Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons

Christopher R. Fetsch[1], Gregory C. DeAngelis[2] and Dora E. Angelaki[3]**

Abstract | The richness of perceptual experience, as well as its usefulness for guiding behaviour, depends on the synthesis of information across multiple senses. Recent decades have witnessed a surge in our understanding of how the brain combines sensory cues. Much of this research has been guided by one of two distinct approaches: one is driven primarily by neurophysiological observations, and the other is guided by principles of mathematical psychology and psychophysics. Conflicting results and interpretations have contributed to a conceptual gap between psychophysical and physiological accounts of cue integration, but recent studies of visual–vestibular cue integration have narrowed this gap considerably.

Most animals, including humans, function in a complex and dynamic sensory environment in which many events must be detected, interpreted and acted upon. Sensory systems have evolved elegant solutions to cope with this flood of information, but a fundamental and unavoidable aspect of sensory input is its uncertainty: that is, the imperfect mapping between events in the world and the sensory representation thereof. This uncertainty arises from both the physical nature of stimuli (for example, the stochastic arrival of photons at the retina) and the transformation of these physical events into messages carried by noisy devices: namely, neurons[1]. Following in the tradition of nineteenth-century thinkers such as Helmholtz and Fechner, experimental psychologists have long recognized that this inherent uncertainty implies a probabilistic interpretation of sensory function: in the absence of perfect knowledge about the world, the brain must operate with noisy statistical measurements of environmental properties[2,3].

Statistically, a simple way to reduce uncertainty is to combine data from multiple (independent) measurements. Because sensory uncertainty places limits on perceptual performance, it follows that the brain can improve performance by combining sensory measurements, both within and across modalities. This simple fact is both the normative basis and evolutionary advantage of cue integration, which is defined as the combination of multiple sensory cues that arise from the same event or object. In addition to mitigating statistical uncertainty, cue integration can help resolve ambiguities in sensory data. For example, the otolith organs of the inner ear detect linear acceleration of the head, but this can arise either from translational motion of the head (for example, when stepping on the gas pedal in a car) or from tilting the head with respect to gravity. This fundamental ambiguity — a consequence of Einstein's equivalence principle[4] — is resolved by combining otolith signals with information from our rotational motion sensors, the semicircular canals[5,6].

In this Review, we focus on 'multisensory' cue integration, referring to cues that come from different sensory modalities (although many of the same principles apply to within-modality cue integration[7]). Moreover, we will only attempt to shed light on one small corner of this rapidly growing field; for instance, we will not catalogue the many tasks and modalities for which cue integration behaviour has been tested[3,7,8], nor will we survey the intriguing recent findings of cross-modal influences within primary sensory structures[9–14]. Instead, we will build towards answering the following questions: what are (and what should be) the computations performed by individual multisensory neurons and by populations of neurons, and how do these computations give rise to behavioural performance in psychophysical cue integration tasks? These questions have obvious relevance for our general understanding of multisensory processing, but additionally we see them as a path towards unifying two of the main subdisciplines in the field: the theory-driven study of cue integration behaviour in human subjects, and the use of neurophysiology (and more

Normative
A general term referring to an idea, statement or model that describes how something ought to be: that is, relating to an ideal or standard of correctness.

[1]*Department of Neuroscience, Columbia University Medical Center, 630 West 168th Street, New York, New York 10032, USA.*
[2]*Department of Brain and Cognitive Sciences, Center for Visual Science, University of Rochester, 315 Meliora Hall, Rochester, New York 14627, USA.*
[3]*Department of Neuroscience, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA.*
**These authors contributed equally to this work.*
Correspondence to: D.E.A.
e-mail: dangelaki@cns.bcm.edu
doi:10.1038/nrn3503

Cue
Any signal or piece of information bearing on the state of some property of the environment. Examples include binocular disparity in the visual system, interaural time or level differences in audition and proprioceptive signals (for example, from muscle spindles) conveying the position of the arm in space.

recently, neuroimaging) to describe the phenomenology of multisensory interactions within neurons and brain areas. We begin with a brief review of the theory-driven, psychophysical approach.

Psychophysical study of cue integration

Before one can address the neural computations underlying a behaviour of interest, it is helpful to specify a mathematical model of the task that the brain needs to solve. Such models are often designed to be statistically 'optimal' (sometimes called ideal observer models), meaning that the hypothetical observer achieves the best possible performance given the uncertainty of perceptual processing and the constraints of the task. Constructing an ideal observer provides a clear standard against which to test the performance of human or animal subjects and helps to refine our understanding of the relevant computations[7,15].

The use of ideal observer models for studying cue integration has its roots in computational vision research[16–19] but has been extended to include auditory[20,21], somatosensory and proprioceptive[22–25], and vestibular[26–30] modalities. Many of these studies use a simple but highly successful model, which is described in BOX 1. The model states that an optimal observer, when estimating an environmental parameter from multiple sensory cues, performs a weighted average of the estimates derived from the individual cues. Optimality in this case is defined as the estimate that has the lowest possible degree of uncertainty (that is, variance) and also remains unbiased (that is, correct on average). This estimate is achieved by weighting the cues in proportion to their relative reliability (that is, by calculating the inverse variance; see BOX 1)[17,18,31]. With a few simplifying assumptions, this weighting strategy is identical to the one afforded by Bayesian probability theory[3,19,32] (BOX 1), hence the model is often referred to as Bayesian or Bayes-optimal cue integration.

We should emphasize that the predictions of the weighted linear combination scheme (BOX 1) and its extension to Bayesian inference (BOX 1) should be intuitive even without an understanding of the mathematical details. In essence, these models assert that the brain should consider all of the available evidence when making a decision (or estimate), but it should also ensure that more reliable evidence has greater influence. For example, if one were to hear two weather forecasts, one from a reputable meteorologist and the other from an eccentric neighbour, the sensible strategy would be to place more trust in the meteorologist. Of course, this requires knowledge of their reliability; one possible source of such knowledge in the brain during cue integration will be discussed in a later section.

Testing cue integration models with behavioural experiments. Optimal cue integration models can be tested by asking subjects to perform a psychophysical discrimination task using multiple cues simultaneouly as well as each cue in isolation[16,17,20,22,30,33–35]. The reliability of the cues is estimated from the precision with which subjects perform the task under single-cue conditions, establishing the predictions for the optimal weighting scheme (BOX 1). The weights can then be measured and tested against this prediction by placing the cues in conflict and assessing the degree to which each cue dominated the perceptual report (FIG. 1).

Note that although cue-conflict experiments are often used to estimate perceptual weights, this does not imply that cue weighting is only relevant under such artificial experimental conditions. Separate modalities can provide conflicting information under natural conditions: for instance, in tasks that produce a systematic bias in one modality but not in another[36]. More fundamentally, the existence of neuronal noise guarantees conflict between parameter estimates from separate modalities even when stimuli are not actually in conflict[37], just as random draws from independent, identical distributions will rarely be identical.

Despite its simplicity, the basic, optimal cue integration model (BOX 1) has been found to explain psychophysical performance reasonably well across many tasks and systems (reviewed in REF. 7). There are some notable exceptions[38–40], and it is clear that a more complete picture will require schemes that go beyond simple

Box 1 | Optimal cue integration

Studying perceptual processing within a normative framework such as an ideal observer model is useful because it encourages us to think rigorously and quantitatively about the computations required of the nervous system[7,15]. The canonical ideal observer model for optimal cue integration — which is borrowed from a simple statistical method in an unrelated context[31] — predicts a linear weighted sum of single-cue estimates ($\hat{S}_i$), with perceptual weights (w_i) specified by each cue's relative reliability (g_i, which is defined as inverse variance):

$$\hat{S}_{opt} = \sum_{i=1}^{n} w_i \hat{S}_i \qquad (2a)$$

$$w_i = \frac{g_i}{\sum_i^n g_i} \quad g_i = \frac{1}{\sigma_i^2} \qquad (2b)$$

The optimal multisensory estimate ($\hat{S}_{opt}$) has a reliability (g_{opt}) that is greater than the reliability of either single-cue estimate; in fact it is their sum:

$$g_{opt} = \sum_{i=1}^{n} g_i \qquad (3)$$

The same reliability-based weighting scheme can be derived by formalizing the problem in terms of Bayesian inference. In this case the goal is to infer a conditional probability density (known as the posterior) over the parameter of interest (X), given the sensory input from, say, two conditionally independent sources of information C_1 and C_2 (the cues). According to Bayes' theorem:

$$P(X|C_1, C_2) \propto P(C_1|X) \times P(C_2|X) \times P(X) \qquad (4)$$

where $P(C_i|X)$ are called the likelihood functions of each cue (the probability of obtaining the sensory input given each possible value of X) and $P(X)$ is the prior over X (the probability of each particular value of X occurring before any sensory observation is made). If one assumes a uniform prior — meaning that all values of X are considered equally probable before the observation — and independent Gaussian likelihoods, the product of these Gaussians will yield another Gaussian with the mean corresponding to equation 2a and variance corresponding to the inverse of equation 3. But irrespective of any simplifying assumptions, the consequence of multiplying the single-cue likelihood functions (which carry information about cue reliability) is that the greater a cue's reliability, the more it contributes to the final estimate.

