

Published in final edited form as:

Nat Rev Neurosci. 2013 June ; 14(6): . doi:10.1038/nrn3503.

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

Christopher R. Fetsch, Gregory C. DeAngelis[#], and Dora E. Angelaki[#]

[#] These authors contributed equally to this work.

Abstract

The richness of perceptual experience, as well as its usefulness for guiding behavior, depends upon the synthesis of information across multiple senses. Recent decades have witnessed a surge in our understanding of how the brain combines sensory signals, or cues. Much of this research has been guided by one of two distinct approaches, one driven primarily by neurophysiological observations, the other 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 (e.g., the stochastic arrival of photons at the retina) and the transformation of these physical events into messages carried by noisy devices, namely neurons¹. Following in the tradition of 19th 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 represents both the normative basis and evolutionary advantage of cue integration: 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 (e.g., 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 principle⁴ – 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 integration⁷). Moreover, we will only attempt to shed light on one small corner of this rapidly growing field; for instance, we will not catalog the many tasks and modalities for which cue integration behavior has been tested^{3, 7, 8}, nor will we survey the intriguing recent findings of cross-modal influences within early or primary sensory structures⁹⁻¹⁴. Instead,

we will build toward answering the following questions: what are (and what should be) the computations performed by individual multisensory neurons and populations of neurons, and how do these computations give rise to behavioral 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 toward unifying two of the main sub-disciplines in the field: (1) the theory-driven study of cue integration behavior in human subjects, and (2) the use of neurophysiology (and more recently, neuroimaging) to describe the phenomenology of multisensory interactions within neurons and brain areas. We begin with a brief review of the former.

Psychophysical study of cue integration

Before one can address the neural computations underlying a behavior of interest, it is often 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 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 research¹⁶⁻¹⁹, but has been extended to include auditory^{20, 21}, somatosensory/proprioceptive²²⁻²⁵, and vestibular²⁶⁻³⁰ modalities. Many of these studies employ a simple yet highly successful model, 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 with the lowest possible degree of uncertainty, or variance, while also remaining unbiased (i.e., correct on average). This estimate is achieved by weighting the cues in proportion to their relative reliability, or inverse variance (Eq. 1)^{17, 18, 31}. With a few simplifying assumptions, this weighting strategy is identical to one afforded by Bayesian probability theory^{3, 19, 32} (see 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 (Eq. 1) and its extension to Bayesian inference (Eq. 3) are entirely intuitive and do not depend on an understanding of the mathematical details. In essence, these models assert that the brain should consider all available evidence when making a decision (or estimate), while ensuring that more reliable evidence has greater influence. If one were to hear two weather forecasts, one from a reputable meteorologist and the other from an eccentric neighbor, 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 behavioral experiments

Optimal cue integration models can be tested by asking subjects to perform a psychophysical discrimination task using multiple cues, 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 (Eq. 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 while 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 another³⁶. More fundamentally, the existence of neuronal noise guarantees conflict between parameter estimates from separate modalities even when stimuli are not actually in conflict³⁷, 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 by REF. 7). There are some notable exceptions³⁸⁻⁴⁰, and it is clear that a more complete picture will require schemes that go beyond simple 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' inference⁴¹), 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, and more recently nonhuman primates³⁰ and rodents⁴⁵, are able to reduce perceptual uncertainty and improve their performance by combining multiple cues in a manner that approximates a statistically optimal observer (Eq. 1). Since 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 by REFS ^{46, 47}).

A region that has received the bulk of neurophysiologists' attention is the mammalian superior colliculus (SC), a midbrain structure involved primarily in orienting the eyes and head toward salient stimuli⁴⁸⁻⁵⁰. 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. But beyond mere convergence of multi-modal information, it soon became clear that these signals interact functionally within the SC, often in dramatic fashion⁴⁷. For example, weak visual or auditory stimuli presented alone might elicit a very small response from an SC neuron, yet when presented together 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 fMRI 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 as follows⁵⁵: (1) The spatial/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. In contrast, sufficient separation in space or time can suppress the multisensory response relative to the best unisensory response. (2) 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 occurs⁵⁵.

Importantly, these physiological response properties bear an intriguing resemblance to behavior in alert animals. For example, stimulus combinations that were more effective in activating neurons were also more effective at driving behavioral detection of stimuli⁵⁶⁻⁵⁹. 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 trajectory⁶²⁻⁶⁴. 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 colleagues⁵⁵. These authors noted that most 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 behavior. 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 mechanisms⁵⁵. Expanding the repertoire to three levels of intensity, 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 as stimulus strength increases, consistent with the principle of inverse effectiveness (see also REFS ⁶⁵⁻⁶⁷). Whereas nearly all neurons demonstrated inverse effectiveness, superadditivity was generally seen only for neurons with weak responses to unisensory stimuli. These studies helped clarify that superadditivity is not a ubiquitous property of multisensory integration by neurons. However, they still left in doubt the specific mathematical rules by which SC neurons combine their inputs.

Computational modeling and neural theories of cue integration

In the same spirit of attempting to pin down the computations underlying multisensory integration, several computational models⁶⁸⁻⁷⁰ have been presented to account for SC responses and their dependence on cortical input (reviewed by 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 models⁶⁸⁻⁷¹ focused mainly on explaining the physiology, rather than providing quantitative predictions linking neuronal activity with behavior. One issue is that the behavioral 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 the related ideas of signal-detection theory⁷² – as used by psychophysicists (but see REFS ^{73, 74}), and hence the neurophysiological data were not interpreted in that context.

Meanwhile and in contrast, a different theoretical framework⁷⁵ was developed from the perspective of probabilistic (statistical) inference – the process of drawing conclusions based on uncertain data – of which cue integration is a special case (see previous section and Box

1). In a landmark set of studies, Ma, Beck, Latham and Pouget showed how populations of neurons can implicitly represent probability distributions and perform Bayesian computations⁷⁵ and optimal decision-making⁷⁶. The key insight of this “probabilistic population code” (PPC) framework (Fig. 2) is that the type of variability observed in most neuronal responses – termed “Poisson-like” variability⁷⁵ – 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 Eq. 3) can be achieved by simple linear summation of population activity (Fig. 2a; see also 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 (simulated in Fig. 2b, blue and red points) The height, or gain, 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 number of neurons. The summation of unisensory activity (Fig. 2b, green) produces a third hill that is shifted toward the more reliable cue, reflecting the optimal weighting of the cues (Eq. 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 Fig. 2, insets), then propagated to a downstream multisensory area via linear summation⁷⁵.

Thus, the PPC theory predicts that brain regions involved in optimal cue integration should exhibit additive⁷⁵ or subadditive⁷⁸ responses regardless of stimulus strength, in contrast with the emphasis on superadditivity and inverse effectiveness (a dependence on stimulus strength) in the SC literature. Compared to models specifically targeting the SC⁷¹, the key advantage of the PPC framework is its generality – it can be 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, as well as 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: (a) despite a wealth of data (summarized by the ubiquitous empirical principles⁴⁶) and several detailed models of the SC⁷¹, there has been no consensus as to the mathematical rules by which single neurons combine multisensory inputs, and few if any model-based behavioral predictions that would permit coupling the SC literature to the modern psychophysical paradigm of statistical decision-making; (b) traditional neurophysiology studies did not measure behavior in a psychophysical task while recording from multisensory neurons and attempting to link the two; and (c) an influential theory⁷⁵ designed to capture the probabilistic nature of cue integration made a clear prediction (additive or subadditive summation, independent of stimulus strength) that was at odds with reported nonlinear 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 sub-disciplines 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, utilizing 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 the SC studies, but, as explained below, reveals a putative computational mechanism underlying key aspects of multisensory integration in both areas.

Visual-vestibular cue integration for heading perception

All animals that navigate through their environment need to estimate their direction of self-motion, or 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 at the same time, 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 space⁸⁴⁻⁸⁷.

Several areas in the primate brain receive both vestibular and visual signals related to self-motion⁸⁸⁻⁹⁰. In the context of cue integration, the best studied of these areas is the dorsal subdivision of the medial superior temporal area (MSTd), 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 (‘visual’ condition), or synchronous combinations of the two cues (‘combined’ condition).

Each neuron was tested with many different combinations of visual and vestibular headings at different levels of stimulus strength, or 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 center of the neurons’ receptive fields (thereby focusing on multisensory enhancement), and varied stimulus strength over a modest range, resulting in about a twofold difference in firing rate on average. In contrast, Morgan et al.⁹⁴ presented all combinations of congruent and conflicting stimuli across the full 360-degree range of possible headings (i.e., spanning the complete tuning curve of each neuron, including both preferred and non-preferred stimuli), and varied visual cue reliability (motion coherence of the optic flow display) over a fourfold range.

