Article

Neuron

Neural Correlates of Optimal Multisensory Decision Making under Time-Varying Reliabilities with an Invariant Linear Probabilistic Population Code

Highlights

- Macaque LIP neurons encode heterogeneous choice and sensory modality signals
- LIP momentary evidence is proportional to visual speed and vestibular acceleration
- Summing spike across cue and time with iIPPC achieves optimal multisensory decision
- LIP responses are consistent with a network model implementing the iIPPC solution

Authors

Han Hou, Qihao Zheng, Yuchen Zhao, Alexandre Pouget, Yong Gu

Correspondence

guyong@ion.ac.cn

In Brief

Hou et al. reveal that activity in the macaque lateral intraparietal area is consistent with a theoretical framework (iIPPC) that implements optimal multisensory decision making with a simple accumulation of time-varying momentary evidence with time-invariant synaptic weights.

Hou et al., 2019, Neuron *104*, 1–12
December 4, 2019
https://doi.org/10.1016/j.neuron.2019.08.038

CellPress

Please cite this article in press as: Hou et al., Neural Correlates of Optimal Multisensory Decision Making under Time-Varying Reliabilities with an Invariant Linear Probabilistic Population Code, Neuron (2019), https://doi.org/10.1016/j.neuron.2019.08.038

Neuron
Article

Neural Correlates of Optimal Multisensory Decision Making under Time-Varying Reliabilities with an Invariant Linear Probabilistic Population Code

Han Hou,[1,2] Qihao Zheng,[1,2,4] Yuchen Zhao,[1,2,4] Alexandre Pouget,[3,5,6] and Yong Gu[1,2,5,6,7,*]
[1]Institute of Neuroscience, Key Laboratory of Primate Neurobiology, CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai Center for Brain Science and Brain-Inspired Intelligence Technology, 200031 Shanghai, China
[2]University of Chinese Academy of Sciences, 100049 Beijing, China
[3]Department of Basic Neurosciences, University of Geneva, 1211 Geneva, Switzerland
[4]These authors contributed equally
[5]These authors contributed equally
[6]Senior author
[7]Lead Contact
*Correspondence: guyong@ion.ac.cn
https://doi.org/10.1016/j.neuron.2019.08.038

SUMMARY

Perceptual decisions are often based on multiple sensory inputs whose reliabilities rapidly vary over time, yet little is known about how the brain integrates these inputs to optimize behavior. The optimal solution requires that neurons simply add their sensory inputs across time and modalities, as long as these inputs are encoded with an invariant linear probabilistic population code (ilPPC). While this theoretical possibility has been raised before, it has never been tested experimentally. Here, we report that neural activities in the lateral intraparietal area (LIP) of macaques performing a vestibular-visual multisensory decision-making task are indeed consistent with the ilPPC theory. More specifically, we found that LIP accumulates momentary evidence proportional to the visual speed and the absolute value of vestibular acceleration, two variables that are encoded with close approximations to ilPPCs in sensory areas. Together, these results provide a remarkably simple and biologically plausible solution to near-optimal multisensory decision making.

INTRODUCTION

Most perceptual decisions are based on multiple sensory inputs whose reliabilities vary over time. For instance, a predator can rely on both auditory and visual information to determine when and where to strike a prey, but these two sources of information are not generally equally reliable, nor are their reliabilities constant over time; as the prey gets closer, the quality of the image and sound typically improves, thus increasing their reliabilities. Although such multisensory decision making happens frequently in the real world, the underlying neural mechanisms remain largely unclear.

The so-called drift-diffusion model (DDM) (Ratcliff, 1978; Ratcliff and Rouder, 1998; Ratcliff and Smith, 2004; Ratcliff and McKoon, 2008), a widely used model of perceptual decision making, cannot deal with such decisions optimally in its most standard form. DDMs have been shown to implement the optimal policy for decisions involving just one source of sensory evidence whose reliability is constant over time (Laming, 1968; Bogacz et al., 2006). Under such conditions, DDMs can implement the optimal strategy by simply summing evidence over time until an upper or lower bound, corresponding to the two possible choices, is hit (Bogacz et al., 2006). This type of model lends itself to a straightforward neural implementation in which neurons simply add their sensory inputs until they reach a preset threshold (Gold and Shadlen, 2007; Ratcliff and McKoon, 2008).