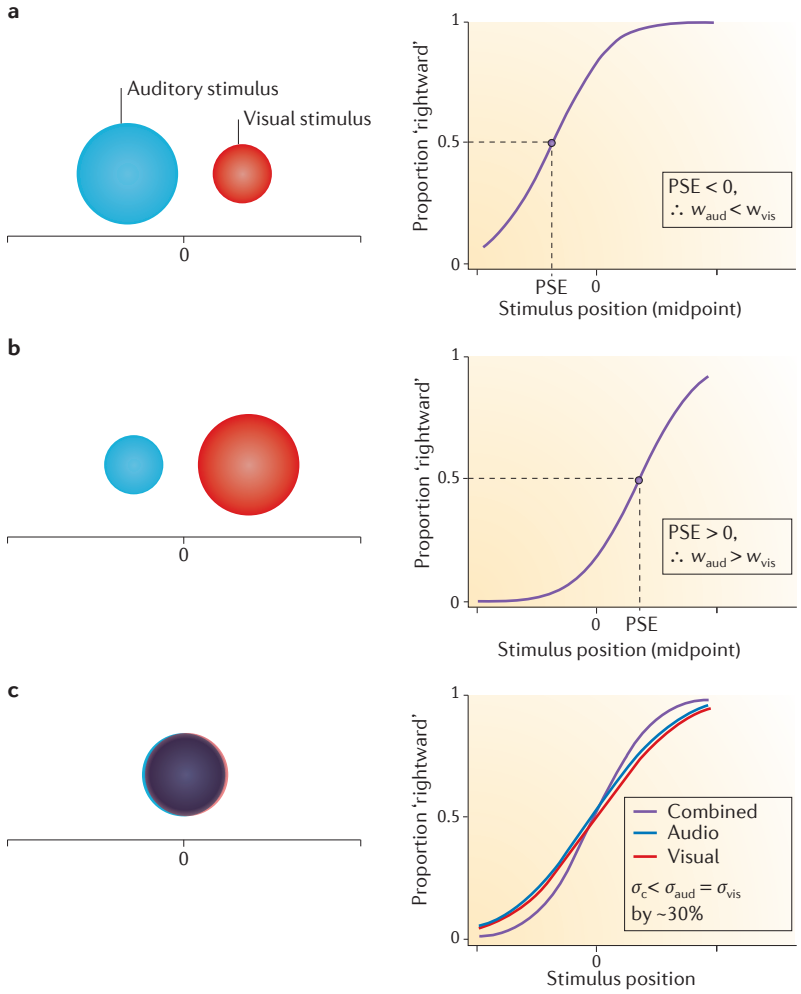


Figure 1 | **Schematic of a generic cue integration cue-conflict psychophysical task.** A simplified version of a visual and auditory localization task[20,41,118] in which the subject reports whether a stimulus was located to the left or right of a reference location (marked '0'). The stimulus can be a visual cue (represented by the red circle) and/or an auditory cue (represented by the blue circle) presented at one of several possible locations in front of the subject. The two cues are presented either at the same location or separated by some amount of distance the cue-conflict), and the reliability of one or both cues is often manipulated experimentally. Here, it is denoted by the width of the circles (where wider is less reliable). **a** | A depiction of cue-conflict trials in which the visual cue is more reliable and also displaced to the right, whereas the auditory cue is less reliable and displaced to the left. For this example, the cue-conflict is kept fixed, and the pair of stimuli are jointly moved to the left or right on different trials, generating a sigmoidal choice curve known as the psychometric function (purple line, shown on the graph on the right), which is plotted relative to the midpoint between the two stimuli. If subjects weight the cues according to their reliability, they will make more rightward choices for a given position of the paired stimuli (relative to non-conflict conditions), and the psychometric curve will be shifted to the left of centre. The stimulus position at which the curve reaches 50% rightward choices (that is, the point of subjective equality (PSE), which is indicated by dashed lines) maps onto a particular set of perceptual weights (w_{aud} and w_{vis}; the w_i in equation 2 in BOX 1), which in this case would have the relationship $w_{aud} < w_{vis}$, as the visual cue is more reliable. **b** | A scenario from a different set of trials with the same cue-conflict but reversed reliability (the auditory cue is more reliable than the visual cue). Here, the subject should make more leftward choices, shifting the curve to the right ($w_{aud} > w_{vis}$). **c** | In addition to measuring shifts of the psychometric function, performance with combined visual and auditory stimuli (purple curve) can be compared with single-cue conditions (red and blue curves), testing the prediction of increased reliability (see BOX 1; here denoted by a decrease in the SD (σ) of the purple cumulative Gaussian psychometric function by a factor of the square root of 2).

reliability-weighted linear combination. For example, recent work has modified the basic model to include inference about whether two cues arise from the same real-world event ('causal' inference[41]), the ability to discount highly discrepant information (robust estimation[17,42]) and consideration of the accuracy of cues in addition to their precision[43,44]. Nevertheless, it is fairly well established that human subjects are able to reduce perceptual uncertainty and improve their performance by combining multiple cues in a manner that approximates a statistically optimal observer (BOX 1). This ability has recently been shown in non-human primates[30] and rodents[45] too. As real-world objects and events are rarely sensed with just a single modality, these studies suggest that weighting sensory information by its reliability may be fundamental to our everyday experience of the world and the actions we take in it.

The neurophysiology of multisensory integration

Years before the aforementioned theoretical and psychophysical tools were brought to bear on cue integration, several laboratories had begun to characterize the properties of multisensory neurons in experimental animals. Although neuroscience has historically considered each sensory modality as a distinct information channel with its own dedicated brain structures, it was known fairly early on that neurons with converging sensory inputs could be found in many regions throughout the brain and across species (reviewed in REFS 46, 47).

A region that has received the bulk of neurophysiologists' attention is the mammalian superior colliculus (SC), a midbrain structure primarily involved in orienting the eyes and head towards salient stimuli[48–50]. Because the motor system needs to react to stimuli regardless of the modality with which they are detected, it makes sense that such a structure would contain multisensory neurons. In particular, cells in the deep layers of the SC are spatially selective for visual and auditory targets as well as tactile stimulation of the face and body. However, beyond mere convergence of multimodal information, it soon became clear that these signals interact functionally within the SC, often in dramatic fashion[47]. For example, weak visual or auditory stimuli presented alone might elicit a very small response from an SC neuron, but when presented together, it can cause a vigorous neural response.

The seminal work of Stein and colleagues (for reviews, see REFS 46,51) characterized this and other functional interactions in the SC, outlining a number of empirical principles that have guided multisensory research for over two decades, including a recent surge in human functional MRI studies[11,52,53]. The extensive evidence supporting these empirical principles has been reviewed elsewhere[14,46,51,54]. For our purposes, the key findings can be adequately summarized by the spatial and temporal principle and the principle of inverse effectiveness[55]. In the spatial and temporal principle, neurons that receive input from multiple sensory modalities typically show enhanced responses to multisensory stimuli (relative to the largest unisensory response) provided that the stimuli from the two modalities are close together in space and

time. By contrast, sufficient separation in space or time can suppress the multisensory response relative to the best unisensory response. In inverse effectiveness, multisensory response enhancement is proportionally larger when the same stimuli presented individually (unisensory stimuli) only weakly activate the neuron. Moreover, weak unisensory stimuli can cause a substantial fraction of multisensory neurons in the SC to display 'superadditivity', a phenomenon in which the multisensory response of a particular neuron is greater than the arithmetic sum of unisensory responses. Note, however, that inverse effectiveness can hold regardless of whether superadditivity occurs[55].

Importantly, these physiological response properties bear an intriguing resemblance to behaviour in alert animals. For example, stimulus combinations that were more effective in activating neurons were also more effective at driving the behavioural detection of stimuli[56–59]. In addition, this line of research has assembled an impressively thorough account of the anatomical origins of multisensory integration in the SC (namely, via descending input from specific cortical regions[60,61]) as well as its developmental trajectory[62–64]. However, until recently, less attention had been paid to the computations that underlie multisensory integration in SC neurons.

A significant advance on this front came in a report by Stanford and colleagues[55]. These authors noted that most of the earlier studies tested only a limited set of stimulus intensities, focusing on weaker stimuli for which multisensory responses were most impressive and seemingly most relevant for behaviour. Without a more thorough characterization of SC responses to different combinations of stimulus intensities, the mathematical operations performed by SC neurons remained obscure, hampering efforts to understand and model the underlying neuronal mechanisms[55]. Expanding the repertoire to three levels of intensity, which were chosen to span the dynamic range of each neuron's response to both visual and auditory stimuli, Stanford *et al.*[55] confirmed the basic response patterns outlined above and systematically characterized transitions from superadditive to additive to subadditive responses in single neurons with increases in stimulus strength, which is consistent with the principle of inverse effectiveness (also see REFS 65–67). Whereas nearly all neurons demonstrated inverse effectiveness, superadditivity was generally seen only in neurons with weak responses to unisensory stimuli. These studies helped to clarify that superadditivity is not a ubiquitous property of multisensory integration by neurons. However, they still left the specific mathematical rules by which SC neurons combine their inputs in doubt.

Computational modelling and neural theories of cue integration. In the same spirit of attempting to determine the computations underlying multisensory integration, several computational models[68–70] have been presented to account for SC responses and their dependence on cortical input (reviewed in REF. 71). Recent SC-based network models provide a detailed explanation of multisensory responses in the SC, emphasizing the role of the association cortex and laying a strong predictive foundation for future studies. However, despite this progress, a parsimonious description of the basic neural computations involved in multisensory integration has eluded consensus. More importantly, these models focused mainly on explaining the physiology rather than providing quantitative predictions linking neuronal activity with behaviour[68–71]. One issue is that the behavioural task associated with multisensory integration in the SC (orienting to spatial targets) has not typically been defined in the language of statistical decision-making — or in the related ideas of signal-detection theory[72] — as used by psychophysicists (but see REFS 73,74), and hence the neurophysiological data were not interpreted in that context.

Meanwhile (and by contrast), a different theoretical framework[75] was developed from the perspective of statistical inference — the process of drawing conclusions based on uncertain evidence — of which cue integration is a special case (see previous section and BOX 1). In a landmark set of studies, Ma *et al.*[75] showed how populations of neurons can implicitly represent probability distributions and perform Bayesian computations[75] and optimal decision-making[76]. The key insight of this 'probabilistic population coding' (PPC) framework (FIG. 2) is that the type of variability observed in most neuronal responses — termed Poisson-like variability[75] — makes these computations surprisingly easy to perform. Specifically, the Bayes-optimal combination of multiple sensory cues (an operation that mathematically requires multiplication of probability distributions; see BOX 1) can be achieved by a simple linear summation of population activity (FIG. 2a; also see REF. 77 for a related approach).

To make this idea more concrete, consider a cue-conflict task such as the one depicted in FIG. 1a,b. Visual and auditory cues are presented with a small spatial separation (conflict) and unequal reliability. When the activity of each neuron is plotted as a function of its preferred stimulus, the population response in unisensory areas takes the form of noisy 'hills' of activity (simulations of which are shown in FIG. 2b by the blue and red points in the graph in the top panel). The gain (also called the height) of each hill of activity is proportional to the reliability of the corresponding cue, with a proportionality constant that depends on the width of the tuning curves and the number of neurons. The summation of unisensory activity (indicated by the purple points in the top graph of FIG. 2b) produces a third hill that is shifted towards the more reliable cue, reflecting the optimal weighting of the cues (BOX 1). There is no need for cue reliability to be learned or even represented explicitly in the brain; instead, it is encoded by the unisensory population activity itself (in the form of probability distributions over stimuli; see the bottom panels of FIG. 2a,b) and then propagated to a downstream multisensory area via linear summation[75].

Thus, the PPC theory predicts that the brain regions involved in optimal cue integration should exhibit additive[75] or subadditive[78] responses regardless of stimulus strength, which is in contrast to the emphasis on superadditivity and inverse effectiveness (a dependence on stimulus strength) in the SC literature. Compared with models specifically targeting the SC[71], the key advantages of the PPC framework are its generality — it can be

Ideal observer
A theoretical construct used to quantify optimal performance in a given task, where optimality is defined according to a pre-defined mathematical function (for example, minimizing a cost function or maximizing a utility function). The term 'ideal' does not imply perfect (error-free) performance, which is generally impossible given the uncertainty associated with all sensory data.

Reliability
Although the term reliability can mean different things in different fields, here we use it as a synonym for the precision of a measurement, which is defined mathematically as its inverse variance.

