Computational principles and models of multisensory integration
Chandramouli Chandrasekaran

Combining information from multiple senses creates robust percepts, speeds up responses, enhances learning, and improves detection, discrimination, and recognition. In this review, I discuss computational models and principles that provide insight into how this process of multisensory integration occurs at the behavioral and neural level. My initial focus is on drift-diffusion and Bayesian models that can predict behavior in multisensory contexts. I then highlight how recent neurophysiological and perturbation experiments provide evidence for a distributed redundant network for multisensory integration. I also emphasize studies which show that task-relevant variables in multisensory contexts are distributed in heterogeneous neural populations. Finally, I describe dimensionality reduction methods and recurrent neural network models that may help decipher heterogeneous neural populations involved in multisensory integration.

Address
Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA

Corresponding author: Chandrasekaran, Chandramouli (mouli@stanford.edu)

Introduction
At every moment, our brains are combining cues from multiple senses to form robust percepts of signals in the world, a process termed multisensory integration [1]. For instance, the percept of flavor involves integration of gustatory and olfactory cues [2,3]. Multisensory integration also enhances our ability to discriminate between sensory stimuli and guides efficient action [1,4]. In noisy auditory environments, visual cues from facial motion are combined with the auditory components of speech to improve comprehension [5–7]. Similarly, when interacting with the world, the motor system integrates visual and proprioceptive feedback to guide limb movements and respond to unexpected perturbations [8,9]. Finally,

multisensory training is more effective for learning tasks and information than unisensory learning [10].

These examples highlight the importance of multisensory integration for everyday life. However, the effortless nature of multisensory integration hides the underlying complexity of this process [11–17]. Sensory signals vary in intensity, have different noise profiles, are delayed relative to each other [5,8,18], and exist in different reference frames [9,19]. Seminal experiments in the anesthetized cat in combination with recent experiments in many other species have established several fundamental principles of multisensory integration [Box 1, [20]]. However, recent studies suggest that we need to go beyond these principles [21,22]. The first goal of this review is to highlight new efforts that are facilitating the rapid development of newer computational frameworks to understand behavioral and neural multisensory integration. Another goal is to suggest that dimensionality reduction techniques and recurrent neural network (RNN) models might help us better understand neuronal responses in multisensory brain areas and thus facilitate the development of newer computational principles underlying multisensory integration.

Bayesian decision theory and the drift-diffusion model can explain multisensory behavior
Multiple sensory systems provide independent estimates that can be combined to reduce noise and uncertainty thereby improving precision [1]. An improvement in the precision of the percept, in turn, improves behavior. For instance, many species, including fish, discriminate and recognize multisensory stimuli better than unisensory stimuli [23•,24–27] and appear to combine sensory cues in an optimal (or near-optimal) manner [28]. Near-optimal behavior in this context means that observers consider the variation in cue reliability on a trial-by-trial basis to maximize their behavioral accuracy. Bayesian decision theory models near-optimal discrimination and other behaviors such as identifying the underlying cause of events in the world [29,30••,31] as a statistical inference process performed by the brain [32] (Figure 1a). Bayesian models assume that probability distributions of prior knowledge and sensory information (likelihood) are known [1,24,33]. The Bayesian framework's success in explaining behavior has inspired the development of neural network models that can represent probability distributions [34] and perform near-optimal Bayesian inference as well as other important probabilistic

Box 1 Seminal Principles of Multisensory Integration

Our understanding of multisensory integration has largely relied on seminal experiments in the anesthetized cat and in recent years from both anesthetized and awake experiments in many other species [20]. These experiments established the fundamental principles of superadditivity, the principle of inverse effectiveness and the temporal window of integration. These principles are 'word level' models to understand and coarsely describe experimental results. They are not formal computational models that can quantitatively fit behavioral or neural data.

Superadditivity, subadditivity, and additivity: These refer to the relationship between responses (usually firing rate of neurons, but can be any neural response) to multisensory stimuli and the unisensory responses. As an example, assume that the sensory signals are visual and auditory. The response to the visual stimulus is V, the response to the auditory stimulus is A, and the audiovisual response is AV. If $AV > A + V$, the response is superadditive. If $AV < A + V$, then the response is subadditive. If $AV \sim A + V$, then the response is additive. Neuronal responses are largely in the additive regime. Only for very strong or very weak stimuli do the responses become sub-additive or superadditive. The observation that neurons largely operate in an additive (or linear) regime is consistent with modeling studies, which show that an additive mechanism can explain behavior in multisensory tasks [46,48••].

The law of inverse effectiveness: The law of inverse effectiveness states that multisensory stimuli are more likely or effectively integrated when the best unisensory response is relatively weak. Again let us assume A and V are the unisensory responses. The law of inverse effectiveness states that the benefit from multisensory integration: AV — (best response for (A, V)) is maximal when this best response is weak. Empirically, this manifests most often when the intensities of the unisensory stimuli are very weak. This principle was originally developed to explain multisensory responses in neurons and was considered to be relevant for behavior as well. However, recent studies suggest that behavior may depart from this principle [7,36].

Temporal window of integration: The temporal window of integration is a time window over which multisensory stimuli are more likely or effectively integrated in behavior (e.g., improvement in accuracy, reduction in RT, so on). The principle suggests that maximal integration occurs when sensory stimuli occur approximately simultaneously. The assumption is that this behaviorally defined temporal window of integration is a result of the finite time over which neurons integrate sensory inputs. However, the size of this temporal window of integration varies across studies, task contexts, stimulus type, and stimulus intensities. Computational models are attempting to formalize and precisely define this window of integration [91].

computations [17,35]. Efforts using machine learning have also described learning rules and architectures to learn near-optimal behaviors [19].

Other computational models aim to describe the multisensory reaction time (RT) benefits observed in redundant signal detection paradigms where two (or more) sensory stimuli (often from the auditory and visual modalities) are presented simultaneously or with a short delay in between them [18,36–40]. Faster multisensory compared to unisensory RTs in these tasks can emerge from two candidate mechanisms. The first mechanism, the race model, argues that the RT benefits emerge due