The aim of the Morgan et al. study⁹⁴ was to define a mathematical ‘combination rule’ that could adequately describe multisensory responses in MSTd. By combination rule we mean an arithmetic expression that describes the response to multisensory stimulus combinations (call it R_{comb}) as a function of the responses to vestibular and visual stimuli presented

separately (R_{vestib} and R_{visual}). To accurately measure the combination rule required testing a broad range of preferred and non-preferred stimuli, probing as much of the stimulus space and dynamic range of the neurons as possible. This comprehensive strategy is important because the contribution of nonlinearities 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 toward super- or subadditivity^{94, 95}. But perhaps the most crucial aspect of this strategy was the presentation of conflicting stimuli (i.e., 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 a variety of 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 behavioral response.

Although in principle the neural combination rule could be fairly complex (e.g., involving nonlinearities), Morgan et al.⁹⁴ found that a simple weighted sum plus a constant ($R_{\text{comb}} = A_{\text{vestib}} * R_{\text{vestib}} + A_{\text{visual}} * R_{\text{visual}} + C$) 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 MSTd is compatible with the PPC theory of Bayes-optimal cue integration⁷⁵. However, a key feature of the MSTd combination rule was not predicted by the basic PPC theory, nor any 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 coherence⁹⁴.

At first blush this outcome seems perfectly reasonable: behavioral choices show evidence for weighting of cues according to their reliability (Fig. 1a,b & Box 1), why shouldn’t single neurons? Here we must clarify the distinctions between different uses of the term ‘weight’, e.g., the weights computed at the level of single neurons versus at the level of behavior. Box 2 explains these distinctions with a simple conceptual model. The take-home message is that the neural weights (combination rule) measured in studies like that of Morgan et al.⁹⁴ do not map onto perceptual weights (Box 1; Fig. 1a,b) in any straightforward way – a caveat that also applies to the empirical principles of Stein and colleagues. 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 behavior – 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 (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 an encoded probability distribution reflecting the sum of unisensory cue reliabilities – precisely the prediction of optimal weighting schemes (Eq. 2). If the summation of activity needs to be counteracted by a baseline shift (i.e., via global inhibition) to keep neurons in their

dynamic range, then single neurons will appear subadditive without compromising the optimality of the computations⁷⁸. 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 weights⁷⁵, the changes in neural weights with reliability described by Morgan et al.⁹⁴ remained a puzzle. In 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 behavior

We and others⁵⁵ have argued that deciphering the computations reflected in single-neuron activity is crucial for constraining models of multisensory integration. However, equally important is to seek evidence supporting the involvement of such neurons in a particular behavior, and a potent way to collect such evidence is to record or manipulate neural activity during a psychophysical task⁹⁶. Until recently, this had not been done for cue integration tasks, as there were few if any suitable animal models for this purpose.

We²⁶ developed a paradigm enabling the simultaneous recording of MSTd neuron activity while monkeys performed a multisensory heading discrimination task. As in the Morgan et al. study⁹⁴, 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 of two choice targets presented at the end of each trial (Fig. 3c).

Behaviorally, Gu et al. found that monkeys show near-optimal (Eq. 2) improvements in perceptual sensitivity when the cues were presented together compared with when they were presented singly (Fig. 3d)²⁶. Subjects also weighted the cues in proportion to their reliability (Fig. 3g)³⁰ as predicted by the standard optimal cue integration model (Eq. 1). We then related neural activity to behavior by converting firing rates into simulated choices made by an ideal observer, using a common technique called ROC analysis⁹⁷⁻⁹⁹. 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 to the behavioral data. The results indicated that MSTd neurons show close parallels with behavior, with respect to both the improvement in sensitivity²⁶ (compare Fig. 3d with 3f) and cue weighting¹⁰⁰ (compare Fig. 3g with 3i).

What should neurons do? Deriving the optimal combination rule

Although we showed that subjects weight visual and vestibular cues in proportion to their reliability³⁰, the perceptual weights were in fact slightly but consistently sub-optimal: both humans and monkeys modestly over-weighted vestibular information in this task, compared with the optimal predictions from Eq. 1. Remarkably, this specific deviation from optimality was reflected in MSTd neuronal activity¹⁰⁰. 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 behavior. As suggested in Box 2, this is not as simple as mapping neural firing rates onto the

choices of a simulated observer (i.e., the ROC analysis strategy discussed above) because such a 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 (the operation shown in Box 2, panel b) is to maximize the information carried by neurons about the variable of interest (here, heading direction). Using Fisher information^{101, 102}, a quantity related to the precision with which an ideal observer can discriminate small changes in a stimulus, we derived a simple expression for the optimal neural weights. Without going into the 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 al⁹⁴: 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 study⁷⁵ – 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 neurons¹⁰⁰ 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 finding⁹⁴ in a completely new light: as part of a neural strategy for combining sensory signals in a near-optimal fashion within a probabilistic population coding 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, panel b) were unlikely to reflect changes in synaptic weights onto individual cells (Box 2, panel a), for two reasons. First, because the animals in Morgan et al.⁹⁴ were not performing a perceptual task, there is no clear basis for expecting reliability to modulate synaptic strength (e.g., 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 time scale 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 rather sophisticated mechanisms for modulating synaptic weights that 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 normalization¹⁰³, a computation in which the response of 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 range of 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:

$$R = \frac{E^n}{\alpha^n + \frac{1}{N} \sum_{j=1}^N E_j^n} \quad [4]$$

In this standard form of the normalization equation¹⁰³, the excitatory drive E is passed through an expansive nonlinearity (the exponent n , simulating spike generation) then divided by the aggregate activity of all neurons in the normalization pool (E_j), plus a constant (α^n).

Divisive normalization is believed to be a near-universal feature of neural computation across many species and brain areas (reviewed by REF. ¹⁰³). It has been implicated in dozens of studies within vision, audition, olfaction, spatial attention, and even higher cognitive processes such as economic decision-making¹⁰⁴. It could be that normalization is so common because it can be realized with a variety of biophysical mechanisms, and yet it provides several key advantages for information processing¹⁰³. A recent theoretical study¹⁰⁵ showed how divisive normalization can implement an important component of many probabilistic computations (marginalization), which may also help explain its ubiquity in neural systems.

Ohshiro et al.¹⁰⁶ recently developed a model in which divisive normalization occurs at the level of multisensory integration (Fig. 4a,b). This model expands on the basic architecture shown schematically in Box 2; namely, that (i) 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. ¹⁰⁶), and that (ii) 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 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 neurons¹⁰⁶.

In an intriguing convergence, it turns out that the normalization model¹⁰⁶ also accounts for the classic empirical principles of multisensory integration⁴², including the spatial/temporal principle, inverse effectiveness, and the relationship between stimulus strength and sub-versus super-additivity⁵⁵. 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 SC¹⁰⁷. 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 different locations (illustrated schematically at the top of Fig. 4c). One stimulus is presented at the center of the neuron's receptive field, and produces responses that increase sharply with stimulus intensity (Fig. 4c, red curve). A second stimulus is presented near the edge of the receptive field, and produces much weaker excitatory responses that also increase with intensity (Fig. 4c, blue curve). Surprisingly, when both stimuli are presented together at high intensities, they elicit a weaker response than the more effective stimulus alone (Fig. 4c, black curve). Normalization produces this phenomenon because the non-optimal second stimulus contributes more strongly to the normalization pool (the denominator of Eq. 4, which includes neurons with receptive fields aligned with the second stimulus) than to the driving input onto the neuron being studied (the numerator of Eq. 4).

This scenario predicts that cross-modal suppression can be triggered by non-optimal stimuli that are excitatory to the neuron when presented alone (Fig. 4c, blue curve), analogous to empirical observations in primary visual cortex¹⁰⁸ and consistent with preliminary results in MSTd (T. Ohshiro, D.E.A., and G.C.D., unpublished observations).

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

Concluding remarks

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 behavior requires a detailed examination of the intervening computations¹⁰⁹. 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^{110, 111}).