Bayesian probability theory
The branch of statistics and probability theory in which probability is interpreted as the 'degree of belief' that an event will occur (or that a hypothesis is true) rather than the relative frequency with which it has occurred. It is chiefly associated with the process of updating a prior belief about a hypothesis in light of new data, but the essence of Bayesian theory is this way of thinking about probability itself, which permits the estimation of a statistical parameter (or property of the environment) from experimental observations (or sensory information).

Accuracy
Accuracy refers to how close the measurement is to the true value: that is, how unbiased it is.

Poisson-like variability
Neurons respond differently to repeated presentations of the same stimulus, and this variability often resembles a family of probability distributions that includes the Poisson distribution (hence the term 'Poisson-like'). A prominent feature of this family is that the variance of neuronal responses (that is, the variance of the number of action potentials across repeated stimulus presentations) is proportional to the mean response.

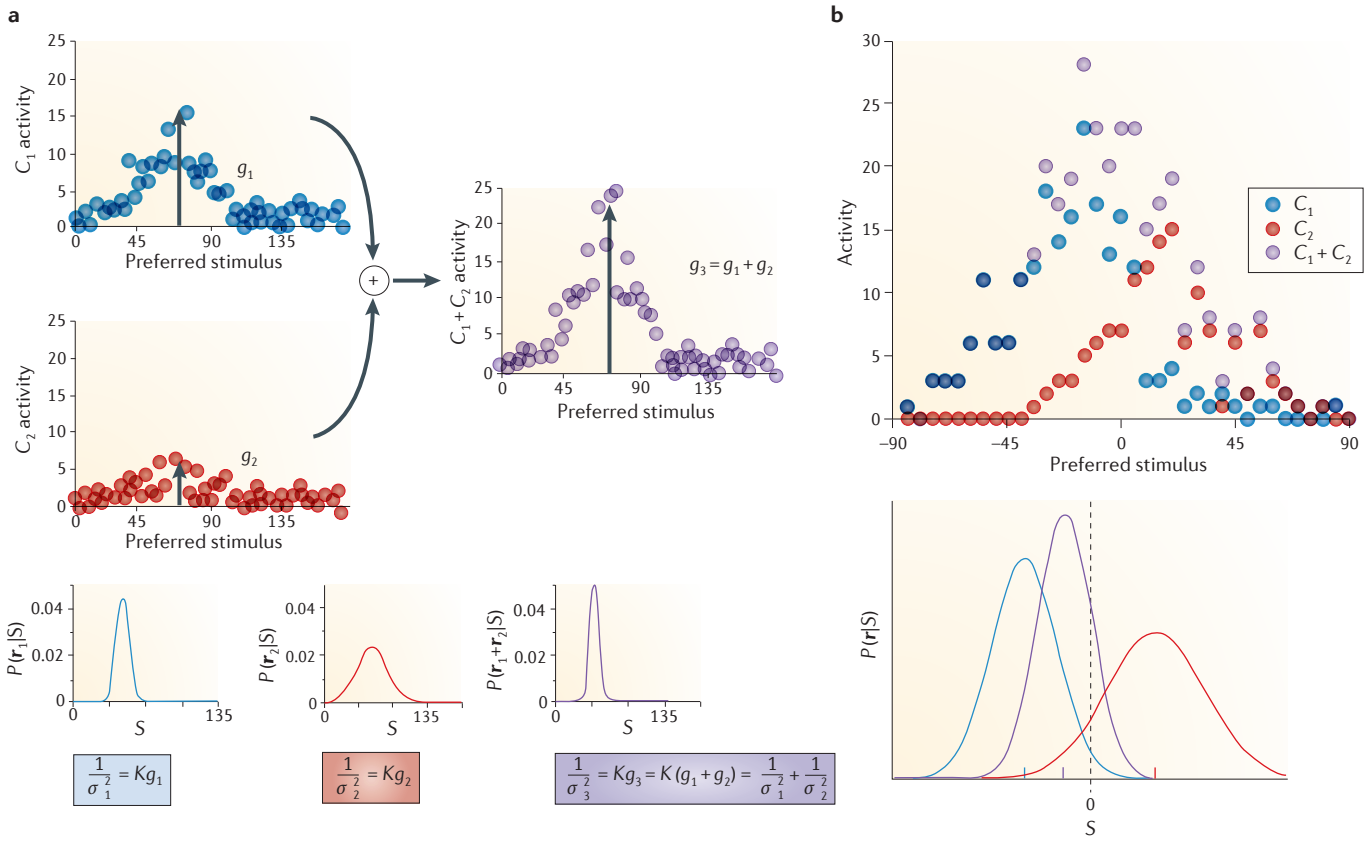


Figure 2 | **A probabilistic population code framework accounts for optimal cue integration by summation of unisensory population activity.** **a** | In this model, sensory cues C_1 (blue) and C_2 (red) each generate a 'hill' of population activity in their respective unisensory areas, which could be, for example, regions in the visual cortex and auditory cortex. Each data point indicates a single neuron, and these cells are arranged by their preferred stimulus value (for example, receptive field location). The hills are noisy and not smooth because of variability in neuronal responses. Owing to the particular kind of variability in these model neurons (which is also commonly found in real neurons), each hill of activity encodes a conditional probability distribution (the probability of response pattern $\mathbf{r}$ given stimulus value S ($P(\mathbf{r}|S)$; see the bottom panel) whose variance is inversely proportional to the gain (height) of the hill ($Kg = 1/\sigma^2$, where K is a constant). The gain of the hill is indicated by the vertical arrows (note the weaker response and consequently broader distribution for the less reliable cue, C_2). The inverse variance of this distribution is the quantity needed to perform optimal reliability-weighted cue integration (BOX 1). Summing the two unisensory populations neuron by neuron generates a third population (purple) whose gain is the sum of the unisensory gains g_1 and g_2. Therefore, the inverse variance of the probability distribution $P(\mathbf{r}_1 + \mathbf{r}_2)|S)$ encoded by the multisensory population is equal to the sum of the individual cues' inverse variances (that is, their reliabilities) — the same operation prescribed by the optimal integration model (BOX 1). **b** | A simulated cue-conflict trial in which sensory cue C_1 specifies, in arbitrary units, a stimulus value of –20 and C_2 specifies a stimulus value of +20. The C_1 response has a greater gain than the C_2 response, simulating a more reliable cue being presented along with a less reliable one, respectively. After summation, the resulting hill of activity (purple) is skewed towards the more reliable cue, as shown schematically by the encoded probability distributions (bottom graph). A downstream brain area that optimally decodes this multisensory activity would produce behavioural responses consistent with optimal cue integration theory (BOX 1). Note that the shape of the multisensory hill — which depends on parameters such as the shape and width of tuning curves and the size of the cue-conflict — need not mimic the shape of the encoded distributions. Optimal cue integration can still occur via a linear combination of unisensory activity for various tuning widths or shapes, provided that the linear combination is appropriately tailored to these tuning properties[75]. Part **a** is reproduced, with permission, from REF. 75 © (2006) Macmillan Publishers Ltd. All rights reserved.

applied to neurons in multisensory brain regions other than the SC — and its ability to account for a range of psychophysical results using straightforward and biologically plausible linear operations. One caveat to the PPC model, however, is that the brain must 'know' (or be able to learn) the shape, width and distribution of neuronal tuning curves in order to perform the correct linear operations and to decode the probability distributions encoded in population activity.

The story so far. In summary, the gulf between the two main approaches in multisensory research — the empirical and neurophysiology-driven approach versus the psychophysical and theory-driven approach — can be boiled down to a few key historical facts. First, despite a wealth of data (summarized by the ubiquitous empirical principles[46]) and several detailed models of the SC[71], there has been no consensus on the mathematical rules by which single neurons combine multisensory inputs, and few (if

any) model-based behavioural predictions that would permit coupling the SC literature to the modern psychophysical paradigm of statistical decision-making. Second, traditional neurophysiology studies did not measure behaviour in a psychophysical task while recording from multisensory neurons and attempting to link the two. Last, an influential theory[75] designed to capture the probabilistic nature of cue integration made a clear prediction (multisensory interactions should be additive or subadditive and independent of stimulus strength) that was at odds with reported non-linear interactions in the SC, such as superadditivity and inverse effectiveness (although the emphasis on superadditivity has waned as estimates of its prevalence evolved over time[55,67,79–81]).

It should be noted that distinct approaches operating in parallel can be productive for scientific discovery, and it goes without saying that many important insights have been gained in the two subdisciplines that do not require any sort of reconciliation or unification of ideas. Nevertheless, in the following section, we will describe a series of studies that bridges some of the gaps described above by using an ecologically relevant task that is both inherently multisensory and amenable to simultaneous psychophysical and neurophysiological investigation. This research focuses on a different brain area and task than those in the SC studies but, as explained below, reveals a putative computational mechanism underlying key aspects of multisensory integration in both areas.

Heading perception from visual–vestibular cues

All animals that navigate through their environment need to estimate their direction of self-motion, which is referred to as heading. Heading perception is an intriguing example of a fundamentally multisensory task. When we move, vision provides cues such as optic flow[82,83] — the apparent movement of the visual scene caused by relative movement between an observer and their surroundings — while several non-visual modalities signal the physical motion of the head or body. In particular, the otolith organs of the vestibular system act as inertial sensors, providing a directional self-motion cue during translation (as opposed to rotation) of the head in space[84–87].

Several areas in the primate brain receive both vestibular and visual signals related to self-motion[88–90]. In the context of cue integration, the best studied of these areas is the dorsal medial superior temporal area (MSTd), which is located in the extrastriate visual cortex of the macaque monkey. Neurons in this region are selective for heading based on optic flow and/or vestibular cues[89,91], and artificially manipulating MSTd activity via electrical stimulation or reversible inactivation can affect perceptual decisions in heading discrimination tasks[92,93].

A comprehensive strategy for determining how neurons combine sensory information. With the goal of understanding the computations performed by multisensory neurons, Morgan *et al.*[94] recorded from individual MSTd cells while monkeys were presented with naturalistic heading stimuli delivered using a virtual-reality system (FIG. 3a). The stimuli consisted of either physical movement of the body by a motion platform (the 'vestibular' condition), computer-generated optic flow simulating observer movement through a three-dimensional field of random dots (the visual condition) or synchronous combinations of the two cues (the combined condition).

Each neuron was tested with many different combinations of visual and vestibular headings at different levels of stimulus strength (that is, cue reliability). This is similar to the approach taken by Stanford *et al.*[55] in the SC but with a few notable differences. Stanford *et al.* presented both visual and auditory targets together in the centre of the neurons' receptive fields (thereby focusing on multisensory enhancement) and varied stimulus strength over a modest range, resulting in an approximately twofold difference in the firing rate on average. By contrast, Morgan *et al.*[94] presented all combinations of congruent and conflicting stimuli across the full 360-degree range of possible headings (that is, spanning the complete tuning curve of each neuron, including both preferred and non-preferred stimuli) and varied visual cue reliability (by varying the motion coherence of the optic flow display) over a fourfold range.

