' refers to the derivative of the tuning functions. If we assume that the spike count variability on single-cue trials is 'Poisson-like' (i.e., can be well approximated by a broad family of functions known as the exponential family with linear sufficient statistics[14-16], a type of variability believed to be common across cortical areas[15, 17-19], including in our MSTd recordings (Y. Gu, CRF, DEA and GCD, unpublished observations)), the variance σ^2 will be proportional to the mean, where the proportionality constant is the Fano factor (F = variance/mean). Because the variance of a sum of independent random variables is the sum of their respective variances, and applying the identity $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$, we have:

$$\sigma^2_{comb} = a^2(\rho^2\sigma^2_{ves} + \sigma^2_{vis}) = a^2(\rho^2 F f_{ves} + F f_{vis}) \quad \text{(S4)}$$

What we seek is the value of ρ that maximizes linear Fisher information (the information that can be recovered by an optimal linear decoder) from the combined responses (I_{comb}), which is defined as[20]:

$$I_{comb} = \frac{f'^2_{comb}}{\sigma^2_{comb}} = \frac{a^2(\rho f'_{ves} + f'_{vis})^2}{a^2(\rho^2 F f_{ves} + F f_{vis})} \quad \text{(S5)}$$

To find the weight that maximizes information we need to solve:

$$\frac{\partial I_{comb}}{\partial \rho} = 0 \quad \text{(S6)}$$

Solving this equation (note that the constants a and F drop out) gives the following expression for the optimal neural weight ratio ρ_{opt}:

$$\rho_{opt} = \frac{f'_{ves}(0) f_{vis}(0,c)}{f_{ves}(0) f'_{vis}(0,c)} \quad \text{(Eq. 7 in text)}$$

Nature Neuroscience: doi:10.1038/nn.2983

(with θ and c dependence restored, to illustrate that this ratio is evaluated at $\theta = 0$ and computed separately for the two coherence levels (**Fig. 8a**)). Note: although it appears that we have derived the weights maximizing linear Fisher information only, Eq. 7 in fact maximizes the full Fisher information. Indeed, it is easy to show that for the weight ratio above, the combined responses also belong to the Poisson-like family, and for the Poisson-like family it turns out that Fisher information is equal to linear Fisher information[20].

As mentioned in the main text, Eq. 7 describes the neural operation one would expect to see in an area performing optimal multisensory integration within a probabilistic population code (PPC) framework[16]. By taking into account the observed tuning parameters (which define the mean of the likelihood as a function of heading), this optimal weight ratio allows a population to implicitly encode the posterior $P(\theta|\mathbf{r})$, even if tuning is not simply scaled multiplicatively by coherence (see Eq. 6). Thus, the fact that measured weights vary with coherence (**Fig. 7c,d**) in a manner roughly similar to optimal weights (**Fig. 8a**) constitutes indirect evidence for a probabilistic coding scheme such as PPC.

To estimate the terms on the right side of Eq. 7 (i.e., for **Fig. 8a-c**), responses in each single-cue condition were fit separately via linear regression, and the slope and intercept were taken as f' and f, respectively. Because the derivation above assumes linear tuning, these linear fits were used in place of real single-cue data for the decoding simulations depicted in **Fig. 8b,c**; this accounts for the slight difference in single-cue thresholds (and thus optimal weights) in Fig. **8b,c** versus **d,e**.

Nature Neuroscience: doi:10.1038/nn.2983

References

1. Fetsch, C.R., Turner, A.H., DeAngelis, G.C. & Angelaki, D.E. Dynamic reweighting of visual and vestibular cues during self-motion perception. *J Neurosci* **29**, 15601-15612 (2009).
2. Gu, Y., Angelaki, D.E. & Deangelis, G.C. Neural correlates of multisensory cue integration in macaque MSTd. *Nat Neurosci* **11**, 1201-1210 (2008).
3. Jazayeri, M. & Movshon, J.A. Optimal representation of sensory information by neural populations. *Nat Neurosci* **9**, 690-696 (2006).
4. Sanger, T.D. Probability density estimation for the interpretation of neural population codes. *J Neurophysiol* **76**, 2790-2793 (1996).
5. Dayan, P. & Abbott, L.F. *Theoretical Neuroscience* (MIT press, Cambridge, MA, 2001).
6. Foldiak, P. The 'ideal homunculus': statistical inference from neural population responses. in *Computation and Neural Systems* (ed. F.H. Eeckman & J.M. Bower) 55-60 (Kluwer Academic Publishers, Norwell, MA, 1993).
7. Shadlen, M.N., Britten, K.H., Newsome, W.T. & Movshon, J.A. A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J Neurosci* **16**, 1486-1510 (1996).
8. Gu, Y.*, et al.* Perceptual learning reduces interneuronal correlations in macaque visual cortex. *Neuron* **71**, 750-761 (2011).
9. Zohary, E., Shadlen, M.N. & Newsome, W.T. Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* **370**, 140-143 (1994).
10. Averbeck, B.B., Latham, P.E. & Pouget, A. Neural correlations, population coding and computation. *Nat Rev Neurosci* **7**, 358-366 (2006).
11. Morgan, M.L., Deangelis, G.C. & Angelaki, D.E. Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron* **59**, 662-673 (2008).
12. Ohshiro, T., Angelaki, D.E. & DeAngelis, G.C. A normalization model of multi-sensory integration. *Nat Neurosci* **In Press.** (2010).
13. Churchland, A.K.*, et al.* Variance as a signature of neural computations during decision making. *Neuron* **69**, 818-831 (2011).
14. Beck, J., Ma, W.J., Latham, P.E. & Pouget, A. Probabilistic population codes and the exponential family of distributions. *Prog Brain Res* **165**, 509-519 (2007).
15. Beck, J.M.*, et al.* Probabilistic population codes for Bayesian decision making. *Neuron* **60**, 1142-1152 (2008).
16. Ma, W.J., Beck, J.M., Latham, P.E. & Pouget, A. Bayesian inference with probabilistic population codes. *Nat Neurosci* **9**, 1432-1438 (2006).
17. Shadlen, M.N. & Newsome, W.T. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J Neurosci* **18**, 3870-3896 (1998).
18. Carandini, M. Amplification of trial-to-trial response variability by neurons in visual cortex. *PLoS Biol* **2**, E264 (2004).
19. Tolhurst, D.J., Movshon, J.A. & Dean, A.F. The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Res* **23**, 775-785 (1983).
20. Beck, J., Bejjanki, V.R. & Pouget, A. Insights from a simple expression for linear fisher information in a recurrently connected population of spiking neurons. *Neural Comput* **23**, 1484-1502 (2011).

Nature Neuroscience: doi:10.1038/nn.2983