In our system of interest (visual-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 modeling in a multisensory paradigm. This strategy consisted of (a) training animals in a multisensory perceptual task and comparing their behavior with normative predictions^{26, 30}, (b) identifying a candidate brain area that likely contributes to the behavior^{26, 93, 112, 113}, (c) establishing neural correlates of psychophysical measurements^{26, 100}, (d) quantifying the neuronal ‘combination rule’ across many stimulus conditions and levels of cue reliability⁹⁴, (e) comparing the observed combination rule with one that is mathematically optimal given the observed neural tuning properties¹⁰⁰, and (f) constructing a model of the population that quantitatively reproduces both physiological and behavioral observations¹⁰⁶ using a relatively simple and widespread neural computation¹⁰³.

Much work remains to be done in this system, including following up untested predictions of the normalization model¹⁰⁶, defining the specific contributions of the various cortical areas believed to participate in self-motion perception^{88, 90, 93, 114}, and understanding how visual-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 problem⁴¹, as it is necessary to determine whether a given pattern of retinal image motion was generated by self-motion alone or self-motion combined with object motion. In the mean time, 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 perception¹¹⁵⁻¹¹⁷ 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.

Glossary Box

reliability, precision, and accuracy	While the term reliability can mean different things in different fields, here we use it as a synonym for the precision, or inverse variance, of a measurement. Accuracy, on the other hand, refers to how close the measurement is to the true value, i.e., how unbiased it is.
normative	A general term referring to an idea, statement, or model that describes how something ought to be; i.e., relating to an ideal or standard of correctness.
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/level differences in audition, and proprioceptive signals (e.g., from muscle spindles) conveying the position of the arm in space.
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 (e.g., 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.
Bayesian theory	The branch of statistics and probability theory in which probability is interpreted as ‘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).
Poisson-like neuronal 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 termed “Poisson-like”). A prominent feature of this family is that the variance of neuronal responses (i.e., the variance of the number of action potentials across repeated stimulus presentations) is proportional to the mean response.
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’).
divisive normalization	A neural computation in which the would-be response of an individual neuron (i.e., its excitatory drive) is divided by the summed activity of a pool of neurons prior to generating an output.

abbreviations

SC	superior colliculus
MSTd	medial superior temporal area, dorsal portion
PPC	probabilistic population code
ROC	receiver operating characteristic
P(A B)	a conditional probability distribution, read as “the probability of A given B.”

REFERENCES

1. Faisal AA, Selen LP, Wolpert DM. Noise in the nervous system. *Nat Rev Neurosci.* 2008; 9:292–303. [PubMed: 18319728]
2. Gepshtein S. Two psychologies of perception and the prospect of their synthesis. *Philosophical Psychology.* 2010; 23:217–281.
3. Knill DC, Pouget A. The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* 2004; 27:712–9. [PubMed: 15541511]
4. Einstein A. Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen. *Jahrbuch der Radioaktivität und Elektronik.* 1907; 4:411–462.
5. Angelaki DE, Shaikh AG, Green AM, Dickman JD. Neurons compute internal models of the physical laws of motion. *Nature.* 2004; 430:560–4. [PubMed: 15282606]
6. Merfeld D, Zupan L, Peterka R. Humans use internal models to estimate gravity and linear acceleration. *Nature.* 1999; 398:615–618. [PubMed: 10217143]
7. Landy, MS.; Banks, MS.; Knill, DC. *Sensory Cue Integration.* Trommershäuser, J.; Kording, KP.; Landy, MS., editors. Oxford University Press; New York: 2011. p. 5-29.
8. Angelaki DE, Gu Y, DeAngelis GC. Multisensory integration: psychophysics, neurophysiology, and computation. *Curr Opin Neurobiol.* 2009; 19:452–8. [PubMed: 19616425]
9. Kajikawa, Y.; Falchier, A.; Musacchia, G.; Lakatos, P.; Schroeder, CE. *The Neural Bases of Multisensory Processes.* Murray, M.; Wallace, M., editors. CRC Press; Boca Raton (FL): 2012.
10. Lakatos P, Chen CM, O’Connell MN, Mills A, Schroeder CE. Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron.* 2007; 53:279–92. [PubMed: 17224408]
11. Kayser, C.; Petkov, CI.; Remedios, R.; Logothetis, NK. *The Neural Bases of Multisensory Processes.* Murray, M.; Wallace, M., editors. CRC Press; Boca Raton (FL): 2012.
12. Kayser C, Logothetis NK. Do early sensory cortices integrate cross-modal information? *Brain Struct Funct.* 2007; 212:121–32. [PubMed: 17717687]
13. Ghazanfar AA, Schroeder CE. Is neocortex essentially multisensory? *Trends Cogn Sci.* 2006; 10:278–285. [PubMed: 16713325]
14. Driver J, Noesselt T. Multisensory interplay reveals crossmodal influences on ‘sensory-specific’ brain regions, neural responses, and judgments. *Neuron.* 2008; 57:11–23. [PubMed: 18184561]
15. Geisler WS. Contributions of ideal observer theory to vision research. *Vision Res.* 2011; 51:771–81. [PubMed: 20920517]
16. Jacobs RA. Optimal integration of texture and motion cues to depth. *Vision Res.* 1999; 39:3621–9. [PubMed: 10746132]
17. Landy MS, Maloney LT, Johnston EB, Young M. Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Res.* 1995; 35:389–412. [PubMed: 7892735]
18. Maloney, LT.; Landy, MS. *Proceedings of the SPIE.* Pearlman, WA., editor. SPIE; Bellingham, WA: 1989. p. 1154-1163.
19. Clark, JJ.; Yuille, AL. *Data fusion for sensory information processing systems.* Kluwer; Norwell, MA: 1990.
20. Alais D, Burr D. The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol.* 2004; 14:257–62. [PubMed: 14761661]

21. Battaglia PW, Jacobs RA, Aslin RN. Bayesian integration of visual and auditory signals for spatial localization. *J Opt Soc Am A Opt Image Sci Vis.* 2003; 20:1391–7. [PubMed: 12868643]
22. Ernst MO, Banks MS. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature.* 2002; 415:429–33. [PubMed: 11807554]
23. Sober SJ, Sabes PN. Flexible strategies for sensory integration during motor planning. *Nat Neurosci.* 2005; 8:490–97. [PubMed: 15793578]
24. van Beers RJ, Sittig AC, Gon JJ. Integration of proprioceptive and visual position-information: An experimentally supported model. *J Neurophysiol.* 1999; 81:1355–64. [PubMed: 10085361]
25. Körding KP, Wolpert DM. Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences.* 2006; 10:319–326. [PubMed: 16807063]
26. Gu Y, Angelaki DE, DeAngelis GC. Neural correlates of multisensory cue integration in macaque MSTd. *Nat Neurosci.* 2008; 11:1201–10. [PubMed: 18776893]
27. de Winkel KN, Weesie J, Werkhoven PJ, Groen EL. Integration of visual and inertial cues in perceived heading of self-motion. *J Vis.* 2010; 10:1. [PubMed: 21047733]
28. Edwards M, O’Mahony S, Ibbotson MR, Kohlhausen S. Vestibular stimulation affects optic-flow sensitivity. *Perception.* 2010; 39:1303–10. [PubMed: 21180352]
29. Butler JS, Smith ST, Campos JL, Bulthoff HH. Bayesian integration of visual and vestibular signals for heading. *J Vis.* 2010; 10:23. [PubMed: 20884518]
30. Fetsch CR, Turner AH, DeAngelis GC, Angelaki DE. Dynamic reweighting of visual and vestibular cues during self-motion perception. *J Neurosci.* 2009; 29:15601–12. [PubMed: 20007484]
31. Cochran WG. Problems arising in the analysis of a series of similar experiments. *Journal of the Royal Statistical Society.* 1937; 4(Suppl.):102–118.
32. Yuille, AL.; Bulthoff, HH. Perception as Bayesian Inference. Knill, DC.; Richards, W., editors. Cambridge University Press; New York, NY: 1996. p. 123-162.
33. Young MJ, Landy MS, Maloney LT. A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision Res.* 1993; 33:2685–96. [PubMed: 8296465]
34. Knill DC, Saunders JA. Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Res.* 2003; 43:2539–58. [PubMed: 13129541]
35. Hillis JM, Watt SJ, Landy MS, Banks MS. Slant from texture and disparity cues: optimal cue combination. *J Vis.* 2004; 4:967–92. [PubMed: 15669906]
36. Clemens IA, De Vrijer M, Selen LP, Van Gisbergen JA, Medendorp WP. Multisensory processing in spatial orientation: an inverse probabilistic approach. *J Neurosci.* 2011; 31:5365–77. [PubMed: 21471371]
37. Ma, W.; Beck, JM.; Pouget, A. Sensory Cue Integration. Trommershäuser, J.; Körding, KP.; Landy, MS., editors. Oxford University Press; New York: 2011.
38. Oruc I, Maloney LT, Landy MS. Weighted linear cue combination with possibly correlated error. *Vision Res.* 2003; 43:2451–68. [PubMed: 12972395]
39. Rosas P, Wagemans J, Ernst MO, Wichmann FA. 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.* 2005; 22:801–9. [PubMed: 15898539]
40. Rosas P, Wichmann FA, Wagemans J. Texture and object motion in slant discrimination: failure of reliability-based weighting of cues may be evidence for strong fusion. *J Vis.* 2007; 7:3. [PubMed: 17685786]
41. Körding KP, et al. Causal inference in multisensory perception. *PLoS ONE.* 2007; 2:e943. [PubMed: 17895984]
42. Knill DC. Robust cue integration: a Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. *J Vis.* 2007; 7(5):1–24.
43. Ernst, MO.; Di Luca, M. Sensory Cue Integration. Trommershäuser, J.; Landy, MS.; Körding, KP., editors. Oxford University Press; New York: 2011. p. 224-50.
44. Zaidel A, Turner AH, Angelaki DE. Multisensory calibration is independent of cue reliability. *Journal of Neuroscience.* 2011; 31:13949–13962. [PubMed: 21957256]