The aim of the Morgan *et al.* study[94] was to define a mathematical 'combination rule' that could adequately describe multisensory responses in area MSTd. In this case, combination rule refers to an arithmetic expression that describes the response to multisensory stimulus combinations (R_{comb}) as a function of the responses to vestibular and visual stimuli (R_{vestib} and R_{visual}, respectively) presented separately. To accurately measure the combination rule, a broad range of preferred and non-preferred stimuli need to be tested, probing as much of the stimulus space and dynamic range of the neurons as possible. This comprehensive strategy is important because the contribution of non-linearities such as a response threshold or saturation (a ceiling effect on high firing rates) can vary widely across different stimulus regimes, potentially biasing the outcome towards super- or subadditivity[94,95]. Perhaps the most crucial aspect of this strategy was the presentation of conflicting stimuli (that is, different headings specified by visual and vestibular cues), as the neural combination rule is not well constrained otherwise. If a neuron has similar tuning for the two modalities, a congruent multimodal stimulus can elicit a response that is consistent with various combination rules. In other words, the same response could arise from a large weight applied to one cue and a small weight to the other, or vice versa. By analogy, the perceptual weights in a multisensory psychophysical task cannot be measured unless the cues are somehow placed in conflict (FIG. 1); otherwise, the subject could completely ignore one or the other cue and still give the same behavioural response.

Although, in principle, the neural combination rule could be fairly complex (for example, involving non-linearities), Morgan *et al.*[94] found that a simple weighted sum plus a constant ($R_{comb} = A_{vestib} \times R_{vestib} + A_{visual} \times R_{visual} + C$) provided a good fit to the data, with 'neural weights' A_{vestib} and A_{visual} that were subadditive (less than 1) on average. This finding of linear subadditivity in area MSTd is compatible with the PPC theory of Bayes-optimal cue integration[75]. However, a key feature of the MSTd combination rule was not predicted by the basic PPC theory or any

Heading
An organism's instantaneous direction of translational movement.

Motion coherence
A property of random-dot motion stimuli — used in visual psychophysics and neurophysiology — that is often varied to control stimulus strength and therefore task difficulty. Motion coherence is the percentage of dots moving in the prescribed direction (the 'signal') while the remaining dots are re-plotted randomly on every video frame (the 'noise').

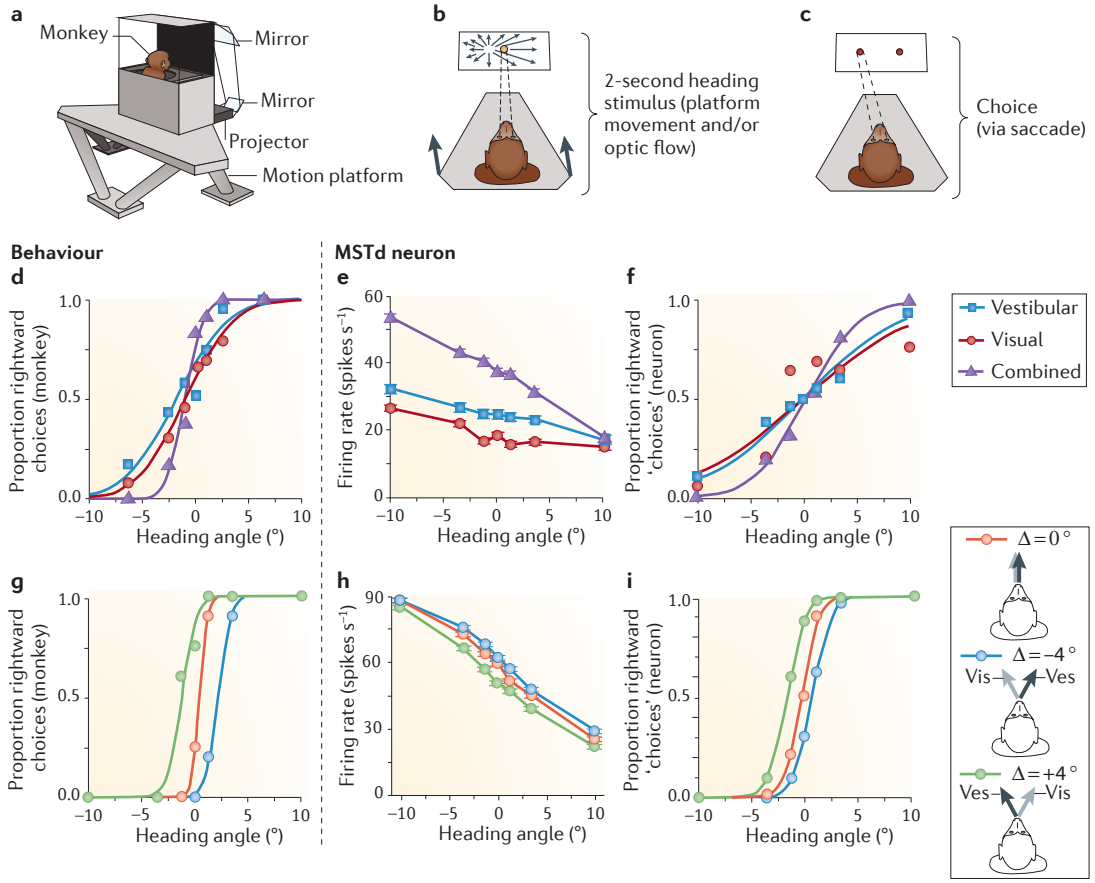


Figure 3 | **Combined psychophysical and neurophysiological studies of visual–vestibular cue integration in the macaque.** **a** | Monkeys were trained to report their perceived heading (direction of self-motion relative to straight ahead) while seated in a virtual-reality setup[26]. The apparatus consists of a motion platform that can translate in any direction, upon which is mounted a projector and rear-projection screen for displaying optic flow patterns that simulate movement of the observer through a random-dot 'cloud'. **b** | While the monkey fixated his gaze (indicated by the dashed lines) on a spot at the centre of the screen (orange), heading stimuli were delivered in one of three conditions: vestibular (platform motion only, indicated by arrows on the platform), visual (optic flow only, indicated by arrows on the screen) or combined (platform motion and optic flow). **c** | Following each 2-second motion stimulus (here, a heading to the left of straight ahead), the monkey indicated his choice by making a saccadic eye movement to one of two targets (red). **d** | Behavioural data (psychometric functions) for a single session are plotted showing the proportion of rightward choices as a function of signed heading angle, where positive heading indicates rightward motion and negative heading indicates leftward motion. The slope of the fitted curve is a measure of the animal's sensitivity to small changes in heading, which is a proxy for the reliability of the cue (or cues). The slope was greater in the combined condition (purple curve) than in the single-cue conditions (blue and red curves), indicating an improvement in sensitivity (that is, a reduction in uncertainty or variance). The average improvement across sessions was close to the optimal prediction (BOX 1). **e** | The firing rate responses (tuning curves) of a single example neuron from the dorsal medial superior temporal area (MSTd) are plotted using the same conventions as the behavioural data. Note the steeper slope of the tuning curve in the combined condition (purple curve), suggesting an increase in sensitivity of the neuron under multisensory stimulation. **f** | The firing rates depicted in panel **e** were converted into simulated choices by an ideal observer using receiver operating characteristic (ROC) analysis. The resulting 'neurometric' functions quantify the sensitivity of the neuron to small changes in heading during the vestibular (blue), visual (red) and combined (purple) conditions. Similar to the behavioural effect, the slope of the neurometric curve is steeper in the combined condition than in the single-cue conditions. **g** | In a separate study[100], the cues were placed in conflict to test for reliability-based cue weighting, analogous to FIG. 1a,b. Here, the visual cue was more reliable, hence the monkey made more rightward choices when the visual heading was displaced to the right (conflict angle (Δ) = +4 °, green curve) and more leftward choices when the visual heading was displaced to the left (Δ = −4 °, blue curve) compared with straight ahead (Δ = 0 °, orange curve). **h** | Tuning curves from the same neuron as in panel **e**, which were recorded under cue-conflict conditions. The curves are offset from one another because the more reliable visual cue drives the cell to fire more spikes (Δ = −4 °, blue curve) or fewer spikes (Δ = +4 °, green curve) for a given heading angle. **i** | Conversion of these firing rates into neurometric functions reveals a pattern similar to the behavioural result in panel **g**; the shift of the curves for different values of Δ reflects the trial-by-trial weighting of cues (favouring the more reliable visual cue, as predicted from optimal cue integration). Panels **a**, **e** and **f** are modified, with permission, from REF. 119 © (2010) John Wiley and Sons. Panels **g**–**i** are modified, with permission, from REF. 100 © (2011) Macmillan Publishers Ltd. All rights reserved.

other existing model at the time: the neural weights were found to vary with cue reliability. Specifically, the visual weight (A_{visual}) increased and the vestibular weight (A_{vestib}) decreased with increasing visual motion coherence[94].

At first, this outcome seems perfectly reasonable: behavioural choices show evidence for the weighting of cues according to their reliability (BOX 1; FIG. 1a,b), and neurons apparently showed a similar pattern. However, it is important to clarify the distinctions between different uses of the term 'weight': for example, the weights computed at the level of single neurons versus weights computed at the level of behaviour. BOX 2 explains these distinctions with a simple conceptual model. The take-home message is that the neural weights (combination rule) measured in studies similar to that of Morgan *et al.*[94] do not map onto perceptual weights (BOX 1; FIG. 1a,b) in any straightforward way — this caveat also applies to the empirical principles of Stein and colleagues[46]. Rather, their relationship depends on many factors, including the tuning properties of the neurons and the read-out mechanism.

With this caveat in mind, let us return to the question of whether neural weights should vary with cue reliability. The model of Ma *et al.*[75] asserts that these weights need not vary with reliability for the population to account for reliability-based weighting in behaviour — in fact, the optimal neural weights are fixed at 1 (simple summation) under a reasonable set of assumptions.

How can cue reliability shape the output of the multisensory population without affecting single-neuron weights? The simple answer is that cue reliability (as manipulated by stimulus strength), by definition, alters the strength of unisensory inputs going into the summation. The Poisson-like neural variability, which entails a response variance proportional to its mean, ensures that the population hill of activity encodes a probability distribution with inverse variance (that is, reliability) proportional to the amplitude (or gain) of the hill (FIG. 2a). Adding two such hills yields a third (multisensory) hill with gain $g_3 = g_1 + g_2$ and thus encodes a probability distribution reflecting the sum of unisensory cue reliabilities — precisely the prediction of optimal weighting schemes (BOX 1). If the summation of activity needs to be counteracted by a baseline shift (via global inhibition) to keep neurons in their dynamic range, then single neurons will appear subadditive without compromising the optimality of the computations[78]. However, this still does not predict neural weights that vary with reliability.