When multiple sensory inputs are involved, the standard DDMs can accumulate sensory evidence optimally as long as the reliabilities of the evidence stay constant during a single trial and across trials. Under this scenario, optimal integration of evidence over time can be achieved by first taking a weighted sum of the momentary evidence at each time step, with weights proportional to the reliability of each sensory stream, followed by temporal integration (Drugowitsch et al., 2014). However, this strategy no longer works when the reliabilities change over time within a single trial. In this case, the momentary evidence must be linearly combined with weights proportional to the time-varying reliabilities, which requires that the synaptic weights change on a very fast timescale, since, in the real life, reliability can change significantly over tens of milliseconds. Moreover, when the reliabilities of the sensory inputs are not known in advance, which is typically the case in real-world situations, neurons cannot determine how to appropriately modulate their synaptic weights until after the sensory inputs have been observed. Therefore, even if it is possible to extend standard DDMs to time-varying reliability (Drugowitsch et al., 2014), it is unclear how such a solution could be implemented biologically.

In contrast, there exists another class of models that does not necessarily involve changes in synaptic strength. As long as the sensory inputs are encoded with what is known as invariant linear probabilistic population codes (ilPPCs), the neural solution for optimal multisensory integration is remarkably simple: it only requires that neurons compute linear combinations of their inputs across time or modalities using fixed—reliability-independent—synaptic weights (Ma et al., 2006; Beck et al., 2008). This solution relies on one specific property of ilPPC: the reliability of the neural code is proportional to the amplitude of the neural responses. As a result, when summing two sensory inputs with unequal reliability, the sensory input with the lowest reliability contribute less to the sum because of its lower firing rate. This is formally equivalent to weighting Gaussian samples with their reliability in an extended DDM, except that there is no need for actual weight changes with ilPPC (Ma et al., 2006).

Hence, the ilPPC framework is a promising solution to multisensory decision-making tasks, but it lacks physiological support. The only previous attempt at testing this framework with time-varying reliabilities was performed by Fetsch et al. (2011), who recorded from the dorsal medial superior temporal (MSTd) area in macaques trained to discriminate their heading direction of self-motion based on multiple sensory inputs: vestibular signals, visual optic flow, or both. In this experiment, the vestibular and visual stimuli followed a Gaussian-shape velocity temporal profile, producing naturally varied cue reliability over time. They reported that MSTd neurons take linear combination of their sensory input but that the weights change with the reliability of the input, which is inconsistent with the ilPPC prediction. However, this analysis ignored correlations among sensory neurons, which turns out to be critical for the structure of the code. Moreover, this work focused on the sensory areas and did not suggest any solution to the problem of temporal integration of sensory evidence with time varying reliabilities.

Here, we recorded the activity of single neurons in the lateral intraparietal area (LIP) under a similar task, and report that the population activity in LIP integrates its sensory inputs across modalities and time in a way consistent with the ilPPC prediction. We focus on LIP because it is the most extensively studied brain region where buildup choice-related activity has been found during visuomotor decisions in macaques (Shadlen and Newsome, 1996, 2001; Roitman and Shadlen, 2002; Gold and Shadlen, 2007; Huk et al., 2017). In addition, LIP receives abundant anatomical inputs (Boussaoud et al., 1990) from areas encoding momentary vestibular and visual self-motion information for heading discrimination, such as the MSTd area (Gu et al., 2006, 2008) and the ventral intraparietal area (VIP) (Chen et al., 2011b, 2013). It is therefore expected that the activity of LIP neurons should carry buildup choice signals germane to the formation of multisensory decisions. Note that two recent rodent studies (Raposo et al., 2014; Nikbakht et al., 2018) also have described multisensory decision signals in rat posterior parietal cortex, a region analogous to its primate counterpart. However, these studies did not characterize the computational solution implemented by these neural circuits, which is precisely the question we investigate here. Specifically, we explored whether the response of LIP neurons is consistent with the ilPPC theory in which neurons take fixed linear combinations of their sensory inputs without any need for complex, time-dependent, modality-specific reweighting of the sensory inputs during multisensory decision making. We also reanalyzed MSTd data and found that it is unlikely to be the primary site of multisensory decision making.

RESULTS

Optimal Multisensory Decision-Making Behavior on Macaques

We trained two macaque monkeys to perform a vestibular-visual multisensory decision-making task (Gu et al., 2008) (Figure 1A). On each trial, the monkeys experienced a 1.5-s fixed-duration forward motion with a small deviation either to the left or to the right of the dead ahead. At the end of the trial, the animals were required to report the perceived heading direction by making a saccade decision to one of the two choice targets (Figure 1B). We randomly interleaved three cue conditions over trials: a vestibular condition and a visual condition in which heading information was solely provided by inertial cues and optic flow, respectively, and a combined condition consisting of congruent vestibular and visual cues. Importantly, both the vestibular and visual stimuli followed a Gaussian-shape velocity temporal profile, peaking at the middle of the 1.5-s stimulus duration.