45. Raposo D, Sheppard JP, Schrater PR, Churchland AK. Multisensory decision-making in rats and humans. *J Neurosci.* 2012; 32:3726–35. [PubMed: 22423093]
46. Stein, BE.; Meredith, MA. *The Merging of the Senses.* MIT Press; Cambridge, MA: 1993.
47. Meredith M, Stein B. Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J Neurophysiol.* 1986; 56:640–62. [PubMed: 3537225]
48. Sparks DL. Translation of sensory signals into commands for control of saccadic eye movements: role of primate superior colliculus. *Physiol Rev.* 1986; 66:118–71. [PubMed: 3511480]
49. Stein BE. Development of the superior colliculus. *Annu Rev Neurosci.* 1984; 7:95–125. [PubMed: 6370084]
50. Wurtz RH, Albano JE. Visual-motor function of the primate superior colliculus. *Annu Rev Neurosci.* 1980; 3:189–226. [PubMed: 6774653]
51. Stein BE, Stanford TR. Multisensory integration: current issues from the perspective of the single neuron. *Nat Rev Neurosci.* 2008; 9:255–66. [PubMed: 18354398]
52. James, TW.; Stevenson, RA. *The Neural Bases of Multisensory Processes.* Murray, M.; Wallace, M., editors. CRC Press; Boca Raton (FL): 2012.
53. Beauchamp MS, Lee KE, Argall BD, Martin A. Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron.* 2004; 41:809–23. [PubMed: 15003179]
54. Sarko, DK.; et al. Murray, MM.; Wallace, MT. *The Neural Bases of Multisensory Processes.* CRC Press; Boca Raton (FL): 2012.
55. Stanford TR, Quessy S, Stein BE. Evaluating the operations underlying multisensory integration in the cat superior colliculus. *J Neurosci.* 2005; 25:6499–6508. [PubMed: 16014711]
56. Jiang W, Jiang H, Stein BE. Two corticotectal areas facilitate multisensory orientation behavior. *J Cogn Neurosci.* 2002; 14:1240–55. [PubMed: 12495529]
57. Stein BE, Huneycutt WS, Meredith MA. Neurons and behavior: the same rules of multisensory integration apply. *Brain Res.* 1988; 448:355–8. [PubMed: 3378157]
58. Wilkinson LK, Meredith MA, Stein BE. The role of anterior ectosylvian cortex in cross-modality orientation and approach behavior. *Exp Brain Res.* 1996; 112:1–10. [PubMed: 8951401]
59. Frens MA, Van Opstal AJ, Van der Willigen RF. Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Percept Psychophys.* 1995; 57:802–16. [PubMed: 7651805]
60. Wallace MT, Meredith MA, Stein BE. Converging influences from visual, auditory, and somatosensory cortices onto output neurons of the superior colliculus. *J Neurophysiol.* 1993; 69:1797–1809. [PubMed: 8350124]
61. Jiang W, Wallace MT, Jiang H, Vaughan JW, Stein BE. Two cortical areas mediate multisensory integration in superior colliculus neurons. *J Neurophysiol.* 2001; 85:506–22. [PubMed: 11160489]
62. Wallace MT, Stein BE. Development of multisensory neurons and multisensory integration in cat superior colliculus. *J Neurosci.* 1997; 17:2429–44. [PubMed: 9065504]
63. Wallace MT, Stein BE. Onset of cross-modal synthesis in the neonatal superior colliculus is gated by the development of cortical influences. *J Neurophysiol.* 2000; 83:3578–82. [PubMed: 10848574]
64. Wallace MT, Stein BE. Sensory and multisensory responses in the newborn monkey superior colliculus. *J Neurosci.* 2001; 21:8886–94. [PubMed: 11698600]
65. Perrault TJ Jr. Vaughan JW, Stein BE, Wallace MT. Neuron-specific response characteristics predict the magnitude of multisensory integration. *J Neurophysiol.* 2003; 90:4022–6. [PubMed: 12930816]
66. Alvarado JC, Vaughan JW, Stanford TR, Stein BE. Multisensory versus unisensory integration: contrasting modes in the superior colliculus. *J Neurophysiol.* 2007; 97:3193–205. [PubMed: 17329632]
67. Perrault TJ Jr. Vaughan JW, Stein BE, Wallace MT. Superior colliculus neurons use distinct operational modes in the integration of multisensory stimuli. *J Neurophysiol.* 2005; 93:2575–86. [PubMed: 15634709]

68. Cuppini C, Ursino M, Magosso E, Rowland BA, Stein BE. An emergent model of multisensory integration in superior colliculus neurons. *Front Integr Neurosci.* 2010; 4:6. [PubMed: 20431725]
69. Patton PE, Anastasio TJ. Modeling cross-modal enhancement and modality-specific suppression in multisensory neurons. *Neural Comput.* 2003; 15:783–810. [PubMed: 12689387]
70. Alvarado JC, Rowland BA, Stanford TR, Stein BE. A neural network model of multisensory integration also accounts for unisensory integration in superior colliculus. *Brain Res.* 2008; 1242:13–23. [PubMed: 18486113]
71. Rowland, BA.; Stein, BE.; Stanford, TR. *Sensory Cue Integration.* Trommershäuser, J.; Kording, KP.; Landy, MS., editors. Oxford University Press; New York: 2011. p. 333-344.
72. Green, DM.; Swets, JA. *Signal Detection Theory and Psychophysics.* Wiley; New York: 1966.
73. Kim B, Basso MA. Saccade target selection in the superior colliculus: a signal detection theory approach. *J Neurosci.* 2008; 28:2991–3007. [PubMed: 18354003]
74. Kim B, Basso MA. A probabilistic strategy for understanding action selection. *J Neurosci.* 2010; 30:2340–55. [PubMed: 20147560]
75. Ma WJ, Beck JM, Latham PE, Pouget A. Bayesian inference with probabilistic population codes. *Nat Neurosci.* 2006; 9:1432–8. [PubMed: 17057707]
76. Beck JM, et al. Probabilistic population codes for Bayesian decision making. *Neuron.* 2008; 60:1142–52. [PubMed: 19109917]
77. Jazayeri M, Movshon JA. Optimal representation of sensory information by neural populations. *Nat Neurosci.* 2006; 9:690–6. [PubMed: 16617339]
78. Ma WJ, Pouget A. Linking neurons to behavior in multisensory perception: a computational review. *Brain Res.* 2008; 1242:4–12. [PubMed: 18602905]
79. Populin LC, Yin TC. Bimodal interactions in the superior colliculus of the behaving cat. *J Neurosci.* 2002; 22:2826–34. [PubMed: 11923447]
80. Stanford TR, Stein BE. Superadditivity in multisensory integration: putting the computation in context. *Neuroreport.* 2007; 18:787–92. [PubMed: 17471067]
81. Perrault TJ Jr, Stein BE, Rowland BA. Non-stationarity in multisensory neurons in the superior colliculus. *Front Psychol.* 2011; 2:144. [PubMed: 21772824]
82. Gibson, JJ. *The Perception of the Visual World.* Houghton-Mifflin; Boston, MA: 1950.
83. Warren WH Jr, Morris MW, Kalish M. Perception of translational heading from optical flow. *J Exp Psychol Hum Percept Perform.* 1988; 14:646–60. [PubMed: 2974874]
84. Benson AJ, Spencer MB, Stott JR. Thresholds for the detection of the direction of whole-body, linear movement in the horizontal plane. *Aviat Space Environ Med.* 1986; 57:1088–96. [PubMed: 3790028]
85. Fernandez C, Goldberg JM. Physiology of peripheral neurons innervating otolith organs of the squirrel monkey. II. Directional selectivity and force-response relations. *J Neurophysiol.* 1976; 39:985–95. [PubMed: 824413]
86. Fernandez C, Goldberg JM. Physiology of peripheral neurons innervating otolith organs of the squirrel monkey. I. Response to static tilts and to long-duration centrifugal force. *J Neurophysiol.* 1976; 39:970–84. [PubMed: 824412]
87. Guedry, FE. *Handbook of Sensory Physiology, The Vestibular System.* Kornhuber, HH., editor. Springer-Verlag; New York: 1974.
88. Chen A, DeAngelis GC, Angelaki DE. Representation of vestibular and visual cues to self-motion in ventral intraparietal cortex. *J Neurosci.* 2011; 31:12036–52. [PubMed: 21849564]
89. Gu Y, Watkins PV, Angelaki DE, DeAngelis GC. Visual and nonvisual contributions to three-dimensional heading selectivity in the medial superior temporal area. *J Neurosci.* 2006; 26:73–85. [PubMed: 16399674]
90. Chen A, DeAngelis GC, Angelaki DE. Convergence of vestibular and visual self-motion signals in an area of the posterior sylvian fissure. *J Neurosci.* 2011; 31:11617–27. [PubMed: 21832191]
91. Duffy CJ. MST neurons respond to optic flow and translational movement. *J Neurophysiol.* 1998; 80:1816–27. [PubMed: 9772241]
92. Britten KH, van Wezel RJ. Electrical microstimulation of cortical area MST biases heading perception in monkeys. *Nat Neurosci.* 1998; 1:59–63. [PubMed: 10195110]