Because reliability-weighted cue integration is readily achievable in a population with fixed neural weights[75], the changes in neural weights as a function of reliability described by Morgan *et al.*[94] remained a puzzle. With hindsight, this neural combination rule was actually a clue, hinting at a particular network-level computation that would unify several seemingly unrelated findings. Before relating that part of the story, however, we must first establish that these computations are relevant for a behaving animal actively engaged in cue integration. Thus, the next section will summarize evidence linking MSTd activity to psychophysical performance in a multisensory heading discrimination task.

Comparison of MSTd responses with cue integration behaviour. We and others[55] have argued that deciphering the computations reflected in single-neuron activity is crucial for constraining models of multisensory integration. However, it is equally important to seek evidence supporting the involvement of such neurons in a particular behaviour, and a potent way to collect such evidence is to record or manipulate neural activity during a psychophysical task[96]. Until recently, this had not been done for cue integration tasks, as there were few if any suitable animal models for this purpose.

We developed a paradigm enabling the simultaneous recording of MSTd neuron activity while monkeys performed a multisensory heading discrimination task[26]. As in the Morgan *et al.* study[94], a virtual-reality apparatus (FIG. 3a) was used to deliver stimuli in visual-only, vestibular-only and combined conditions, and cue reliability was controlled by varying the visual motion coherence. Monkeys were trained to report their perceived heading (left or right relative to straight ahead; left in the example of FIG. 3b) by making an eye movement to one out of two choice targets presented at the end of each trial (FIG. 3c).

Behaviourally, Gu *et al.*[26] found that monkeys show near-optimal (BOX 1) improvements in perceptual sensitivity when the cues were presented together compared with when they were presented alone (FIG. 3d). Subjects also weighted the cues in proportion to their reliability[30] (FIG. 3g), as predicted by the standard optimal cue integration model (BOX 1). We then related neural activity to behaviour by converting firing rates into simulated choices made by an ideal observer, using a common technique called receiver operating characteristic (ROC) analysis[97–99]. In this analysis, the simulated observer effectively performs the discrimination task by comparing distributions of firing rates (the means of which are plotted in FIG. 3e,h) recorded in response to different stimuli. The resulting 'neurometric' curves (FIG. 3f,i) quantify the sensitivity and pattern of cue weighting by a single neuron in units that are comparable with the behavioural data. The results indicated that MSTd neurons show close parallels with behaviour with respect to both the improvement in sensitivity[26] (compare FIG. 3d with FIG. 3f) and cue weighting[100] (compare FIG. 3g with FIG. 3i).

What should neurons do? Deriving the optimal combination rule. Although we showed that subjects weight visual and vestibular cues in proportion to their reliability[30], the perceptual weights were in fact slightly but consistently suboptimal: both humans and monkeys modestly over-weighted vestibular information in this task, when compared with the optimal predictions from equation 2 in BOX 1. Remarkably, this specific deviation from optimality was reflected in MSTd neuronal activity[100]. To explain this surprising finding, we needed to return to the concept of a neural combination rule[94] and compare the observed neural weights with those that would be required to produce optimal cue integration at the level of behaviour. As suggested in BOX 2, this is not as simple as mapping neural firing rates onto the choices of a simulated observer (that is, the ROC

Box 2 | **Levels of analysis in multisensory integration**

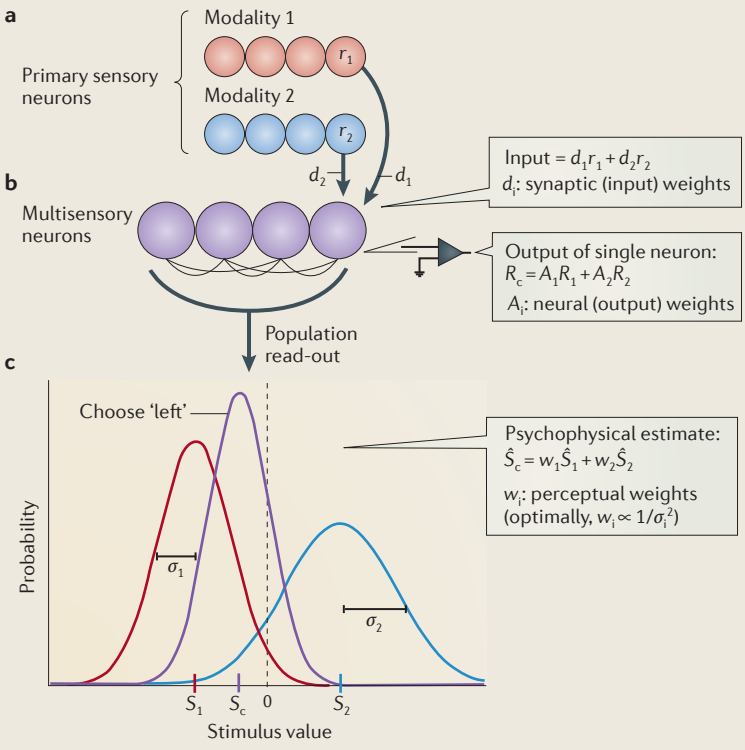
Descriptions of how single neurons combine multiple sensory inputs, such as those provided by Stein and colleagues for the superior colliculus (SC)[46,51,55] and Morgan *et al.*[94] for dorsal medial superior temporal area (MSTd), are often compared with patterns of behavioural cue integration despite the complex and poorly understood connection between these two levels of analysis. Here, we attempt to clarify some of the issues involved by illustrating three distinct uses of the term 'weights'.

In this highly simplified scheme, we assume the existence of two populations of primary sensory neurons (shown in red and blue), with each receiving unisensory information from different modalities (for example, the visual (red) and auditory (blue) stimuli in FIG. 1). These signals are transmitted to a multisensory neural population (purple), the activity of which will generate a particular behaviour or perceptual choice.

In the figure, the output of two primary sensory neurons (r_1 and r_2 in panel **a**) converges onto a multisensory neuron with synaptic (or 'input') weights d_1 and d_2. These weights could reflect the number and/or efficacy of synaptic connections associated with each modality and are generally inaccessible to the neurophysiologist recording extracellular action potentials in a multisensory area (although they often can be inferred from the relative strength of unisensory responses[105,117]). Instead, what we actually measure is the output of a network computation — probably involving lateral and perhaps feedback connections — which generates firing rates denoted R_1 and R_2 (the responses to each modality presented in isolation) and R_c (the response to multisensory stimuli; see panel **b**). We can then ask how R_c is best predicted from R_1 and R_2, for instance, via a weighted sum with neural (or 'output') weights A_1 and A_2[94,100].

Lastly, the population activity of the multisensory layer is read out (decoded) by downstream circuits to generate a behavioural response (panel **c**). The details of this step are not critical; here, similar to the model shown in FIG. 2, the population activity is shown giving rise to bell-shaped probability distributions (likelihoods or posteriors; see BOX 1): blue, red and purple corresponding to trials of modality 1-only, modality 2-only and combined multisensory stimuli, respectively. It is assumed that the brain uses these distributions to choose a single stimulus value (usually the peak value or the most likely value) as its estimate on a particular trial, leading to a behavioural choice made by the subject (in this case, left relative to a reference value of zero). Regardless of the neural implementation, the perceptual weights w_1 and w_2 can be estimated by recording these choices over many trials (that is, the sigmoidal choice curves in FIG. 1). The optimal perceptual weights are given by the reliability (that is, inverse variance) of the unisensory evidence (BOX 1), as illustrated here by the shift of the combined (purple) distribution towards the more reliable cue (modality 1, red; note that SD $\sigma_1 < \sigma_2$ and therefore $w_1 > w_2$ and the observer more often chooses 'left').

With or without the equations, we hope this exercise makes it clear that the neural weights depicted in panel **b** (that is, what is measured in most single-unit studies of multisensory integration[51,94]) are decoupled from both the synaptic weights onto a multisensory neuron (panel **a**) and the perceptual weights (panel **c**) that are commonly measured in cue integration psychophysics[7]. In particular, the relationship between neural weights (A_1 and A_2) and perceptual weights (w_1 and w_2) is far from straightforward and is the subject of ongoing investigations (see the main text).



analysis strategy discussed above) because such mapping is not bidirectional: one cannot take a hypothetical instance of optimal performance and directly infer its underlying neuronal activity. Instead, we needed to derive a more basic mathematical formula for how multisensory neurons should — in a normative sense — combine their inputs.

Our approach was to assume that the goal of neural cue combination (BOX 2) is to maximize the information carried by neurons about the variable of interest (here, this is heading direction). Using Fisher information[101,102], a quantity related to the precision with which an ideal observer can discriminate between small changes in a stimulus, we derived a simple expression for the optimal

neural weights. In summary (for details, see REF. 100), we found that the optimal neural weights do in fact vary with cue reliability in a manner similar to the pattern observed by Morgan *et al.*[94]: as the reliability of the visual cue increases, so does its influence on the multisensory neural response. This differs from the prediction of the original PPC study[75] — in which neural weights were fixed at 1 irrespective of cue reliability — but the discrepancy can be explained by an assumption in the basic PPC framework about the effect of cue reliability on neuronal responses, an assumption that does not hold for our MSTd data. Thus, the derivation of an optimal combination rule for neurons[100] was essentially a modification of the PPC theory to accommodate MSTd-like response patterns. The bottom line is that, by making a mathematical appeal to optimality, we were able to understand an initially puzzling empirical finding[94] in a completely new light: that is, as part of a neural strategy for combining sensory signals in a near-optimal fashion within a PPC framework.

Closing the loop: a normalization model unifies old and new observations. Although we had gained a better understanding of the neuronal combination rule in MSTd and why it might exist, we still had not addressed the 'how' question: what cellular and/or circuit mechanisms could plausibly explain the changes in how neurons weight their inputs as a function of cue reliability? We surmised that the changes in neural weights (BOX 2) were unlikely to reflect changes in synaptic weights onto individual cells (BOX 2) for two reasons. First, because the animals in Morgan *et al.*[94] were not performing a perceptual task, there was no clear basis for expecting reliability to modulate synaptic strength (for example, via some kind of reward signal), even if such changes were desirable. Second, Fetsch *et al.*[100] showed that the neural weight changes occur even when reliability is varied randomly on a trial-by-trial basis. These effects seem unlikely to be mediated by changes in synaptic weights because the strengths of synaptic inputs to the neuron would need to be modulated from moment to moment (on a timescale of tens or hundreds of milliseconds) based on a rapid assessment of cue reliability at the beginning of each trial. Although such a possibility cannot be ruled out, it would require sophisticated mechanisms for modulating synaptic weights; such mechanisms are currently unknown.