The monkeys' behavioral performance was quantified by psychometric functions for each cue condition (Figure 1C). Consistent with the previous results (Gu et al., 2008), the monkeys made more accurate decisions in the combined condition, as evidenced by a steeper psychometric function and a smaller psychophysical threshold (Figure 1C). Across all recording sessions, both monkeys' psychophysical thresholds of the combined condition were significantly smaller than those of single cue conditions and close to the prediction from the optimal cue integration theory (Knill and Richards, 1996) (Figure 1D), although monkey M was significantly supra-optimal (Raposo et al., 2012; Nikbakht et al., 2018) (see more detail at the end of the section entitled "Network model implementing ilPPC for multisensory decision making"). Therefore, the monkeys can integrate vestibular and visual cues near optimally in our multisensory decision-making task.

Heterogeneous Multisensory Choice Signals in LIP

Next, we set out to explore how these optimal decisions were formed in the brain. We recorded from 164 single, well-isolated neurons in LIP of two monkeys while they were performing the task (Figure S1). As expected, we found buildup-choice-related signals in LIP neurons under all cue conditions. As shown in peristimulus time histograms (PSTHs) of the example cells (Figures 2A and S1D–S1G), there was generally an increasing divergence between the neuron's firing rate on trials in which the monkey chose the target in the neuron's response field (IN choices) and trials in which the opposite target was chosen (OUT choices). Importantly, in all conditions, the buildup choice signals tended to be stronger for heading directions more distant away from straight ahead (Figure S2), suggesting that the response of LIP neurons reflects the accumulation of visual and vestibular sensory evidence for heading judgments.

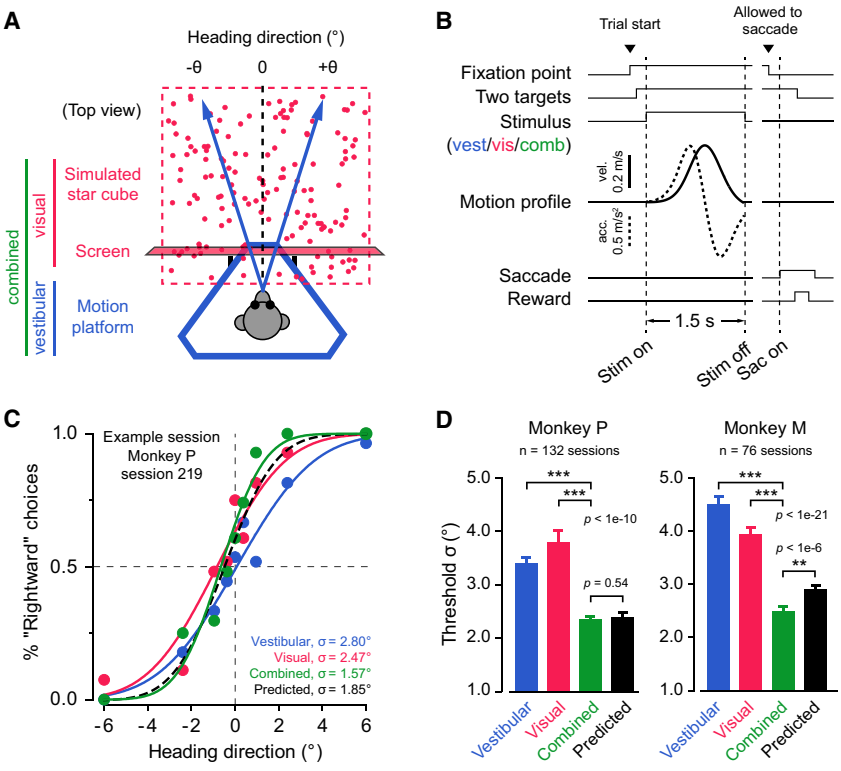

Figure 1. Optimal Cue Integration in a Vestibular-Visual Multisensory Decision-Making Task

(A) Schematic drawing of the experimental setup (top view). The monkey's task was to discriminate whether the heading direction was to the left or the right of the straight ahead (black dashed line) based on vestibular inputs (blue), visual optic flow (red), or both (green).