93. Gu Y, DeAngelis GC, Angelaki DE. Causal links between dorsal medial superior temporal area neurons and multisensory heading perception. *J Neurosci.* 2012; 32:2299–313. [PubMed: 22396405]
94. Morgan ML, DeAngelis GC, Angelaki DE. Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron.* 2008; 59:662–73. [PubMed: 18760701]
95. Holmes NP, Spence C. Multisensory integration: space, time and superadditivity. *Curr Biol.* 2005; 15:R762–4. [PubMed: 16169476]
96. Parker AJ, Newsome WT. Sense and the single neuron: probing the physiology of perception. *Annu Rev Neurosci.* 1998; 21:227–77. [PubMed: 9530497]
97. Bradley A, Skottun BC, Ohzawa I, Sclar G, Freeman RD. Visual orientation and spatial frequency discrimination: a comparison of single neurons and behavior. *J Neurophysiol.* 1987; 57:755–772. [PubMed: 3559700]
98. Green DM, Swets JA. The Relative Operating Characteristic in Psychology. Book. 2004:1–13.
99. Britten KH, Shadlen MN, Newsome WT, Movshon JA. The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J Neurosci.* 1992; 12:4745–4765. [PubMed: 1464765]
100. Fetsch CR, Pouget A, DeAngelis GC, Angelaki DE. Neural correlates of reliability-based cue weighting during multisensory integration. *Nat Neurosci.* 2011; 15:146–54. [PubMed: 22101645]
101. Paradiso MA. A theory for the use of visual orientation information which exploits the columnar structure of striate cortex. *Biol Cybern.* 1988; 58:35–49. [PubMed: 3345319]
102. Seung HS, Sompolinsky H. Simple models for reading neuronal population codes. *Proc Natl Acad Sci U S A.* 1993; 90:10749–53. [PubMed: 8248166]
103. Carandini M, Heeger DJ. Normalization as a canonical neural computation. *Nat Rev Neurosci.* 2011; 13:51–62. [PubMed: 22108672]
104. Louie K, Gratton LE, Glimcher PW. Reward value-based gain control: divisive normalization in parietal cortex. *J Neurosci.* 2011; 31:10627–39. [PubMed: 21775606]
105. Beck JM, Latham PE, Pouget A. Marginalization in neural circuits with divisive normalization. *J Neurosci.* 2011; 31:15310–9. [PubMed: 22031877]
106. Ohshiro T, Angelaki DE, DeAngelis GC. A normalization model of multisensory integration. *Nat Neurosci.* 2011; 14:775–82. [PubMed: 21552274]
107. Meredith MA, Stein BE. Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Res.* 1986; 365:350–4. [PubMed: 3947999]
108. Carandini M, Heeger D, Movshon J. Linearity and normalization in simple cells of the macaque primary visual cortex. *J Neurosci.* 1997; 17:8621–44. [PubMed: 9334433]
109. Carandini M. From circuits to behavior: a bridge too far? *Nat Neurosci.* 2012; 15:507–509. [PubMed: 22449960]
110. Beauchamp MS, Pasalar S, Ro T. Neural substrates of reliability-weighted visual-tactile multisensory integration. *Front Syst Neurosci.* 2010; 4:25. [PubMed: 20631844]
111. Helbig HB, et al. The neural mechanisms of reliability weighted integration of shape information from vision and touch. *NeuroImage.* 2012; 60:1063–72. [PubMed: 22001262]
112. Takahashi K, et al. Multimodal coding of three-dimensional rotation and translation in area MSTd: comparison of visual and vestibular selectivity. *J Neurosci.* 2007; 27:9742–56. [PubMed: 17804635]
113. Gu Y, DeAngelis GC, Angelaki DE. A functional link between area MSTd and heading perception based on vestibular signals. *Nat Neurosci.* 2007; 10:1038–47. [PubMed: 17618278]
114. Chen A, DeAngelis GC, Angelaki DE. A comparison of vestibular spatiotemporal tuning in macaque parietoinsular vestibular cortex, ventral intraparietal area, and medial superior temporal area. *J Neurosci.* 2011; 31:3082–94. [PubMed: 21414929]
115. Shams L, Kamitani Y, Shimojo S. Illusions. What you see is what you hear. *Nature.* 2000; 408:788. [PubMed: 11130706]
116. Shimojo S, Shams L. Sensory modalities are not separate modalities: plasticity and interactions. *Curr Opin Neurobiol.* 2001; 11:505–9. [PubMed: 11502399]

117. Sekuler R, Sekuler AB, Lau R. Sound alters visual motion perception. *Nature*. 1997; 385:308. [PubMed: 9002513]
118. Avillac M, Denève S, Olivier E, Pouget A, Duhamel J-R. Reference frames for representing visual and tactile locations... supplement. *Nat Neurosci*. 2005; 8:941–949. [PubMed: 15951810]
119. Wallace MT, et al. Unifying multisensory signals across time and space. *Exp Brain Res*. 2004; 158:252–8. [PubMed: 15112119]
120. Fetsch CR, Deangelis GC, Angelaki DE. Visual-vestibular cue integration for heading perception: applications of optimal cue integration theory. *Eur J Neurosci*. 2010; 31:1721–1729. [PubMed: 20584175]

Highlighted references

121. Knill DC, Pouget A. The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci*. 2004; 27:712–9. [PubMed: 15541511] A concise review that provides a good introduction to the idea of sensory uncertainty and the Bayesian perspective on behavior and neural coding, including the incorporation of priors and studies of motor control.
122. Landy MS, Maloney LT, Johnston EB, Young M. Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Res*. 1995; 35:389–412. [PubMed: 7892735] 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, while also introducing a psychophysical procedure that has become a standard method for testing such models.
123. Ernst MO, Banks MS. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*. 2002; 415:429–33. [PubMed: 11807554] . 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, 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.
124. Gu Y, Angelaki DE, DeAngelis GC. Neural correlates of multisensory cue integration in macaque MSTd. *Nat Neurosci*. 2008; 11:1201–10. [PubMed: 18776893] 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 (MSTd) that could underlie the behavior.
125. Meredith M, Stein B. Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J Neurophysiol*. 1986; 56:640–62. [PubMed: 3537225] This paper was 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 (the spatial and temporal principles, inverse effectiveness, etc.).
126. Stanford TR, Quessy S, Stein BE. Evaluating the operations underlying multisensory integration in the cat superior colliculus. *J Neurosci*. 2005; 25:6499–6508. [PubMed: 16014711] This study explored multisensory interactions in the SC in a more systematic fashion than previous work, varying the intensity and timing of visual and auditory stimuli. The results suggested that most SC neurons combine their inputs additively, and that the oft-cited phenomenon of ‘superadditivity’ may only occur for very weak stimuli.
127. Ma WJ, Beck JM, Latham PE, Pouget A. Bayesian inference with probabilistic population codes. *Nat Neurosci*. 2006; 9:1432–8. [PubMed: 17057707] This theoretical study outlined 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 showed that simple summation of population activity in sensory areas can be sufficient to implement Bayes-optimal cue integration.