to 'statistical facilitation' and not multisensory integration [41–43]. Statistical facilitation refers to the mathematical property that the mean of the minimum of two distributions is less than the mean of the individual distributions. Concretely, if A and V are two random variables denoting RT distributions, then the mean of the distribution formed by min (A, V) is less than the *minimum of the means* of A and V distributions. Thus, even without explicitly combining sensory cues, it is statistically possible to obtain faster RTs for multisensory compared to unisensory stimuli. When appropriate care is taken to arrange experimental conditions, an inequality based statistical test can be used to reject the race model [44,45•].
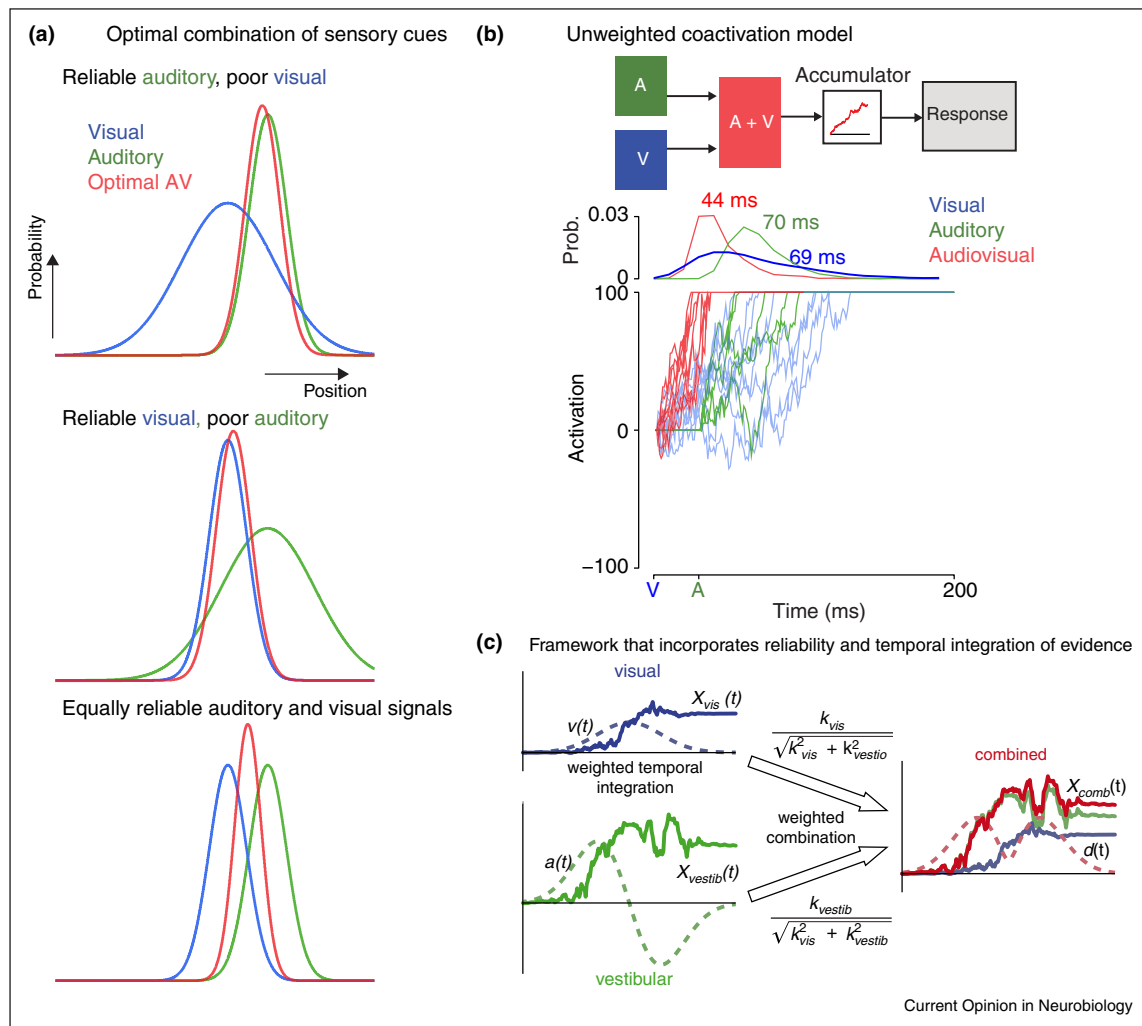
Typically, after rejecting the race model, 'coactivation' models (Figure 1b), which assume that both components of the redundant signal influence responses on a single trial are used to model RT. The simplest coactivation models assume that activation from sensory inputs are summed and integrated over time according to a drift-diffusion process to a response criterion ([46,47], Figure 1b). These models explain multisensory RTs in detection tasks across a large range of stimulus delays [46,47] and can model behavior in choice and go/no-go tasks [39].

One issue with simple coactivation models is that they combine sensory inputs without weighting them according to their reliability — a clearly sub-optimal strategy when the reliability of sensory inputs varies widely across contexts as well as a function of time. A recent study addressed this issue by fusing the coactivation framework and the principles of Bayesian cue combination to explain behavior in multisensory RT discrimination tasks [48••]. The model weights the input from different sensory cues using their reliability at every moment during the integration of sensory evidence (Figure 1c) and explains both accuracy as well as RT of human participants in multisensory heading discrimination tasks. Extensions to this model will likely explain discrimination behavior across both a range of stimulus difficulties as well as delays between time-varying sensory signals.

A distributed network with multiple redundant pathways mediates multisensory integration

The Bayesian framework's success in explaining behavior has inspired studies to examine the relationship between its core principles and neuronal responses in different brain regions during multisensory discrimination tasks [12,15]. Monkeys, when trained to report their heading direction, also near-optimally combine visual (optic flow), and vestibular cues [27]. They are also sensitive to changes in the reliability of these cues on a trial-by-trial basis [49]. Consistent with this near-optimal behavior, a subpopulation of medial superior temporal area (MSTd) neurons with congruent tuning to visual and vestibular