(B) Task timeline. Note that the self-motion speed followed a Gaussian-shape profile.

(C) Example psychometric functions from one session. The proportion of "rightward" choices is plotted against the headings for three cue conditions, respectively. Smooth solid curves represent best-fitting cumulative Gaussian functions. Black dashed curve represents the optimal prediction.

(D) Average psychophysical thresholds from two monkeys for three conditions and predicted thresholds calculated from optimal cue integration theory (black bars).

Error bars indicate SEM; p values are from a paired t test.

To quantify the strength of choice-related signals, we used receiver operating characteristic (ROC) analysis to generate an index of choice divergence (CD) (Raposo et al., 2014) (Figure 2B). The four cells illustrated in Figure 2 exhibited canonical ramping choice signals, but their CDs varied greatly across cue conditions. For example, for cell 1, the CD was largest in the combined condition, modest in the visual condition, and smallest in the vestibular condition. By contrast, for cell 4, the CD was largest in the vestibular condition. This heterogeneity was also manifest at the population level (Figures 2C and 2D). Approximately half of LIP neurons exhibited significant CD ($p < 0.05$, two-sided permutation test) in each condition (vestibular, 52%; visual, 46%; combined, 59%; Figure 2D), but they did not fully overlap (Figure 2C). Although vestibular and visual CD is significantly correlated across the population, the correlation was relatively weak ($p = 5.2e\text{-}17$, $r^2 = 0.35$; Figure S3B). In addition, while more than two-thirds of neurons (76%) had significant CD in any of the three conditions, only one-third of neurons (30%) had significant CD in all of the three conditions (Figure 2D).

Apart from the heterogeneous choice signals, LIP also encodes heterogeneous sensory modality signals. As shown in Figure S3A, most LIP neurons carried mixed choice and modality signals, exhibiting a category-free-like neural representation, as previously seen in rat posterior parietal cortex (Raposo et al., 2014). However, although randomly mixed at the single-neuron level, the choice and modality signals can still be linearly decoded from the LIP population, as shown by the results from demixed principal-component analysis (dPCA) (Kobak et al., 2016; Rossi-Pool et al., 2017) (Figures S3D–S3F). Therefore, we ignore the mixed modality signals thereafter, since they are orthogonal to the decision signals and are irrelevant to our task.

Another potential difficulty in interpreting LIP activity arises from the fact that LIP neurons also multiplex a combination of temporally overlapping decision- and non-decision signals (Meister et al., 2013; Park et al., 2014). In particular, the signal of saccade preparation may interfere with the one reflecting evidence accumulation (Shadlen and Newsome, 2001). However, this was not likely to be an issue in our study. In our fixed-duration task, we introduced a 300- to 600-ms delay between the stimulus offset and the time at which the monkey was allowed to saccade (STAR Methods). Moreover, the monkeys tended to stop integrating evidence ~500 ms prior to the stimulus offset (see Figure 3B and below), further separating in time the processes of evidence accumulation and saccade preparation. Therefore, the premotor activity of LIP should not play a significant role in our analysis of multisensory evidence accumulation.

LIP Integrates the Visual Speed and Absolute Value of Vestibular Acceleration

Despite the high degree of heterogeneity among the LIP population, the choice signals consistently ramped up earlier in the vestibular and combined conditions than in the visual condition (Figure 3). This was evident not only in population analyses, such as the percentage of cells with significant choice signals (Figure 3A), the averaged PSTH (Figure 3B), and the demixed choice components (Figure S3F), but also in the cell-by-cell analysis (Figure 3C; see also Figure 2C). Specifically, we used divergence time to quantify the onset of choice signals for each neuron in each cue condition (STAR Methods). It turned out that while the vestibular and combined divergence times were similar ($p = 0.054$, $n = 73$; two-tailed paired t test), both of them were significantly shorter than the visual one (vestibular-visual: $p = 8e\text{-}16$, $n = 59$; combined-visual: $p = 7e\text{-}19$, $n = 67$;