128. Morgan ML, DeAngelis GC, Angelaki DE. Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron*. 2008; 59:662–73. [PubMed: 18760701] The authors recorded neuronal responses in area MSTd to a wide array of stimulus combinations, finding 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.
129. Ohshiro T, Angelaki DE, DeAngelis GC. A normalization model of multisensory integration. *Nat Neurosci*. 2011; 14:775–82. [PubMed: 21552274] This paper presented a computational model that relies on a widespread neural computation called divisive normalization. Normalization of responses at the level of multisensory integration helps explain several key empirical findings: the reliability-dependent combination rule in MSTd, as well as the ubiquitous empirical principles that were initially described in classic studies of the SC.
130. Fetsch CR, Pouget A, DeAngelis GC, Angelaki DE. Neural correlates of reliability-based cue weighting during multisensory integration. *Nat Neurosci*. 2011; 15:146–54. [PubMed: 22101645] Here the authors recorded from MSTd while monkeys performed a cue-conflict version of the visual-vestibular heading task. Behaviorally, the animals were able to re-weight the cues on a trial-by-trial basis as reliability varied, and neuronal activity accounted well for the behavioral 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 behavior and neural responses.
131. Carandini M, Heeger DJ. Normalization as a canonical neural computation. *Nat Rev Neurosci*. 2011; 13:51–62. [PubMed: 22108672] 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.

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 – borrowed from a simple statistical method in an unrelated

context³¹ – 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 , defined as inverse variance):

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

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

The reliability of the final estimate is greater than the reliability of either single-cue estimate; in fact it is their sum:

$$g_{opt} = \sum_{i=1}^n g_i \quad (2)$$

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' rule,

$$P(X|C_1, C_2) \propto P(C_1|X) * P(C_2|X) * P(X) \quad (3)$$

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 mean corresponding to Eq. 1a and variance corresponding to the inverse of Eq. 2. 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.

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

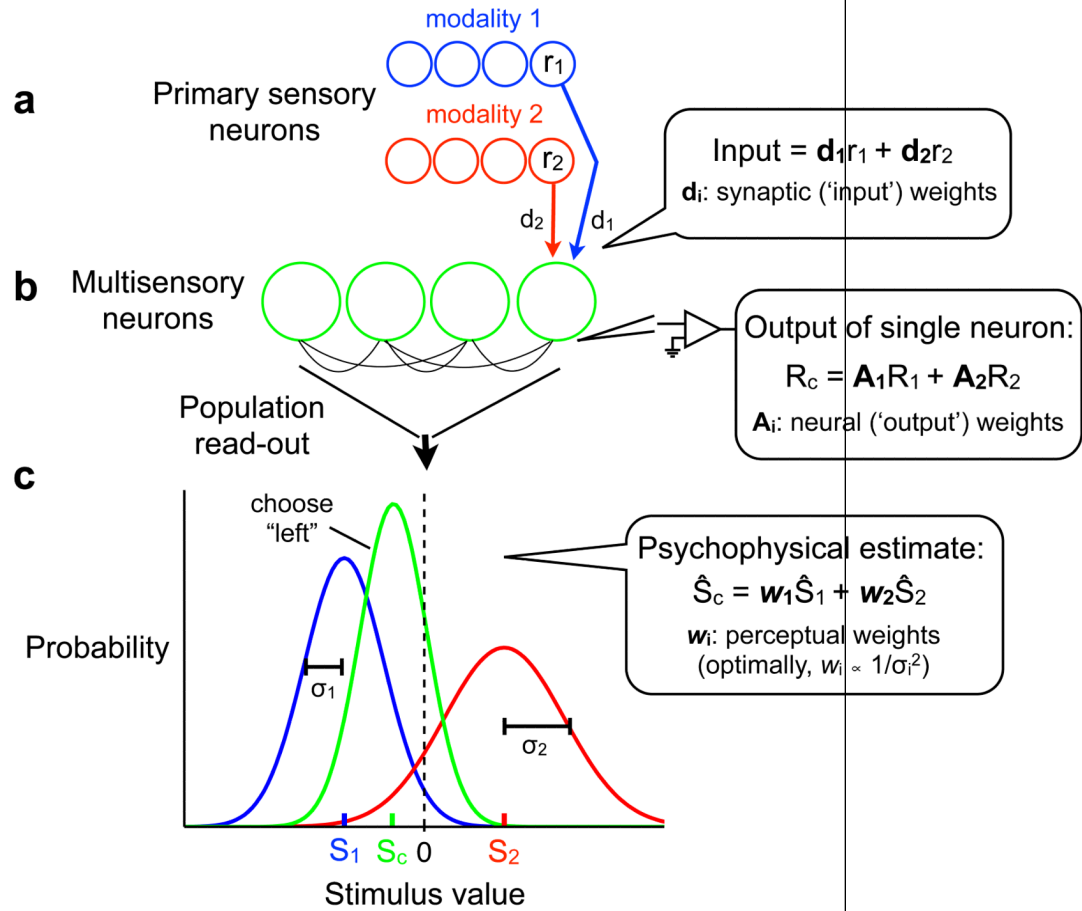
Descriptions of how single neurons combine multiple sensory inputs, such as those provided by Stein and colleagues for the SC^{46, 51, 55} and Morgan et al.⁹⁴ for MSTd (medial superior temporal area, dorsal portion), are often compared to patterns of behavioral cue integration, despite the complex and 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), each receiving unisensory information from different modalities, (e.g., the visual and auditory stimuli in Fig. 1). These signals are transmitted to a multisensory neural population (green) whose activity will generate a particular behavior or perceptual choice.

In the figure, the output of two primary sensory neurons (r_1, r_2 ; 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^{106, 118}). Instead, what we actually measure is the output of a network computation – likely involving lateral and perhaps feedback connections – which generates firing rates denoted R_1, R_2 (the responses to each modality presented in isolation) and R_c (the response to multisensory stimuli; 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 behavioral 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 green 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, or most likely value) as its estimate on a particular trial, leading to a behavioral 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 (i.e., the sigmoidal choice curves in Fig. 1). The optimal perceptual weights are given by the reliability (inverse variance) of the unisensory evidence (Eq. 1), as illustrated here by the shift of the combined (green) distribution toward the more reliable cue (modality 1, blue; note that $\sigma_1 < \sigma_2$ and therefore $w_1 > w_2$ and the observer more often chooses “left”).

With or without the equations and symbols, we hope this exercise makes it clear that the neural weights depicted in panel **b** (i.e., 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 psychophysics⁷. In particular, the relationship between neural weights (A_1, A_2) and perceptual weights (w_1, w_2) is far from straightforward and is the subject of ongoing investigations (see below).