For these reasons, it seemed more likely that a network-level computation was responsible for the observed changes in neural weights. By this we simply mean that the response of a given multisensory neuron is shaped in a systematic way by the activity of other neurons in the population without requiring changes in the synaptic inputs to individual cells. A strong candidate for such a network effect is divisive normalization[103], a computation in which the input to each neuron is divided by the summed activity of a population of neurons known as the normalization 'pool'. A normalization pool could consist of all neurons within a functional brain area, all neurons within some distance of the target neuron to be normalized or all neurons that share some stimulus feature preferences with the target neuron. Normalization is similar in concept to lateral inhibition, except that lateral inhibition typically involves subtraction, whereas normalization involves a ratio of the driving input to the target neuron and the summed activity of the normalization pool. Normalization can be stated mathematically as follows:

$$R = \frac{E^n}{\alpha^n + \frac{1}{N}\sum_{j=1}^{N} E_j^n} \quad (1)$$

In this standard form of the normalization equation[103], the excitatory drive E is passed through an expansive non-linearity (the exponent n, representing spike generation) and is then divided by the aggregate activity of all neurons in the normalization pool (E_j) plus a constant (α^n) determining how neurons respond to increased stimulus intensity. R is resulting output of the target neuron.

Divisive normalization is believed to be a near-universal feature of neural computation across many species and brain areas (reviewed in REF. 103). It has been implicated in dozens of studies, including those of vision, audition, olfaction, spatial attention and even higher cognitive processes such as economic decision-making[104]. It could be that normalization is so common because it can be realized with various biophysical mechanisms, and still it provides several key advantages for information processing[103].

Ohshiro *et al.*[105] recently developed a model in which divisive normalization occurs at the level of multisensory integration (FIG. 4a,b). In this model, which is similar in concept to the one described in BOX 2, unisensory inputs converge on a topographically aligned multisensory neuron (FIG. 4a) with synaptic weights d_1 and d_2 (called 'modality dominance weights' in REF. 105), and the spiking output of the multisensory layer is influenced by lateral connections that may implement normalization (FIG. 4b). We found that divisive normalization can give rise to a neural combination rule similar to what we measured experimentally in area MSTd[94,100], the intuition for which is as follows. The pooled normalization signal in the visual condition changes greatly with motion coherence, whereas the effect of coherence on the normalization pool is weaker in the combined condition owing to the contribution of vestibular signals that do not depend on coherence. As a result, the weights in the neural combination rule depend on coherence in a manner similar to that shown by MSTd neurons[105].

In an intriguing convergence, it turns out that the normalization model[105] also accounts for the classic empirical principles of multisensory integration[46], including the relationship between stimulus strength and sub- versus superadditivity[55]. For an example of how a well-known multisensory property follows intuitively from a divisive normalization mechanism, consider the phenomenon of cross-modal suppression in the SC[106]. In this manifestation of the spatial principle, the response of a neuron depends in a peculiar way on the intensities of two stimuli that are presented at

Divisive normalization
A neural computation in which the would-be response of an individual neuron (that is, its excitatory drive) is divided by the summed activity of a pool of neurons before generating an output.

different locations (illustrated schematically in the top row of FIG. 4c). One stimulus is presented at the centre of the neuron's receptive field and produces responses that increase sharply with stimulus intensity (indicated by the red curve in the graph in FIG. 4c). A second stimulus is presented near the edge of the receptive field and produces much weaker excitatory responses that also increase with intensity (indicated by the blue curve in the graph in FIG. 4c). Surprisingly, when both stimuli are presented together at high intensities, they elicit a weaker response than the more effective stimulus alone (indicated by the purple curve in the graph in FIG. 4c). Normalization produces this phenomenon because the non-optimal second stimulus contributes more strongly to the normalization pool (the denominator of equation 1, which includes neurons with receptive fields aligned with the second stimulus) than to the driving input onto the neuron being studied (the numerator of equation 1). This scenario predicts that cross-modal suppression can be triggered by non-optimal stimuli that are excitatory to the neuron when presented alone (blue curve in the graph in FIG. 4c), which is analogous to empirical observations in the primary visual cortex[107] and consistent with preliminary results in area MSTd (T. Ohshiro, D.E.A. and G.C.D., unpublished observations).

To conclude, divisive normalization may provide a unifying computational framework for understanding various empirical observations regarding multisensory integration in both the brainstem and cortex. Further work is needed to flesh out a more complete model that connects optimal probabilistic computations (for example, the PPC theory) to normalization and perhaps also incorporates feedback or modulatory input from anatomically and functionally distinct neuronal populations.

Concluding remarks and future directions

Just as the brain benefits from exploiting multiple sources of sensory information, we hope to have convinced the reader of the benefits of drawing upon multiple experimental and theoretical tools when approaching a difficult problem such as multisensory integration. We believe that a mechanistic understanding of how neural circuits give rise to multisensory perception and behaviour requires a detailed examination of the intervening computations[108]. Because sensory information is inherently probabilistic, it is natural to frame these computations in the language of statistical decision theory and the analysis of ideal observers, an approach that has been largely absent from traditional multisensory neurophysiology and neuroimaging (but see REFS 109,110).

In our system of interest (visual and vestibular integration for heading perception), we have outlined — with the benefit of hindsight — a comprehensive strategy for linking psychophysical performance, physiological measurements in behaving animals and computational modelling in a multisensory paradigm. This strategy consisted of: training animals in a multisensory perceptual task and comparing their behaviour with normative predictions[26,30]; identifying a candidate brain area that is likely to contribute to the behaviour[26,93,111,112]; establishing neural correlates of psychophysical measurements[26,100];

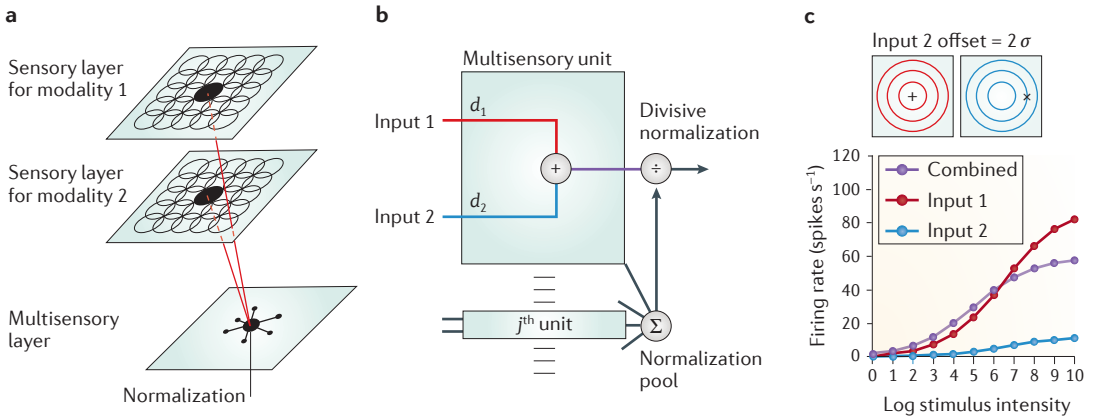


Figure 4 | **The normalization model of multisensory integration.** **a** | In this model, as in the simplified conceptual model of BOX 2, unisensory neurons from separate populations send inputs to a topographically aligned multisensory neuron. **b** | Unisensory inputs are multiplied by synaptic weights d_1 and d_2 (which are fixed for a given neuron) and summed to generate the driving input to a particular multisensory neuron. This driving input is then divided by the summed responses of the rest of the population (the normalization pool; see equation 1 under the section 'Closing the loop'). **c** | In addition to explaining the origin of the multisensory combination rule in the dorsal medial superior temporal area (MSTd)[94,105], divisive normalization can account for the classic empirical principles of multisensory integration made famous by studies of the superior colliculus (SC)[46,51]. One such phenomenon is called the spatial principle, which is illustrated here as a case of cross-modal suppression. One stimulus ('input 1', + in red circles) is presented at the centre of the receptive field of a simulated SC neuron, while a second stimulus ('input 2', x in blue circles) is presented two SDs (2σ) away from the centre. At relatively high (>7) intensities of the two inputs, the response to the combined inputs (purple curve) is less than the response to input 1 alone (that is, cross-modal suppression), even though input 2 alone is excitatory. This results from the contribution of input 2 to the normalization pool. Figure is modified, with permission, from REF. 105 © (2011) Macmillan Publishers Ltd. All rights reserved.

quantifying the neuronal combination rule across many stimulus conditions and levels of cue reliability[94]; comparing the observed combination rule with one that is mathematically optimal, given the observed neural tuning properties in the candidate brain area[100]; and constructing a model of the population that quantitatively reproduces both physiological and behavioural observations[105] using a relatively simple and widespread neural computation[103].

Much work remains to be done in this system, including following up untested predictions of the normalization model[105], defining the specific contributions of the various cortical areas believed to participate in the perception of self-motion [88,90,93,113] and understanding how visual and vestibular integration accounts for the potentially confounding effects of eye, head and object movements on self-motion perception. Indeed, the general problem of inferring self-motion in the presence of moving objects may require neural solutions to the causal inference problem[41], as it is necessary to determine whether a given pattern of retinal image motion was generated by self-motion alone or by self-motion combined with object motion. In the meantime, we suggest that the strategy described in this Review might serve as a roadmap for studying the neural underpinnings of cue integration in other systems. Evidence is rapidly mounting that multisensory interactions are more fundamental to perception[114–116] and cortical information processing[10,12–14] than previously thought. Hence, the quest to understand multisensory integration will be vital for expanding our knowledge of both normal and abnormal brain function.