Figure 1

(a) Optimal combination of sensory cues

Reliable auditory, poor visual

Visual
Auditory
Optimal AV

(b) Unweighted coactivation model

Current Opinion in Neurobiology

Bayesian frameworks and coactivation models to understand multisensory behavior. **(a)** Near-optimal cue (or stimulus) combination according to the Bayesian framework to improve discrimination behavior is well understood with an example (adapted from [92]). Consider a cat seeking a mouse with both visual and auditory cues. The curves in the figure show the hypothetical probability distributions of the mouse's position as estimated by the cat's brain for the three different modalities (Blue-visual, green-auditory, red-audiovisual). Assume that it is dark, and the mouse is in an environment with many gray rocks roughly the size and shape of a mouse. In this context, the optimal Bayesian cat would use auditory cues to estimate the mouse's location (top). By contrast, when it is sunny, to optimize its discrimination behavior the cat would largely rely on visual cues to locate the mouse (middle). One can readily imagine many intermediate scenarios where the optimal strategy for the cat is to combine both visual and auditory cues to have the best chance of catching the mouse (bottom). For the case that the cat's auditory and visually guided estimates of the mouse's position are Gaussian, *independent* and unbiased (means: s_a, s_v and standard deviations σ_a, σ_v), the optimal estimate of the position of the mouse is the weighted average of the auditory and visually guided estimates ($S_{av} = W_a S_a + W_v S_v$). The weights are the normalized reliability of each cue $\left(W_a = \frac{1/\sigma_a^2}{(1/\sigma_a^2)+(1/\sigma_2^2)}, W_v = \frac{1/\sigma_v^2}{(1/\sigma_a^2)+(1/\sigma_2^2)} \right)$. Behavior in multisensory discrimination experiments can be tested to see if they are consistent with this prediction from the optimal framework. **(b)** An example of a simple coactivation model used to explain detection behavior. Top panel, a cartoon of the linear summation coactivation model typically used to explain multisensory detection behavior [46,47]. Auditory and visual inputs are linearly summed to arrive at a new drift-rate, and this undergoes the drift-diffusion process to the criterion to trigger a response. Bottom panel, simulations from the coactivation model of a few trials of an audiovisual stimulus where a visual cue turns on at $t = 0$ and an auditory cue turns on at $t = 30$ ms. The visual and the auditory stimuli were assumed to be modest in intensity. In this hypothetical integrator, the onset of the visual stimulus before the onset of the auditory stimulus results in an increase in activity. The auditory stimulus can build on this activity, and this results in the criterion being reached faster on average for the audiovisual (44 ms) compared to both auditory (~70 ms RT when measured relative to the visual stimulus onset) and visual-only stimuli (~69 ms). Blue lines denote the visual cue. The green lines denote the auditory cue. Red lines denote the audiovisual cue. **(c)** A framework that combined the insights from both A and B to develop an optimal coactivation drift-diffusion model to explain multisensory discrimination behavior [48••]. This model was developed in the context of a heading discrimination task using both visual and vestibular cues. The key innovation in this model is that it integrates evidence in an optimal manner by factoring in the model both the time course of sensory signals as well as the reliability of the signals. The visual and vestibular cues are time varying signals, whose reliabilities change as a function of time. They also can have different reliabilities depending on the context (as in **a**).

cues also combines these cues near-optimally [15]. Association areas such as the frontal eye fields (FEF, [50]) and ventral intraparietal area (VIP, [51]) are also likely to be relevant for heading discrimination. The strongest signals that correlate with the animal's behavioral choice are found in VIP [51].

The neurophysiological studies discussed thus far suggest a feed forward hierarchical process for multisensory integration where sensory cues are processed in unisensory areas (visual or auditory areas) and then subsequently processed in high-level association areas (e.g., parietal or prefrontal cortex) to result in multisensory behavior [13,52]. This feedforward hierarchical view predicts that timing of responses should follow a systematic order from unisensory to high-level association areas and inactivation of high-level areas such as the parietal cortex should result in profound multisensory behavioral deficits. However, simultaneous recordings in several cortical structures reveal that the simple feed forward view provides a poor account of the timing of responses across areas even for flexible unisensory (visually guided) decision-making [53••]. Inactivating VIP had minimal impact on the behavior in the multisensory heading discrimination task [54•] — a result consistent with the mild behavioral consequences of inactivating lateral intraparietal area (LIP) of monkeys performing a unisensory motion discrimination task [55•]. These perturbation studies suggest that multiple redundant pathways, which involve other areas besides the parietal lobe, are likely involved in multisensory integration (and decision-making). The perturbation results also imply that choice-related activity in an area may not always imply a causal role in discrimination behavior and highlight the need for a better understanding of the factors that lead to neural responses in an area to covary with choice [56–58]. Together, the recent results suggest that the feed forward view is insufficient to explain multisensory integration [13].

The feedforward view of multisensory integration (and other behaviors) is also *inconsistent* with connectivity studies and neurophysiological experiments that demonstrate multisensory effects in primary sensory areas. Studies show anatomical connections between putatively unisensory areas [59,60] and neurophysiological studies demonstrate that sensory inputs from other modalities (e.g., visual cues) modulate firing rates (FRs) in classically unisensory cortices [61–64] (e.g., auditory cortex). For example, in monkeys performing an audiovisual redundant signals detection task, the motion of the mouth reduces the response latency of auditory cortical spiking responses to vocalizations [64]. These trial-by-trial changes in auditory cortical response latencies modestly covary with RT and thus potentially contribute to the RT benefits observed in multisensory detection tasks [18,36–40]. Similarly, using multisite electrophysiology and optogenetic inactivation, a recent study showed that projections from gustatory cortex modulate neural responses in the primary olfactory cortex, even in the absence of a taste stimulus [3]. Finally, a groundbreaking study combining electron microscopy, electrophysiology, and behavior demonstrated that a multilevel multisensory circuit contributes to the escape response in larval Drosophila [65••]. All of these studies bolster the view that multisensory integration involves a distributed network with multiple redundant pathways [13,66].

Future neurophysiological studies using multisensory RT detection/discrimination paradigms will allow further quantitative tests of behavioral frameworks (such as the drift-diffusion and Bayesian models) and improve our understanding of this distributed network [67]. Promising candidates for mediating RT and accuracy benefits typically observed for audiovisual integration (often observed for speech and vocalizations) that need further investigation in the monkey temporal lobe include the belt auditory cortex [64,68•] and the upper bank of the superior temporal sulcus [69,70]. The ventrolateral prefrontal cortex [71,72], and the dorsal premotor cortex [73] are two other association areas that need to be further studied. Another important future line of research will be to expand our understanding of the role of oscillations in these brain areas and how they relate to multisensory effects observed in FRs and behavior [74].

Dimensionality reduction and the dynamical systems approach might help achieve a population-level understanding of multisensory processing

The neurophysiological studies describing the activity of single neurons are crucial for establishing that an area is multisensory and to perhaps derive fundamental principles of multisensory integration [20]. If we are to advance beyond these coarse descriptions and derive more insights into the mechanisms underlying multisensory integration, we need to better understand the temporal patterns of neuronal FRs in these multisensory brain areas. The problem is that a closer examination of neuronal FRs in multisensory brain regions

(Figure 1 Legend Continued) Simply adding as in **b** is suboptimal. For instance, you would just be adding noise at the start when the visual signal is low. $X_{vis}(t)$ = integrated evidence for the visual cue (optic flow), $X_{vest}(t)$ is the integrated evidence for the vestibular cue. k_{vis} is a constant that signifies the strength of the visual signal. k_{vestib} is a constant that signifies the strength of the vestibular signal. The combined signal $X_{comb}(t)$ is the reliability-weighted sum of these two signals (weights are shown on top of the arrows). The dashed lines on the left-hand side of the figure denote the velocity and acceleration profiles of the signals. The velocity profile was Gaussian. Both visual and vestibular cues were presented congruently and were temporally synchronized. In this study, momentary evidence for vestibular input is assumed to result from acceleration ($a(t)$). The momentary evidence for visual input is thought to result from velocity cues ($v(t)$).

suggests a great deal of response heterogeneity — a phenomenon that is increasingly obseved in numerous brain regions and many other tasks [75,76•]. The consensus from several experiments is that single neuron responses in both unisensory and multisensory brain regions are often perplexing, temporally complex and can be counterintuitive to expectations [75,77]. For example, neural responses in multisensory brain regions can be additive, sub-additive, or super-additive [20,78]. Similarly, neurons in MSTd [27], VIP [51], and FEF [50] can show either congruent or opposite tuning to the visual and vestibular cues during the heading discrimination task. The role of these opposite tuning neurons is still mysterious.