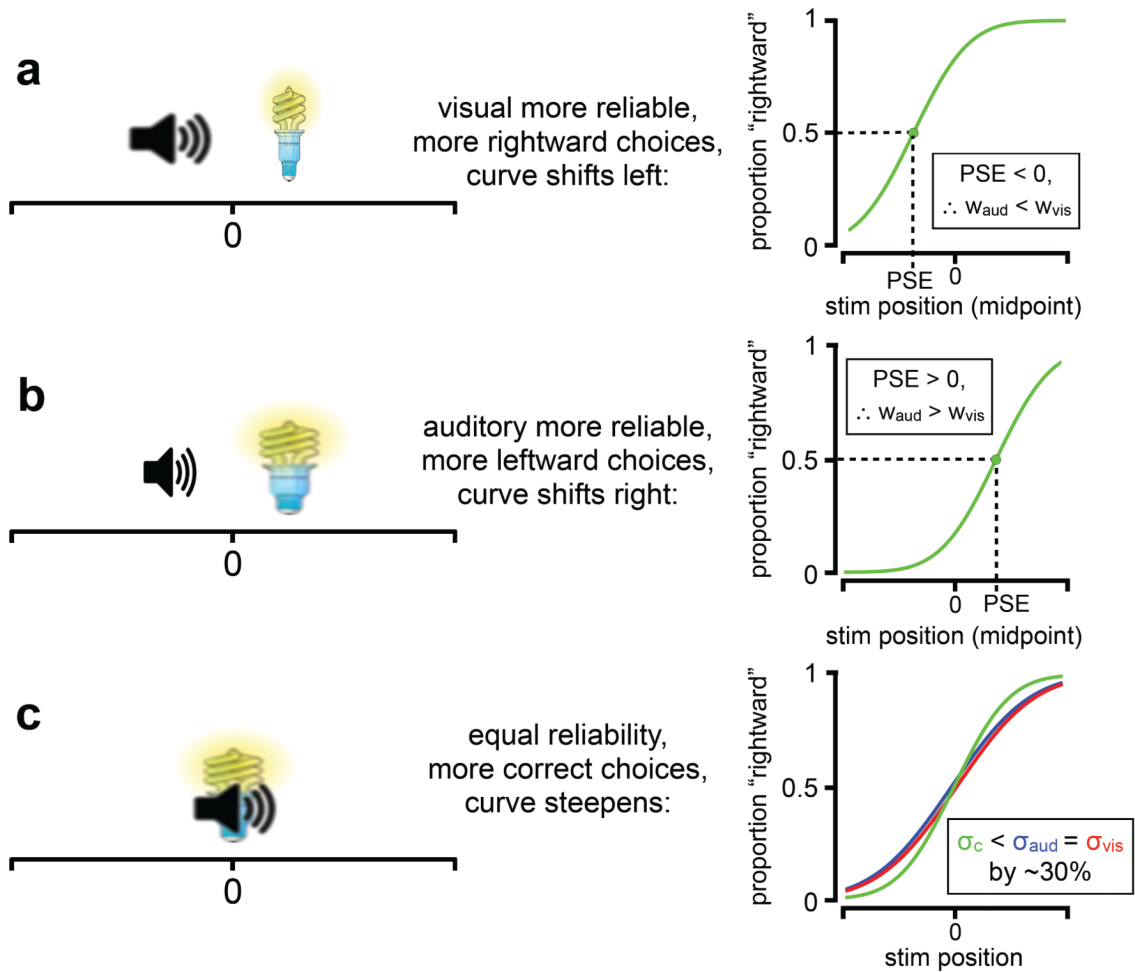


Figure 1. Schematic of a generic cue-integration/cue-conflict psychophysical task

Simplified version of a visual-auditory localization task^{20, 41, 119} 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 flash of light (the visual cue; light bulb icon) and/or a broadband noise burst or click (the auditory cue; speaker icon) 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 (the cue-conflict), and the reliability of one or both cues is often manipulated experimentally, here denoted by the width and blurring of the icons. **a.** Depiction of cue-conflict trials in which the visual cue is more reliable and also displaced to the right, while 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 is jointly moved to the left or right on different trials, generating a sigmoidal choice curve (psychometric function, green, plotted relative to the midpoint between the two stimuli). If the 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 center. The stimulus position at which the curve reaches 50% rightward choices (point of subjective equality, PSE, dashed lines) maps onto a particular set of perceptual weights (w_{aud} and w_{vis} ; the w_i in Eq. 1, Box 1), which in this case would have the relationship $w_{\text{aud}} < w_{\text{vis}}$, since the visual cue is more reliable. **b.** Scenario on a different set of trials with the same cue-conflict but reversed reliability (auditory more reliable than visual). Here the

subject should make more leftward choices, shifting the curve to the right ($w_{\text{aud}} > w_{\text{vis}}$). **c.** In addition to measuring shifts of the psychometric function, performance with combined visual-auditory stimuli (green curve) can be compared to single-cue conditions (red and blue curves), testing the prediction that reliabilities add (Eq. 2 in Box 1; here denoted by a decrease in the standard deviation, σ , of the green cumulative Gaussian psychometric function by a factor of the square root of 2).

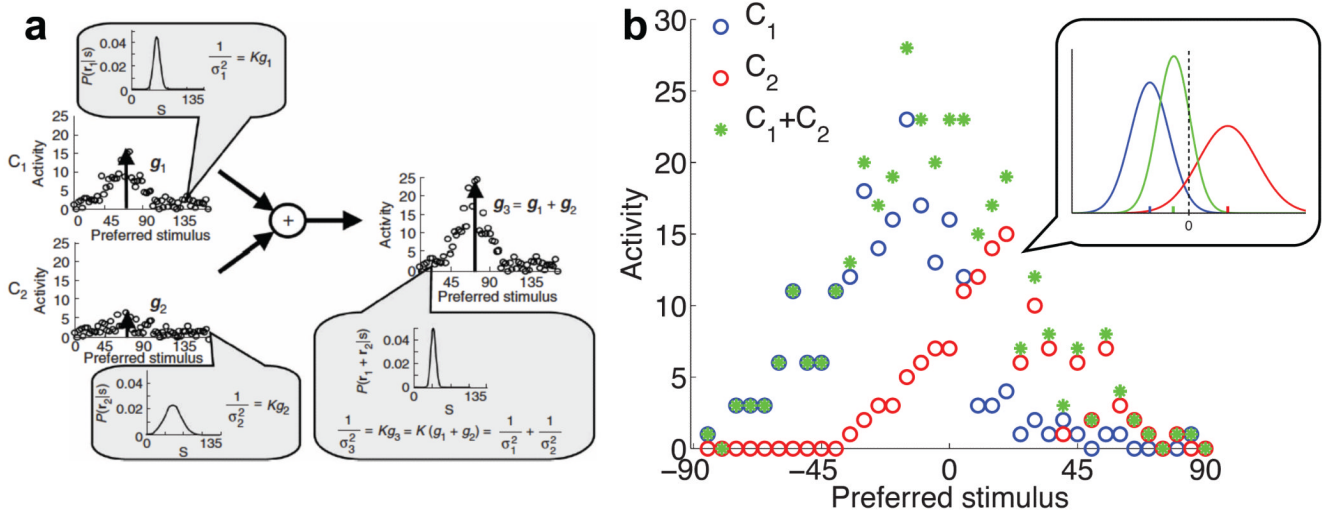


Figure 2. A probabilistic population code (PPC) framework accounts for optimal cue integration by summation of unisensory population activity

a. In this model (reproduced with permission from REF. ⁷⁵), sensory cues C_1 and C_2 each generate a ‘hill’ of population activity in their respective unisensory areas, which could be, for example, regions in visual cortex and auditory cortex. Each data point indicates a single neuron, and these cells are arranged by their preferred stimulus value (e.g., receptive field location). The hills are noisy, not smooth, because of variability in neuronal responses. Owing to the particular kind of variability in these model neurons (also commonly found in real neurons), each hill of activity encodes a conditional probability distribution ($P(r_i|S)$, insets) whose variance is inversely proportional to the gain, or height, of the hill, indicated by the vertical arrows ($\sigma_i^{-2} = K/g_i$; 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 (right side) whose gain is the sum of the unisensory gains g_1 and g_2 . Therefore, the inverse variance of the probability distribution $P(r_1 + r_2|S)$ encoded by the multisensory population is equal to the sum of the individual cues’ inverse variances, or reliabilities – the same operation prescribed by the optimal integration model (Eq. 2 in Box 1). **b.** A simulated cue-conflict trial in which sensory cue C_1 (blue) specifies, in arbitrary units, a stimulus value of -20 and C_2 (red) a 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 (green) is skewed toward the more reliable cue, as shown schematically by the encoded probability distributions (inset). A downstream brain area that optimally decodes this multisensory activity would produce behavioral 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 a variety of tuning widths or shapes, provided that the linear combination is appropriately tailored to these tuning properties⁷⁵.