1. Faisal, A. A., Selen, L. P. & Wolpert, D. M. Noise in the nervous system. *Nature Rev. Neurosci.* **9**, 292–303 (2008).
2. Gepshtein, S. Two psychologies of perception and the prospect of their synthesis. *Philos. Psychol.* **23**, 217–281 (2010).
3. Knill, D. C. & Pouget, A. The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* **27**, 712–719 (2004).
A concise review that provides a good introduction to the idea of sensory uncertainty and the Bayesian perspective on behaviour and neural coding, including the incorporation of priors and studies of motor control.
4. Einstein, A. Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen. *Jahrbuch Radioaktivität Elektronik* **4**, 411–462 (1907).
5. Angelaki, D. E., Shaikh, A. G., Green, A. M. & Dickman, J. D. Neurons compute internal models of the physical laws of motion. *Nature* **430**, 560–564 (2004).
6. Merfeld, D., Zupan, L. & Peterka, R. Humans use internal models to estimate gravity and linear acceleration. *Nature* **398**, 615–618 (1999).
7. Landy, M. S., Banks, M. S. & Knill, D. C. in *Sensory Cue Integration* (eds Trommershäuser, J., Kording, K. P. & Landy, M. S.) 5–29 (Oxford Univ. Press, 2011).
8. Angelaki, D. E., Gu, Y. & DeAngelis, G. C. Multisensory integration: psychophysics, neurophysiology, and computation. *Curr. Opin. Neurobiol.* **19**, 452–458 (2009).
9. Kajikawa, Y., Falchier, A., Musacchia, G., Lakatos, P. & Schroeder, C. E. in *The Neural Bases of Multisensory Processes* (eds Murray, M. & Wallace, M.) (CRC Press, 2012).
10. Lakatos, P., Chen, C. M., O'Connell, M. N., Mills, A. & Schroeder, C. E. Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* **53**, 279–292 (2007).
11. Kayser, C., Petkov, C. I., Remedios, R. & Logothetis, N. K. in *The Neural Bases of Multisensory Processes* (eds Murray, M. & Wallace, M.) (CRC Press, 2012).
12. Kayser, C. & Logothetis, N. K. Do early sensory cortices integrate cross-modal information? *Brain Struct. Funct.* **212**, 121–132 (2007).
13. Ghazanfar, A. A. & Schroeder, C. E. Is neocortex essentially multisensory? *Trends Cogn. Sci.* **10**, 278–285 (2006).
14. Driver, J. & Noesselt, T. Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. *Neuron* **57**, 11–23 (2008).
15. Geisler, W. S. Contributions of ideal observer theory to vision research. *Vision Res.* **51**, 771–781 (2011).
16. Jacobs, R. A. Optimal integration of texture and motion cues to depth. *Vision Res.* **39**, 3621–3629 (1999).
17. 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).
Focusing on the array of visual cues available for the perception of depth, this paper develops several key ideas underlying contemporary ideal observer models of cue integration and also introduces a psychophysical procedure that has become a standard method for testing such models.
18. Maloney, L. T. & Landy, M. S. in *Proceedings of the SPIE* (ed. Pearlman, W. A.) 1154–1163 (SPIE, 1989).
19. Clark, J. J. & Yuille, A. L. *Data Fusion For Sensory Information Processing Systems* (Kluwer, 1990).
20. Alais, D. & Burr, D. The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* **14**, 257–262 (2004).
21. 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).
22. Ernst, M. O. & Banks, M. S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429–433 (2002).
This study provides one of the earliest and clearest psychophysical demonstrations of optimal cue integration across separate sensory modalities. The authors showed that human subjects integrate vision and touch to estimate the width of a grasped object by taking into account the relative reliability of the cues and combining them to improve their performance. Importantly, cue reliability was varied randomly from trial to trial, suggesting that the brain may not need to explicitly learn or represent the uncertainty of the cues to accomplish the task.
23. Sober, S. J. & Sabes, P. N. Flexible strategies for sensory integration during motor planning. *Nature Neurosci.* **8**, 490–497 (2005).
24. van Beers, R. J., Sittig, A. C. & Gon, J. J. Integration of proprioceptive and visual position-information: an experimentally supported model. *J. Neurophysiol.* **81**, 1355–1364 (1999).
25. Körding, K. P. & Wolpert, D. M. Bayesian decision theory in sensorimotor control. *Trends Cogn. Sci.* **10**, 319–326 (2006).
26. Gu, Y., Angelaki, D. E. & DeAngelis, G. C. Neural correlates of multisensory cue integration in macaque MSTd. *Nature Neurosci.* **11**, 1201–1210 (2008).
Using a visual–vestibular heading discrimination task, this study showed that monkeys, like humans, are capable of combining sensory cues to improve perceptual performance. The authors also characterized a population of neurons in extrastriate visual cortex (specifically, in area MSTd) that could underlie the behaviour.
27. de Winkel, K. N., Weesie, J., Werkhoven, P. J. & Groen, E. L. Integration of visual and inertial cues in perceived heading of self-motion. *J. Vis.* **10**, 1 (2010).
28. Edwards, M., O'Mahony, S., Ibbotson, M. R. & Kohlhagen, S. Vestibular stimulation affects optic-flow sensitivity. *Perception* **39**, 1303–1310 (2010).
29. Butler, J. S., Smith, S. T., Campos, J. L. & Bulthoff, H. H. Bayesian integration of visual and vestibular signals for heading. *J. Vis.* **10**, 23 (2010).
30. 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).
31. Cochran, W. G. Problems arising in the analysis of a series of similar experiments. *Appl. Stat.* **4**, 102–118 (1937).
32. Yuille, A. L. & Bülthoff, H. H. in *Perception as Bayesian Inference* (eds Knill, D. C. & Richards, W.) 123–162 (Cambridge Univ. Press, 1996).
33. 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).
34. 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).
35. 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).
36. Clemens, I. A., De Vrijer, M., Selen, L. P., Van Gisbergen, J. A. & Medendorp, W. P. Multisensory processing in spatial orientation: an inverse probabilistic approach. *J. Neurosci.* **31**, 5365–5377 (2011).
37. Ma, W., Beck, J. M. & Pouget, A. in *Sensory Cue Integration* (eds Trommershäuser, J., Kording, K. P. & Landy, M. S.) 393–405 (Oxford Univ. Press, 2011).
38. Oruc, I., Maloney, L. T. & Landy, M. S. Weighted linear cue combination with possibly correlated error. *Vision Res.* **43**, 2451–2468 (2003).
39. 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).
40. Rosas, P., Wichmann, F. A. & Wagemans, J. Texture and object motion in slant discrimination: failure of reliability-based weighting of cues may be evidence for strong fusion. *J. Vis.* **7**, 3 (2007).
41. Körding, K. P. *et al.* Causal inference in multisensory perception. *PLoS ONE* **2**, e943 (2007).
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**, 5 (2007).
43. Ernst, M. O. & Di Luca, M. in *Sensory Cue Integration* (eds Trommershäuser, J., Landy, M. S. & Kording, K. P.) 224–250 (Oxford Univ. Press, 2011).
44. Zaidel, A., Turner, A. H. & Angelaki, D. E. Multisensory calibration is independent of cue reliability. *J. Neurosci.* **31**, 13949–13962 (2011).
45. Raposo, D., Sheppard, J. P., Schrater, P. R. & Churchland, A. K. Multisensory decision-making in rats and humans. *J. Neurosci.* **32**, 3726–3735 (2012).
46. Stein, B. E. & Meredith, M. A. *The Merging of the Senses* (MIT Press, 1993).
47. Meredith, M. & Stein, B. Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J. Neurophysiol.* **56**, 640–662 (1986).
This paper is among the first to demonstrate the impressive capacity of SC neurons to combine visual, tactile and auditory cues, yielding multisensory responses that were often considerably enhanced (and sometimes suppressed) relative to unisensory responses.