How can one make sense of this variation and heterogeneity in neural responses and better understand the population level computations involved in multisensory integration? One recent approach is to be less concerned about single neuron responses and instead, focus on a population-level description of neural responses [79]. In this approach, neuronal FRs are considered noisy signatures of an underlying dynamical system implemented by the network in a brain region. The hallmark of this approach is the use of dimensionality reduction techniques (e.g., principal component analysis) to visualize and describe firing rate variance in lower dimensional subspaces. The goal of the approach is to understand the 'dynamics' of this lower dimensional population neural 'state,' (how it evolves through time) and the relationship of the dynamics to various features of the stimulus and the behavior [75,76•,80,81,82•].

When carefully applied, the dynamical systems approach has minimal assumptions and also involves minimal sorting of neurons into categories. This approach has helped understand population responses in the motor cortex during reaching [80], prefrontal cortex (PFC) during context-dependent decision-making [75], and medial frontal cortex during value-based decisions [83]. Dimensionality reduction and decoding techniques also allow the study of relationships between neural population activity and behavior on single trials [84,85,86•]. Single-trial analysis may help us better understand how principles proposed by the Bayesian frameworks are instantiated in neural populations [17,35].

A recent study applied this population-level approach to understand the neuronal responses in posterior parietal cortex (PPC) of rats performing an audiovisual discrimination task [87••]. Dimensionality reduction showed that PPC neurons are poorly described by functional subcategories (e.g., choice selective, modality-selective, and movement selective neurons). Instead, task parameters such as choice, modality, and temporal response features are distributed in a random, intermixed manner across neurons. Even though these features are intermixed, the

population response can be used to read out information about different task parameters [87••]. Dimensionality reduction suggested that different dimensions were explored during the decision-formation and movement epochs. This result suggests that the PPC network dynamically reorganized to support the evolving demands of decision-making.

Optimized recurrent neural network models provide a candidate framework to model the heterogeneous temporally complex FRs in multisensory brain regions

Using dimensionality reduction methods to analyze FRs is only part of the process for achieving a deeper understanding of multisensory integration. To derive hypotheses about the underlying dynamics in different brain areas, we also need to develop neural networks capable of modeling these heterogeneous, temporally complex neuronal responses. This need becomes even more pressing as we begin to use novel multisensory tasks and newer neural population recording methods that will provide deeper access to the rich FR dynamics involved in multisensory integration. In these settings, the standard approach of hand building neural network models to incorporate a certain mechanism (say an integration operation) and then examining and drawing parallels between the FRs of neurons within the network and neurophysiological responses is insufficient. FR dynamics of neurons in hand-built neural network architectures are often very simple, scaled versions of a prototypical time series (e.g., an integrator). Handcrafted neural network architectures are thus insufficient to describe the complexity we observe in physiological responses.

To bridge the gap between neural network models and physiological data, we need neural network models whose constituent units demonstrate temporally complex responses. The networks should also display rich dynamical behavior at the population level, and task relevant information should be distributed in the population. The optimized recurrent neural network (RNN, Figure 2a) is one candidate framework that possesses these desired properties. Like animals trained to perform a behavioral task, RNNs can be trained through an optimization process to transform an input (e.g., a sensory input) into the desired output (e.g., a category) [88]. The advantage of the optimized RNN approach is that the network is only instructed on *what* it should do and unlike the standard hand-built network model there is no instruction on *how* the network has to do it. Analyzing the network dynamics, in turn, provides a way to generate *quantitative* hypotheses about the dynamical mechanisms that may be implemented in real neural populations to solve the behavioral task. Various constraints can then iteratively be built into these networks and then examined to systematically match the dynamics of the RNN and the neurophysiological data. RNNs have helped demystify neural responses in the