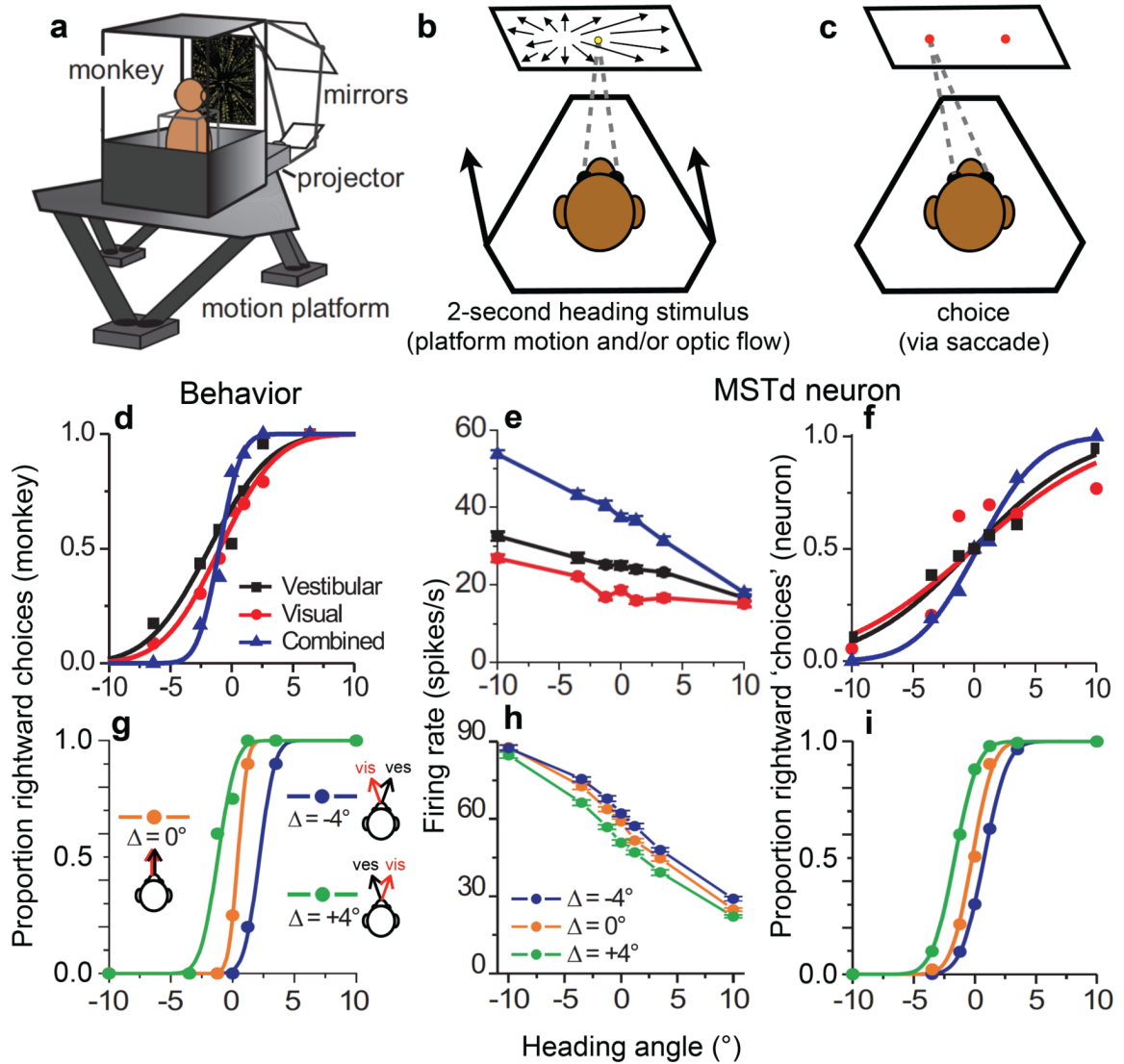


Figure 3. Combined psychophysical and neurophysiological studies of visual-vestibular cue integration in the macaque

a. Monkeys were trained to report their perceived heading (direction of self-motion relative to straight ahead) while seated in a virtual-reality setup²⁶. 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’. Figure modified with permission from REF. ¹²⁰. **b.** While the monkey fixated his gaze (dashed lines) on a spot at the center of the screen (yellow), 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, as shown). **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.** Behavioral 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 indicates leftward. The slope of the fitted curve is a measure of the animal’s sensitivity to small changes in heading, in other words the

reliability of the cue(s). The slope was greater in the combined condition (blue curve, triangles) than in the single-cue conditions (black and red curves), indicating an improvement in sensitivity (i.e., reduction in uncertainty or variance). The average improvement across sessions was close to the optimal prediction (Eq. 2). **e.** The firing rate responses (tuning curves) of a single example neuron from area MSTd are plotted using the same conventions as the behavioral data. Note the steeper slope of the tuning curve in the combined condition (blue, triangles), 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 ROC analysis. The resulting ‘neurometric’ functions quantify the sensitivity of the neuron to small changes in heading during the vestibular (black), visual (red), and combined (blue) conditions. Similar to the behavioral effect, the slope of the neurometric curve is steeper in the combined condition than the single-cue conditions. Panels e and f modified with permission from REF. ¹²⁰**g.** In a separate study¹⁰⁰, 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 ($\theta = +4^\circ$, green curve and symbols) and more leftward choices when the visual heading was displaced to the left ($\theta = -4^\circ$, blue curve and symbols). **h.** Tuning curves from the same neuron as in panel e., 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 ($\theta = -4^\circ$, blue) or fewer spikes ($\theta = +4^\circ$, green) for a given heading angle. **i.** Conversion of these firing rates into neurometric functions reveals a pattern similar to the behavioral result in panel g.; the shift of the curves for different values of θ reflects the trial-by-trial weighting of cues (favoring the more reliable visual cue, as predicted from optimal cue integration). Panels g-i modified with permission from REF. ¹⁰⁰.

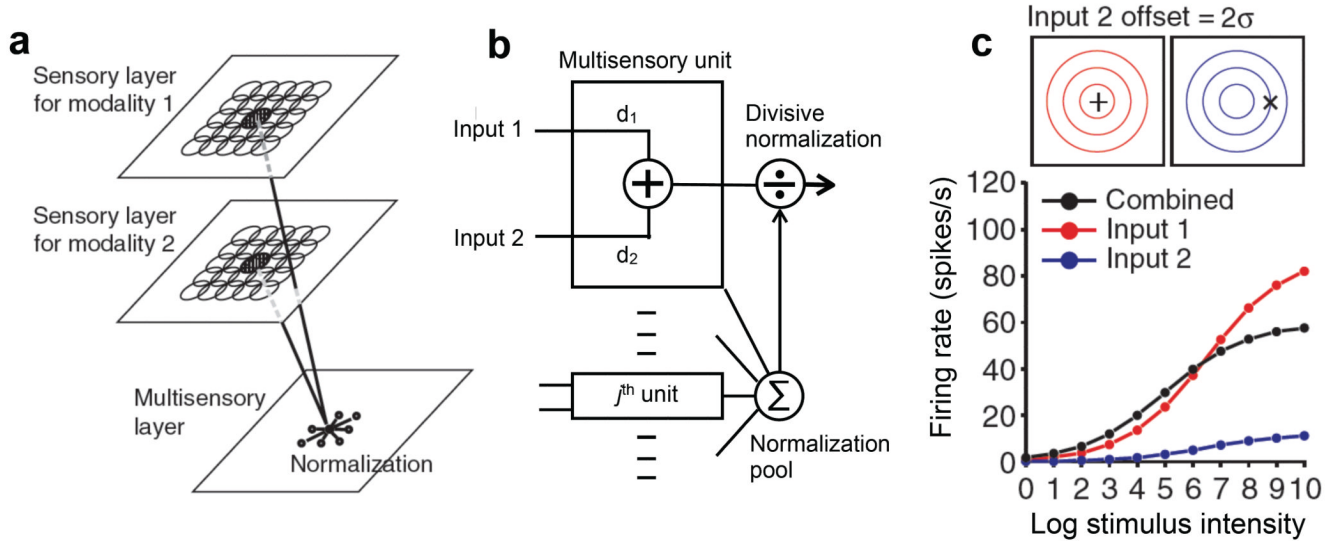


Figure 4. The normalization model of multisensory integration

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 (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 Eq. 4 under “Closing the loop”). **c.** In addition to explaining the origin of the multisensory combination rule in MSTd^{94, 106}, divisive normalization can also 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, illustrated here as a case of cross-modal suppression. One stimulus (‘input 1’, cross in red circles) is presented at the center of the receptive field of a simulated SC neuron, while a second stimulus (‘input 2’, X in blue circles) is presented two standard deviations (2σ) away from center. At relatively high (>7) intensities of the two inputs, the response to the combined inputs (black curve) is less than the response to input 1 alone (i.e., cross-modal suppression), even though input 2 alone is excitatory. This results from the contribution of input 2 to the normalization pool. Panels a-c modified with permission from REF. 106.