These early observations laid the foundation for the well-known empirical 'principles' of multisensory integration, including the spatial and temporal principle and inverse effectiveness.
48. Sparks, D. L. Translation of sensory signals into commands for control of saccadic eye movements: role of primate superior colliculus. *Physiol. Rev.* **66**, 118–171 (1986).
49. Stein, B. E. Development of the superior colliculus. *Annu. Rev. Neurosci.* **7**, 95–125 (1984).
50. Wurtz, R. H. & Albano, J. E. Visual-motor function of the primate superior colliculus. *Annu. Rev. Neurosci.* **3**, 189–226 (1980).
51. Stein, B. E. & Stanford, T. R. Multisensory integration: current issues from the perspective of the single neuron. *Nature Rev. Neurosci.* **9**, 255–266 (2008).
52. James, T. W. & Stevenson, R. A. in *The Neural Bases of Multisensory Processes* (eds Murray, M. & Wallace, M.) (CRC Press, 2012).
53. Beauchamp, M. S., Lee, K. E., Argall, B. D. & Martin, A. Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* **41**, 809–823 (2004).
54. Sarko, D. K. *et al.* in *The Neural Bases of Multisensory Processes* (eds Murray, M. M. & Wallace, M. T.) (CRC Press, 2012).
55. Stanford, T. R., Quessy, S. & Stein, B. E. Evaluating the operations underlying multisensory integration in the cat superior colliculus. *J. Neurosci.* **25**, 6499–6508 (2005).
This study explores multisensory interactions in the SC in a more systematic manner than previous work by varying the intensity and timing of visual and auditory stimuli. The results suggest that most SC neurons combine their inputs additively and that the oft-cited phenomenon of 'superadditivity' may only occur for very weak stimuli.
56. Jiang, W., Jiang, H. & Stein, B. E. Two corticotectal areas facilitate multisensory orientation behavior. *J. Cogn. Neurosci.* **14**, 1240–1255 (2002).
57. Stein, B. E., Huneycutt, W. S. & Meredith, M. A. Neurons and behavior: the same rules of multisensory integration apply. *Brain Res.* **448**, 355–358 (1988).
58. Wilkinson, L. K., Meredith, M. A. & Stein, B. E. The role of anterior ectosylvian cortex in cross-modality orientation and approach behavior. *Exp. Brain Res.* **112**, 1–10 (1996).
59. Frens, M. A., Van Opstal, A. J. & Van der Willigen, R. F. Spatial and temporal factors determine auditory–visual interactions in human saccadic eye movements. *Percept. Psychophys.* **57**, 802–816 (1995).
60. Wallace, M. T., Meredith, M. A. & Stein, B. E. Converging influences from visual, auditory, and somatosensory cortices onto output neurons of the superior colliculus. *J. Neurophysiol.* **69**, 1797–1809 (1993).
61. Jiang, W., Wallace, M. T., Jiang, H., Vaughan, J. W. & Stein, B. E. Two cortical areas mediate multisensory integration in superior colliculus neurons. *J. Neurophysiol.* **85**, 506–522 (2001).
62. Wallace, M. T. & Stein, B. E. Development of multisensory neurons and multisensory integration in cat superior colliculus. *J. Neurosci.* **17**, 2429–2444 (1997).
63. Wallace, M. T. & Stein, B. E. Onset of cross-modal synthesis in the neonatal superior colliculus is gated by the development of cortical influences. *J. Neurophysiol.* **83**, 3578–3582 (2000).
64. Wallace, M. T. & Stein, B. E. Sensory and multisensory responses in the newborn monkey superior colliculus. *J. Neurosci.* **21**, 8886–8894 (2001).
65. Perrault, T. J. Jr, Vaughan, J. W., Stein, B. E. & Wallace, M. T. Neuron-specific response characteristics predict the magnitude of multisensory integration. *J. Neurophysiol.* **90**, 4022–4026 (2003).
66. Alvarado, J. C., Vaughan, J. W., Stanford, T. R. & Stein, B. E. Multisensory versus unisensory integration: contrasting modes in the superior colliculus. *J. Neurophysiol.* **97**, 3193–3205 (2007).
67. Perrault, T. J. Jr, Vaughan, J. W., Stein, B. E. & Wallace, M. T. Superior colliculus neurons use distinct operational modes in the integration of multisensory stimuli. *J. Neurophysiol.* **93**, 2575–2586 (2005).
68. Cuppini, C., Ursino, M., Magosso, E., Rowland, B. A. & Stein, B. E. An emergent model of multisensory integration in superior colliculus neurons. *Front. Integr. Neurosci.* **4**, 6 (2010).
69. Patton, P. E. & Anastasio, T. J. Modeling cross-modal enhancement and modality-specific suppression in multisensory neurons. *Neural Comput.* **15**, 783–810 (2003).
70. Alvarado, J. C., Rowland, B. A., Stanford, T. R. & Stein, B. E. A neural network model of multisensory integration also accounts for unisensory integration in superior colliculus. *Brain Res.* **1242**, 13–23 (2008).
71. Rowland, B. A., Stein, B. E. & Stanford, T. R. in *Sensory Cue Integration* (eds Trommershäuser, J., Kording, K. P. & Landy, M. S.) 333–344 (Oxford Univ. Press, 2011).
72. Green, D. M. & Swets, J. A. *Signal Detection Theory and Psychophysics* (Wiley, 1966).
73. Kim, B. & Basso, M. A. Saccade target selection in the superior colliculus: a signal detection theory approach. *J. Neurosci.* **28**, 2991–3007 (2008).
74. Kim, B. & Basso, M. A. A probabilistic strategy for understanding action selection. *J. Neurosci.* **30**, 2340–2355 (2010).
75. Ma, W. J., Beck, J. M., Latham, P. E. & Pouget, A. Bayesian inference with probabilistic population codes. *Nature Neurosci.* **9**, 1432–1438 (2006).
This theoretical study outlines a simple and flexible strategy for performing optimal Bayesian inference with populations of neurons, using multisensory cue integration as a primary example. By exploiting a mathematical property of neuronal noise, the authors show that simple summation of population activity in sensory areas can be sufficient to implement Bayes-optimal cue integration.
76. Beck, J. M. *et al.* Probabilistic population codes for Bayesian decision making. *Neuron* **60**, 1142–1152 (2008).
77. Jazayeri, M. & Movshon, J. A. Optimal representation of sensory information by neural populations. *Nature Neurosci.* **9**, 690–696 (2006).
78. Ma, W. J. & Pouget, A. Linking neurons to behavior in multisensory perception: a computational review. *Brain Res.* **1242**, 4–12 (2008).
79. Populin, L. C. & Yin, T. C. Bimodal interactions in the superior colliculus of the behaving cat. *J. Neurosci.* **22**, 2826–2834 (2002).
80. Stanford, T. R. & Stein, B. E. Superadditivity in multisensory integration: putting the computation in context. *Neuroreport* **18**, 787–792 (2007).
81. Perrault, T. J. Jr, Stein, B. E. & Rowland, B. A. Non-stationarity in multisensory neurons in the superior colliculus. *Front. Psychol.* **2**, 144 (2011).
82. Gibson, J. J. *The Perception of the Visual World* (Houghton-Mifflin, 1950).
83. Warren, W. H. Jr, Morris, M. W. & Kalish, M. Perception of translational heading from optical flow. *J. Exp. Psychol. Hum. Percept. Perform.* **14**, 646–660 (1988).
84. Benson, A. J., Spencer, M. B. & Stott, J. R. Thresholds for the detection of the direction of whole-body, linear movement in the horizontal plane. *Aviat. Space Environ. Med.* **57**, 1088–1096 (1986).
85. Fernandez, C. & Goldberg, J. M. Physiology of peripheral neurons innervating otolith organs of the squirrel monkey. II. Directional selectivity and force-response relations. *J. Neurophysiol.* **39**, 985–995 (1976).
86. Fernandez, C. & Goldberg, J. M. Physiology of peripheral neurons innervating otolith organs of the squirrel monkey. I. Response to static tilts and to long-duration centrifugal force. *J. Neurophysiol.* **39**, 970–984 (1976).
87. Guedry, F. E. in *Handbook of Sensory Physiology, The Vestibular System* (ed. Kornhuber, H. H.) (Springer, 1974).
88. 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).
89. 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).
90. Chen, A., DeAngelis, G. C. & Angelaki, D. E. Convergence of vestibular and visual self-motion signals in an area of the posterior sylvian fissure. *J. Neurosci.* **31**, 11617–11627 (2011).
91. Duffy, C. J. MST neurons respond to optic flow and translational movement. *J. Neurophysiol.* **80**, 1816–1827 (1998).
92. Britten, K. H. & van Wezel, R. J. Electrical microstimulation of cortical area MST biases heading perception in monkeys. *Nature Neurosci.* **1**, 59–63 (1998).
93. Gu, Y., DeAngelis, G. C. & Angelaki, D. E. Causal links between dorsal medial superior temporal area neurons and multisensory heading perception. *J. Neurosci.* **32**, 2299–2313 (2012).
94. Morgan, M. L., DeAngelis, G. C. & Angelaki, D. E. Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron* **59**, 662–673 (2008).
The authors recorded neuronal responses in area MSTd to a wide array of stimulus combinations, and found a simple mathematical rule by which multisensory neurons combine their inputs. Neurons appeared to take a weighted sum of their inputs, with weights that depend on cue reliability — a surprising finding that conflicted with theoretical predictions and was not well understood until later.
95. Holmes, N. P. & Spence, C. Multisensory integration: space, time and superadditivity. *Curr. Biol.* **15**, R762–R764 (2005).
96. Parker, A. J. & Newsome, W. T. Sense and the single neuron: probing the physiology of perception. *Annu. Rev. Neurosci.* **21**, 227–277 (1998).
97. Bradley, A., Skottun, B. C., Ohzawa, I., Sclar, G. & Freeman, R. D. Visual orientation and spatial frequency discrimination: a comparison of single neurons and behavior. *J. Neurophysiol.* **57**, 755–772 (1987).
98. Green, D. M. & Swets, J. A. *Signal Detection Theory and Psychophysics* (Wiley, 1966).
99. Britten, K. H., Shadlen, M. N., Newsome, W. T. & Movshon, J. A. The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci.* **12**, 4745–4765 (1992).
100. Fetsch, C. R., Pouget, A., DeAngelis, G. C. & Angelaki, D. E. Neural correlates of reliability-based cue weighting during multisensory integration. *Nature Neurosci.* **15**, 146–154 (2011).
In this study, the authors recorded from area MSTd while monkeys performed a cue-conflict version of the visual–vestibular heading task. Behaviourally, the animals were able to re-weight the cues on a trial-by-trial basis as reliability varied, and neuronal activity accounted for the behavioural results. The authors also derived a mathematically optimal combination rule for this task and used it to help explain deviations from optimality at the level of behaviour and neural responses.
101. Paradiso, M. A. A theory for the use of visual orientation information which exploits the columnar structure of striate cortex. *Biol. Cybern.* **58**, 35–49 (1988).
102. Seung, H. S. & Sompolinsky, H. Simple models for reading neuronal population codes. *Proc. Natl Acad. Sci. USA* **90**, 10749–10753 (1993).
103. Carandini, M. & Heeger, D. J. Normalization as a canonical neural computation. *Nature Rev. Neurosci.* **13**, 51–62 (2011).
This review summarizes the large and diverse literature on the divisive normalization, including candidate biophysical mechanisms as well as its benefits for the efficiency of neural coding.
104. Louie, K., Grattan, L. E. & Glimcher, P. W. Reward value-based gain control: divisive normalization in parietal cortex. *J. Neurosci.* **31**, 10627–10639 (2011).
105. Ohshiro, T., Angelaki, D. E. & DeAngelis, G. C. A normalization model of multisensory integration. *Nature Neurosci.* **14**, 775–782 (2011).
This paper presents a computational model that relies on a widespread neural computation called divisive normalization. Normalization of responses at the level of multisensory integration helps to explain several key empirical findings: the reliability-dependent combination rule in area MSTd as well as the ubiquitous empirical principles that were initially described in classic studies of the SC.
106. Meredith, M. A. & Stein, B. E. Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Res.* **365**, 350–354 (1986).
107. Carandini, M., Heeger, D. & Movshon, J. Linearity and normalization in simple cells of the macaque primary visual cortex. *J. Neurosci.* **17**, 8621–8644 (1997).
108. Carandini, M. From circuits to behavior: a bridge too far? *Nature Neurosci.* **15**, 507–509 (2012).
109. Beauchamp, M. S., Pasalar, S. & Ro, T. Neural substrates of reliability-weighted visual-tactile multisensory integration. *Front. Syst. Neurosci.* **4**, 25 (2010).
110. Helbig, H. B. *et al.* The neural mechanisms of reliability weighted integration of shape information from vision and touch. *Neuroimage* **60**, 1063–1072 (2012).
111. 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).

112. Gu, Y., DeAngelis, G. C. & Angelaki, D. E. A functional link between area MSTd and heading perception based on vestibular signals. *Nature Neurosci.* **10**, 1038–1047 (2007).
113. Chen, A., DeAngelis, G. C. & Angelaki, D. E. A comparison of vestibular spatiotemporal tuning in macaque parietoinsular vestibular cortex, ventral intraparietal area, and medial superior temporal area. *J. Neurosci.* **31**, 3082–3094 (2011).
114. Shams, L., Kamitani, Y. & Shimojo, S. Illusions. What you see is what you hear. *Nature* **408**, 788 (2000).
115. Shimojo, S. & Shams, L. Sensory modalities are not separate modalities: plasticity and interactions. *Curr. Opin. Neurobiol.* **11**, 505–509 (2001).
116. Sekuler, R., Sekuler, A. B. & Lau, R. Sound alters visual motion perception. *Nature* **385**, 308 (1997).
117. Avillac, M., Denève, S., Olivier, E., Pouget, A. & Duhamel, J.-R. Reference frames for representing visual and tactile locations in parietal cortex. *Nature Neurosci.* **8**, 941–949 (2005).
118. Wallace, M. T. *et al.* Unifying multisensory signals across time and space. *Exp. Brain Res.* **158**, 252–258 (2004).
119. Fetsch, C. R., DeAngelis, G. C. & Angelaki, D. E. Visual–vestibular cue integration for heading perception: applications of optimal cue integration theory. *Eur. J. Neurosci.* **31**, 1721–1729 (2010).

Acknowledgements
We acknowledge the prodigious contributions of Y. Gu (Institute of Neuroscience, Shanghai, China) to the research discussed in this Review. We also thank M. Morgan, T. Ohshiro and many other current and former laboratory members, collaborators and technicians who made the work possible. Research was supported by US National Institutes of Health grants R01-EY016178 (to G.C.D.) and R01-EY019087 (to D.E.A.).

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
Gregory C. DeAngelis's homepage: http://www.bcs.rochester.edu/people/gdeangelis/index.html
Dora E. Angelaki's homepage: http://neuro.bcm.edu/?sct=gfaculty&prf=62

ALL LINKS ARE ACTIVE IN THE ONLINE PDF