Figure 2
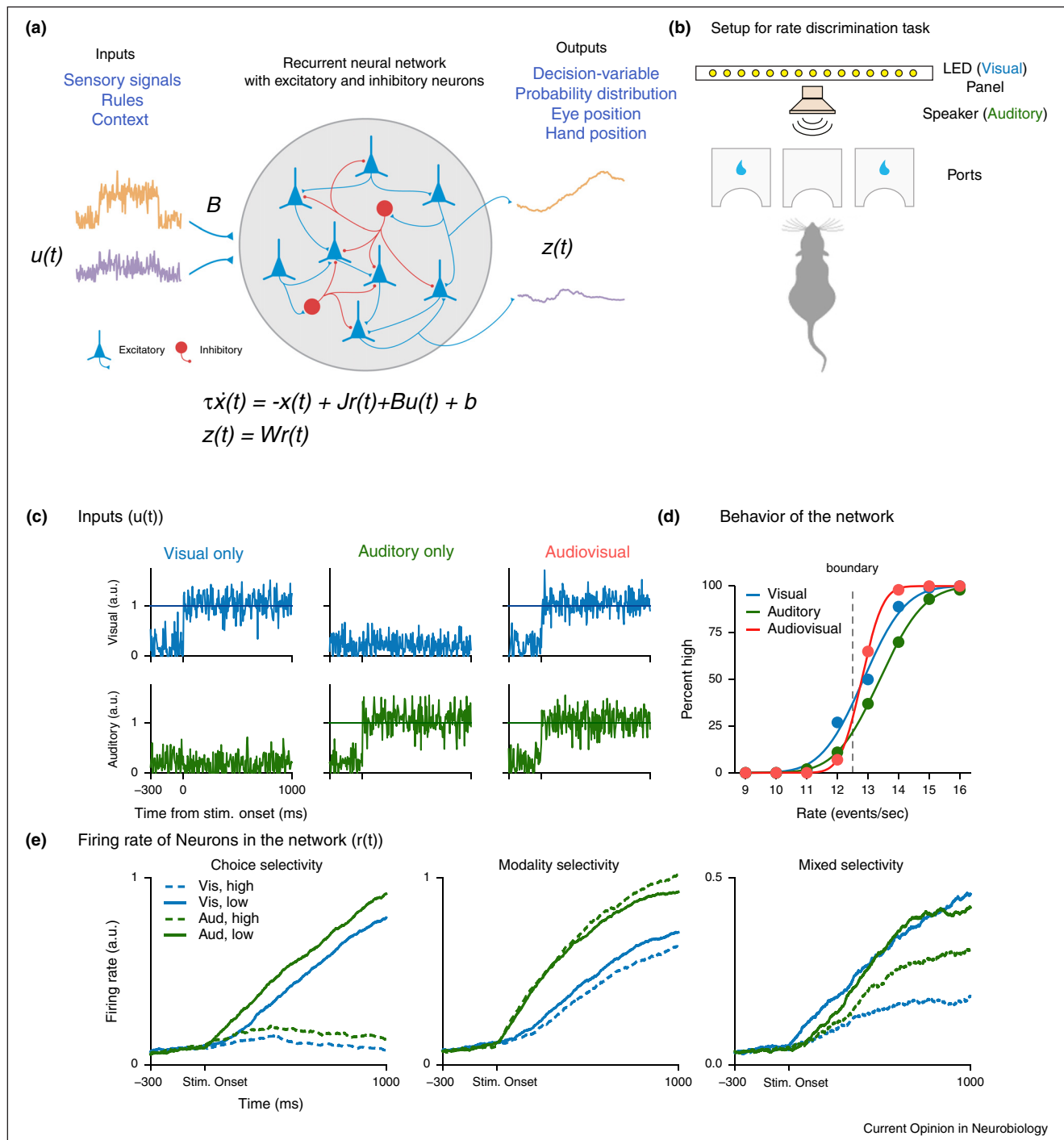
Simulations from the RNN model developed in Song *et al.* [90] to solve the multisensory integration task reported in Raposo *et al.* [87**]. **(a)** A schematic and dynamical equation for a nonlinear RNN. $x(t)$ is a vector with the ith component of this vector describing the summed and filtered synaptic current input for the ith neuron (as in a biological neuron). The continuous variable $r(t)$ is a vector and describes the FR of neurons in the network and are obtained through a nonlinear transform of $x(t)$, typically a saturating nonlinearity or a rectified linear function. The defining feature of these RNNs is the recurrent feedback from one neuron in the network to another. The matrix J parameterizes the connection strength between the units. The network receives external input through the $u(t)$ term weighted by a vector B; every neuron also receives a bias input b_i. τ is a time constant that sets the time scale of the network. The outputs of the network, $z(t)$, are usually obtained by a linear readout operation. Each node in this network is a neuron that receives external input as well as recurrent feedback (through J). Inputs ($u(t)$) can be sensory signals, rules, or context signals. The outputs ($z(t)$), which are obtained by a weighted readout of the firing rates of the neurons in the network can be binary choice [75], continuous decision-variables, probability distributions, or behavioral signals such as hand position/eye position[90]/electromyography

PFC during context-dependent decisions [75] and the dynamical structure in population FRs in motor cortex during reaching [89].

To illustrate this approach, I recapitulate here a network capable of multisensory integration developed as part of a recent methods study aimed at providing a simple, accessible and flexible framework to train and test RNNs [90] (Figure 2a). The network was provided noisy visual and auditory inputs and trained to accomplish the same rate discrimination task performed by the rats in [[87••], Figure 2b]. If the input rate was higher than the standard, then the network produced a high output and vice versa. The network reproduces the behavioral improvement shown by the rats for the multisensory compared to the unisensory stimuli (Figure 2c,d). Given that the network solves this task and mimics the behavior of the animals, we can examine the types of neuronal responses in the network and compare them to the physiological responses observed in [87••]. The network contains neurons tuned to choice and modality. One other key feature of this network is that the FRs of some of the neurons in the network also show the phenomenon of mixed selectivity observed in FRs of the neurons recorded in the PPC of the rat [Figure 2e, [87••]]. More work is necessary to understand the constraints needed to match the population dynamics in the network to that of the network dynamics recorded in the rat PPC [87••].

In the future, several other relevant questions can be asked of RNNs developed for multisensory tasks. For instance, at the behavioral level, do RNNs also benefit from multisensory compared to unisensory learning [10]? At the physiological level, RNNs could be trained to perform the heading discrimination task and then examined to see if they also naturally find solutions with neuronal populations that show both congruent and opposite tuning [27]. These networks can also be examined to see if its constituent neurons also show the well-established neurophysiological observations of superadditive, subadditive and additive multisensory responses in multisensory brain regions [20]. Creating virtual lesions of

the opposite tuning neurons or the superadditive neurons in the network may help clarify their role in multisensory integration. Finally, combining the dynamical features of these RNNs with principles of probabilistic computation [35] could help develop new probabilistic, dynamical models for multisensory integration.

Conflicts of interest statement
Nothing declared.

Acknowledgments

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

• of special interest
•• of outstanding interest

1. Ernst MO, Bulthoff HH: **Merging the senses into a robust percept**. *Trends Cogn Sci* 2004, **8**:162-169.

2. Spence C: **Multisensory flavor perception**. *Cell* 2015, **161**:24-35.

3. Maier JX, Blankenship ML, Li JX, Katz DB: **A multisensory network for olfactory processing**. *Curr Biol* 2015, **25**:2642-2650.

4. Alais D, Newell FN, Mamassian P: **Multisensory processing in review: from physiology to behaviour**. *Seeing Perceiving* 2010, **23**:3-38.

5. Chandrasekaran C, Trubanova A, Stillittano S, Caplier A, Ghazanfar AA: **The natural statistics of audiovisual speech**. *PLoS Comput Biol* 2009, **5**:e1000436.

6. Sumby WH, Pollack I: **Visual contribution to speech intelligibility in noise**. *J Acoust Soc Am* 1954, **26**:212-215.

7. Ross LA, Saint-Amour D, Leavitt VM, Javitt DC, Foxe JJ: **Do you see what i am saying? Exploring visual enhancement of speech comprehension in noisy environments**. *Cereb Cortex* 2007, **17**:1147-1153.

(Figure 2 Legend Continued) responses [89]. The RNN cartoon is adapted from Figure 1 of [90]. **(b)** Schematic of the behavioral apparatus for the multisensory rate discrimination task for the rats in [87••] (redrawn based on Figure 1a in [87••]). The Stimuli were 1s long auditory and/or visual streams delivered either through a speaker or from a LED panel. The rats discriminated whether the presented rate was lower (move to left port) or higher (move to right port) than a decision boundary (12.5 events/sec). The rat cartoon is a recolored version of the one in Figure 1a of [87••]. **(c)** Inputs to a model network trained to solve the audiovisual integration task from [87••]. The network was trained to perform the same task and provided with both positive and negatively tuned visual and auditory inputs ($u(t)$, positive inputs are shown here). The RNN consisted of 150 neurons (120 Excitatory and 30 inhibitory) and used a rectified linear current ($x(t)$) to firing rate ($r(t)$) function. The network was trained to hold a high output value if the input was above the decision boundary (12.5 events/sec) and low if the input was below this decision-boundary. The results shown were obtained from the code provided in [90] (https://github.com/frsong/pycog). **(d)** The RNN solves the task and shows a benefit for multisensory compared to unisensory stimuli and thus demonstrates behavior similar to the rats in the original study [87••]. The psychometric functions show the percentage of high choices made by the network as a function of the event rate for the uni-sensory and multisensory trials. The smooth lines are cumulative Gaussian fits to the psychometric function. **(e)** FR ($r(t)$) of selected simulated neurons in the RNN aligned to the stimulus onset during the period of sensory stimulation and decision-formation. In particular, some neuronal FRs show the main effect of choice (left panel); FRs of other neurons show a main effect for modality (middle panel). Also as in the real data recorded in the posterior parietal cortex [87••], neurons demonstrate FRs best described by an interaction between choice and modality tuning (right panel). High and low denote the choices. Vis-high denotes that the rats chose high for the visual input for example and so on.

8. Cluff T, Crevecoeur F, Scott SH: **A perspective on multisensory integration and rapid perturbation responses**. *Vision Res* 2015, **110**:215-222.

9. Sabes PN: **Sensory integration for reaching: models of optimality in the context of behavior and the underlying neural circuits**. *Prog Brain Res* 2011, **191**:195-209.

10. Shams L, Seitz AR: **Benefits of multisensory learning**. *Trends Cogn Sci* 2008, **12**:411-417.

11. Angelaki DE, Gu Y, DeAngelis GC: **Multisensory integration: psychophysics, neurophysiology, and computation**. *Curr Opin Neurobiol* 2009, **19**:452-458.

12. Seilheimer RL, Rosenberg A, Angelaki DE: **Models and processes of multisensory cue combination**. *Curr Opin Neurobiol* 2014, **25**:38-46.

13. Bizley JK, Jones GP, Town SM: **Where are multisensory signals combined for perceptual decision-making?** *Curr Opin Neurobiol* 2016, **40**:31-37.

14. Arnal LH, Giraud AL: **Cortical oscillations and sensory predictions**. *Trends Cogn Sci* 2012, **16**:390-398.

15. Fetsch CR, DeAngelis GC, Angelaki DE: **Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons**. *Nat Rev Neurosci* 2013, **14**:429-442.

16. Stevenson RA, Ghose D, Fister JK, Sarko DK, Altieri NA, Nidiffer AR, Kurela LR, Siemann JK, James TW, Wallace MT: **Identifying and quantifying multisensory integration: a tutorial review**. *Brain Topogr* 2014, **27**:707-730.

17. Ma WJ, Pouget A: **Linking neurons to behavior in multisensory perception: a computational review**. *Brain Res* 2008, **1242**:4-12.

18. Miller J: **Timecourse of coactivation in bimodal divided attention**. *Percept Psychophys* 1986, **40**:331-343.

19. Makin JG, Fellows MR, Sabes PN: **Learning multisensory integration and coordinate transformation via density estimation**. *PLoS Comput Biol* 2013, **9**:e1003035.

20. Stein BE, Stanford TR: **Multisensory integration: current issues from the perspective of the single neuron**. *Nat Rev Neurosci* 2008, **9**:255-266.

21. Holmes NP: **The principle of inverse effectiveness in multisensory integration: some statistical considerations**. *Brain Topogr* 2009, **21**:168-176.

22. Stanford TR, Quessy S, Stein BE: **Evaluating the operations underlying multisensory integration in the cat superior colliculus**. *J Neurosci* 2005, **25**:6499-6508.

23. Schumacher S, Burt de Perera T, Thenert J, von der Emde G:
• **Cross-modal object recognition and dynamic weighting of sensory inputs in a fish**. *Proc Natl Acad Sci* 2016, **113**:7638-7643.
This study showed that weakly electric fish can weight sensory inputs (visual and active electrical sense) dynamically according to their reliability to improve their object recognition capabilities. This observation in fish suggests that this ability is not unique to mammals and does not require a cerebral cortex.

24. Ernst MO, Banks MS: **Humans integrate visual and haptic information in a statistically optimal fashion**. *Nature* 2002, **415**:429-433.

25. Alais D, Burr D: **The ventriloquist effect results from near-optimal bimodal integration**. *Curr Biol* 2004, **14**:257-262.

26. Raposo D, Sheppard JP, Schrater PR, Churchland AK: **Multisensory decision-making in rats and humans**. *J Neurosci* 2012, **32**:3726-3735.

27. Gu Y, Angelaki DE, DeAngelis GC: **Neural correlates of multisensory cue integration in macaque MSTd**. *Nat Neurosci* 2008, **11**:1201-1210.

28. Ma WJ: **Organizing probabilistic models of perception**. *Trends Cogn Sci* 2012, **16**:511-518.

29. Kording KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L: **Causal inference in multisensory perception**. *PLoS One* 2007, **2**:e943.

30. Rohe T, Noppeney U: **Cortical hierarchies perform Bayesian**
•• **causal inference in multisensory perception**. *PLoS Biol* 2015, **13**:e1002073.
This elegant study combined neuroimaging, decoding, and computational modeling of behavioral data to suggest that causal inference emerges in a hierarchical network in the human brain. Specifically, in an audiovisual spatial localization task, spatial location is represented in sensory areas in a manner that assumes segregated sources. Posterior parietal areas, in contrast, fused these estimates from different senses and represented a common spatial source. The anterior parietal areas, in contrast, appeared to represent the spatial location according to the optimal Bayesian model for causal inference.

31. Kayser C, Shams L: **Multisensory causal inference in the brain**. *PLoS Biol* 2015, **13**:e1002075.

32. Knill DC, Pouget A: **The Bayesian brain: the role of uncertainty in neural coding and computation**. *Trends Neurosci* 2004, **27**:712-719.

33. Ma WJ, Jazayeri M: **Neural coding of uncertainty and probability**. *Annu Rev Neurosci* 2014, **37**:205-220.

34. Ma WJ, Beck JM, Latham PE, Pouget A: **Bayesian inference with probabilistic population codes**. *Nat Neurosci* 2006, **9**:1432-1438.

35. Pouget A, Beck JM, Ma WJ, Latham PE: **Probabilistic brains: knowns and unknowns**. *Nat Neurosci* 2013, **16**:1170-1178.

36. Chandrasekaran C, Lemus L, Trubanova A, Gondan M, Ghazanfar A: **Monkeys and humans share a common computation for face/voice integration**. *PLoS Comput Biol* 2011, **7**:e1002165.

37. Cappe C, Murray MM, Barone P, Rouiller EM: **Multisensory facilitation of behavior in monkeys: effects of stimulus intensity**. *J Cogn Neurosci* 2010, **22**:2850-2863.

38. Miller J: **Divided attention: evidence for coactivation with redundant signals**. *Cogn Psychol* 1982, **14**:247-279.

39. Blurton SP, Greenlee MW, Gondan M: **Multisensory processing of redundant information in go/no-go and choice responses**. *Atten Percept Psychophys* 2014, **76**:1212-1220.

40. Diederich A, Colonius H: **Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time**. *Percept Psychophys* 2004, **66**:1388-1390.

41. Raab DH: **Statistical facilitation of simple reaction times**. *Trans N Y Acad Sci* 1962, **24**:574-590.

42. Otto TU, Mamassian P: **Noise and correlations in parallel perceptual decision making**. *Curr Biol* 2012, **22**:1391-1396.

43. Otto TU, Dassy B, Mamassian P: **Principles of multisensory behavior**. *J Neurosci* 2013, **33**:7463-7474.

44. Miller J: **Statistical facilitation and the redundant signals effect: what are race and coactivation models?** *Atten Percept Psychophys* 2016, **78**:516-519.

45. Gondan M, Minakata K: **A tutorial on testing the race model**
• **inequality**. *Atten Percept Psychophys* 2016, **78**:723-735.
This paper is an important tutorial that describes the race model inequality and how to robustly apply it to behavioral data from multisensory studies. Besides statistical tests, the tutorial also discusses experimental design features that need to be carefully considered for redundant signal tasks to ensure rejection of the race model.

46. Schwarz W: **Diffusion, superposition and the redundant-targets effect**. *J Math Psychol* 1994, **38**:504-520.

47. Diederich A: **Intersensory facilitation of reaction time: evaluation of counter and diffusion coactivation models**. *J Math Psychol* 1995, **39**:197-215.

48. Drugowitsch J, DeAngelis GC, Klier EM, Angelaki DE, Pouget A:
•• **Optimal multisensory decision-making in a reaction-time task**. *Elife (Cambridge)* 2014:e03005.
This landmark study combined the coactivation model framework with the Bayesian-optimality framework and derived an optimal drift-diffusion

model to explain the behavior of observers in multisensory RT discrimination tasks.

49. Fetsch CR, Pouget A, DeAngelis GC, Angelaki DE: **Neural correlates of reliability-based cue weighting during multisensory integration**. *Nat Neurosci* 2012, **15**:146-154.

50. Gu Y, Cheng Z, Yang L, DeAngelis GC, Angelaki DE: **Multisensory convergence of visual and vestibular heading cues in the pursuit area of the frontal eye field**. *Cereb Cortex* 2016, **26(9)**:3785-3801.

51. Chen A, DeAngelis GC, Angelaki DE: **Functional specializations of the ventral intraparietal area for multisensory heading discrimination**. *J Neurosci* 2013, **33**:3567-3581.

52. Lemus L, Hernandez A, Luna R, Zainos A, Romo R: **Do sensory cortices process more than one sensory modality during perceptual judgments?** *Neuron* 2010, **67**:335-348.

53. Siegel M, Buschman TJ, Miller EK: **Cortical information flow**
•• **during flexible sensorimotor decisions**. *Science* 2015, **348**:1352-1355.
This study investigated the flow of task-related signals in monkeys trained to perform a context-dependent decision-making task. By examining neuronal responses in several different areas, the study shows that viewing flexible sensorimotor decisions as a simple hierarchical feed-forward process is an incomplete view. Sensorimotor decisions result from complex temporal dynamics that include feed-forward and feed-back interactions between frontal and posterior cortex.

54. Chen A, Gu Y, Liu S, DeAngelis GC, Angelaki DE: **Evidence for a**
• **causal contribution of macaque vestibular, but not intraparietal, cortex to heading perception**. *J Neurosci* 2016, **36**:3789-3798.
This study used Muscimol, a GABA agonist, to inactivate brain regions in the monkey thought to be involved in multisensory integration. The surprising result was that bilateral inactivation of the ventral intraparietal area, an area that is strongly predictive of the animal's choice in heading discrimination tasks, had little or almost no impact on the behavior of the monkeys. These results suggests the involvement of multiple redundant pathways for multisensory integration.

55. Katz LN, Yates JL, Pillow JW, Huk AC: **Dissociated functional**
• **significance of decision-related activity in the primate dorsal stream**. *Nature* 2016, **535**:285-288.
This paper has many parallels to the study in [54•]. Again, the main experiment in the study was to inject Muscimol into LIP or the Middle Temporal Area (MT) of monkeys performing a motion discrimination task. MT inactivation impacted the discrimination behavior. LIP inactivation, in contrast, resulted in no effects on the discrimination behavior — a result consistent with the observations from Ref. [54•]. Importantly LIP inactivation did result in changes in behavior in a memory-guided free choice task.

56. Cumming BG, Nienborg H: **Feedforward and feedback sources of choice probability in neural population responses**. *Curr Opin Neurobiol* 2016, **37**:126-132.

57. Engel TA, Chaisangmongkon W, Freedman DJ, Wang XJ: **Choice-correlated activity fluctuations underlie learning of neuronal category representation**. *Nat Commun* 2015, **6**:6454.

58. Pitkow X, Liu S, Angelaki DE, DeAngelis GC, Pouget A: **How can single sensory neurons predict behavior?** *Neuron* 2015, **87**:411-423.

59. Bizley JK, Nodal FR, Bajo VM, Nelken I, King AJ: **Physiological and anatomical evidence for multisensory interactions in auditory cortex**. *Cereb Cortex* 2006:bhl128.

60. Cappe C, Rouiller EM, Barone P: **Multisensory anatomical pathways**. *Hear Res* 2009, **258**:28-36.

61. Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK: **Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex**. *J Neurosci* 2005, **25**:5004-5012.

62. Kayser C, Petkov CI, Logothetis NK: **Visual modulation of neurons in auditory cortex**. *Cereb Cortex* 2008:bhm187.

63. Lakatos P, Chen C-M, O'Connell MN, Mills A, Schroeder CE: **Neuronal oscillations and multisensory interaction in primary auditory cortex**. *Neuron* 2007, **53**:279-292.

64. Chandrasekaran C, Lemus L, Ghazanfar AA: **Dynamic faces speed up the onset of auditory cortical spiking responses during vocal detection**. *Proc Natl Acad Sci U S A* 2013, **110**:E4668-E4677.

65. Ohyama T, Schneider-Mizell CM, Fetter RD, Aleman JV,
•• Franconville R, Rivera-Alba M, Mensh BD, Branson KM, Simpson JH, Truman JW *et al.*: **A multilevel multimodal circuit enhances action selection in Drosophila**. *Nature* 2015, **520**:633-639.
This *tour-de-force* study used a combination of electron microscopy, behavior, and electrophysiology to delineate the multi-level multisensory circuit in Drosophila larvae that controls escape locomotor behaviors.

66. Ghazanfar AA, Schroeder CE: **Is neocortex essentially multisensory?** *Trends Cogn Sci* 2006, **10**:278-285.

67. Roitman JD, Shadlen MN: **Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task**. *J Neurosci* 2002, **22**:9475-9489.

68. Tsunada J, Liu AS, Gold JI, Cohen YE: **Causal contribution of**
• **primate auditory cortex to auditory perceptual decision-making**. *Nat Neurosci* 2016, **19**:135-142.
One of the first studies to use the predictions from the drift-diffusion model originally developed for visual discrimination tasks to understand neuronal responses in belt auditory cortex during auditory decision-making. This study when taken together with reports of monkeys [37,64] performing auditory detection tasks also dispels the notion that monkeys are poor at sophisticated auditory behaviors.

69. Ghazanfar AA, Chandrasekaran C, Logothetis NK: **Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys**. *J Neurosci* 2008, **28**:4457-4469.

70. Chandrasekaran C, Ghazanfar AA: **Different neural frequency bands integrate faces and voices differently in the superior temporal sulcus**. *J Neurophysiol* 2009, **101**:773-788.

71. Russ BE, Orr LE, Cohen YE: **Prefrontal neurons predict choices during an auditory same-different task**. *Curr Biol* 2008, **18**:1483-1488.

72. Romanski LM: **Representation and integration of auditory and visual stimuli in the primate ventral lateral prefrontal cortex**. *Cereb Cortex* 2007, **17**:i61-i69.

73. Lanz F, Moret V, Rouiller EM, Loquet G: **Multisensory integration in non-human primates during a sensory-motor task**. *Front Hum Neurosci* 2013, **7**:799.

74. van Atteveldt N, Murray Micah M, Thut G, Schroeder Charles E: **Multisensory integration: flexible use of general operations**. *Neuron* 2014, **81**:1240-1253.

75. Mante V, Sussillo D, Shenoy KV, Newsome WT: **Context-dependent computation by recurrent dynamics in prefrontal cortex**. *Nature* 2013, **503**:78-84.

76. Kobak D, Brendel W, Constantinidis C, Feierstein CE, Kepecs A,
• Mainen ZF, Romo R, Qi XL, Uchida N, Machens CK: **Demixed principal component analysis of neural population data**. *Elife* 2016:5.
This study developed a dimensionality reduction technique that provides a description of the firing rate variance in a neural population as a function of task parameters such as stimuli, decisions, and rewards. The advantage of this method is that it provides 'demixed' components that still explain considerable variance. This method in combination with methods outlined in [75] are examples of supervised dimensionality reduction techniques that may be helpful to understand the heterogeneous neural populations involved in multisensory integration and other behaviors.

77. Rigotti M, Barak O, Warden MR, Wang XJ, Daw ND, Miller EK, Fusi S: **The importance of mixed selectivity in complex cognitive tasks**. *Nature* 2013, **497**:585-590.

78. Dahl CD, Logothetis NK, Kayser C: **Spatial organization of multisensory responses in temporal association cortex**. *J Neurosci* 2009, **29**:11924-11932.

79. Shenoy KV, Sahani M, Churchland MM: **Cortical control of arm movements: a dynamical systems perspective**. *Annu Rev Neurosci* 2013, **36**:337-359.

80. Churchland MM, Cunningham JP, Kaufman MT, Foster JD, Nuyujukian P, Ryu SI, Shenoy KV: **Neural population dynamics during reaching**. *Nature* 2012, **487**:51-56.

81. Machens CK, Romo R, Brody CD: **Functional, but not anatomical, separation of ''what'' and ''when'' in prefrontal cortex**. *J Neurosci* 2010, **30**:350-360.

82. Cunningham JP, Yu BM: **Dimensionality reduction for large-**
• **scale neural recordings**. *Nat Neurosci* 2014, **17**:1500-1509.

83. Chen X, Stuphorn V: **Sequential selection of economic good and action in medial frontal cortex of macaques during value-based decisions**. *Elife* 2015:4.

84. Kaufman MT, Churchland MM, Ryu SI, Shenoy KV: **Vacillation, indecision and hesitation in moment-by-moment decoding of monkey motor cortex**. *Elife* 2015, **4**:e04677.

85. Kiani R, Cueva CJ, Reppas JB, Newsome WT: **Dynamics of neural population responses in prefrontal cortex indicate changes of mind on single trials**. *Curr Biol* 2014, **24**:1542-1547.

86. Rich EL, Wallis JD: **Decoding subjective decisions from**
• **orbitofrontal cortex**. *Nat Neurosci* 2016, **19**:973-980.
This study combined simultaneous multi-electrode recordings and single-trial decoding to show that the network in the orbitofrontal cortex (OFC) of monkeys rapidly transitions between discrete, identifiable activation states that represent the value of each of two choice options of a value-based two-alternative forced-choice decision task.

87. Raposo D, Kaufman MT, Churchland AK: **A category-free neural**
•• **population supports evolving demands during decision-making**. *Nat Neurosci* 2014, **17**:1784-1792.
This study examined the types of multisensory neuronal responses in the posterior parietal cortex (PPC) of rats performing an audiovisual discrimination task. Dimensionality reduction and decoding analyses showed that neurons in the posterior parietal cortex are not easily understood using functional categories. Task-relevant variables are intermixed at the single neuron level. The network in PPC also dynamically reorganized based on task demands.

88. Sussillo D: **Neural circuits as computational dynamical systems**. *Curr Opin Neurobiol* 2014, **25C**:156-163.

89. Sussillo D, Churchland MM, Kaufman MT, Shenoy KV: **A neural network that finds a naturalistic solution for the production of muscle activity**. *Nat Neurosci* 2015, **18**:1025-1033.

90. Song HF, Yang GR, Wang XJ: **Training excitatory-inhibitory recurrent neural networks for cognitive tasks: a simple and flexible framework**. *PLoS Comput Biol* 2016, **12**:e1004792.
This study developed a method to train recurrent neural networks with biological constraints. The code pack provided by the authors also makes it effortless to train such networks. Until recently, these networks were very difficult to train and required considerable expertise in optimization. The network shown in Figure 2 designed to perform the multisensory task is based on the framework and codepack in this paper.

91. Diederich A, Colonius H: **The time window of multisensory integration: relating reaction times and judgments of temporal order**. *Psychol Rev* 2015, **122**:232-241.

92. Beck J, Ma WJ, Latham PE, Pouget A: **Probabilistic population codes and the exponential family of distributions**. *Prog Brain Res* 2007, **165**:509-519.