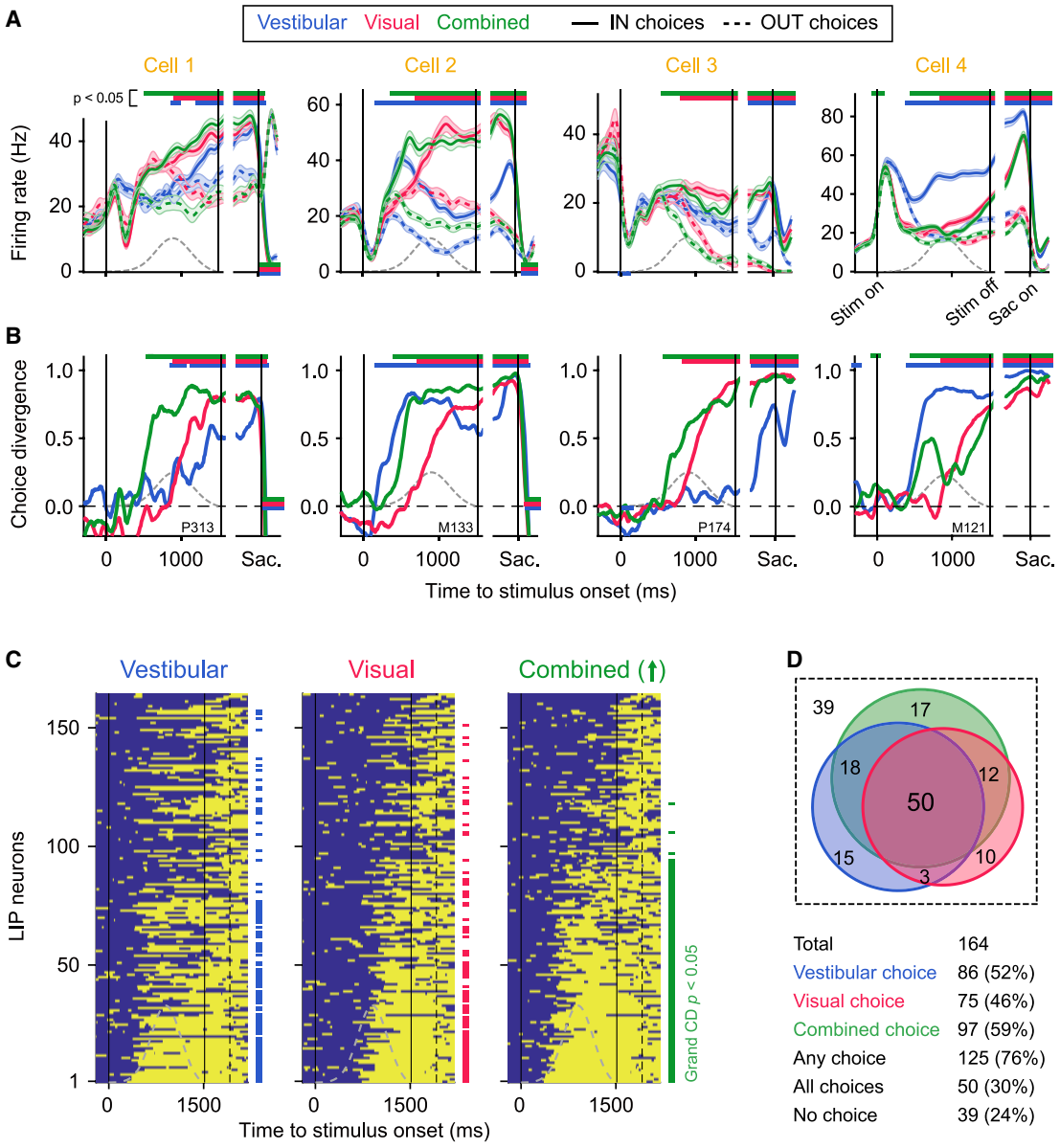

Figure 2. Heterogeneous Choice Signals in the LIP Population

(A) Peri-stimulus time histograms (PSTHs) of four example cells. Spike trains were aligned to stimulus onset (left subpanels) and saccade onset (right subpanels), respectively, and grouped by cue condition and monkey's choice. Vestibular, blue; visual, red; combined, green. Toward the cell's response field (RF), or IN choices, solid curves; away from the cell's RF, or OUT choices, dashed curves. Shaded error bands represent SEM. Horizontal color bars represent time epochs in which IN and OUT trials have significantly different firing rates ($p < 0.05$, two-tailed t test), with the color indicating cue condition and the position indicating the relationship between IN and OUT firings (IN > OUT, top; IN < OUT, bottom). Gray dashed curves represent the actual speed profile measured by an accelerometer attached to the motion platform.

(B) Choice divergence (CD) of the same four cells (STAR Methods). Horizontal color bars are the same as in (A), except that p values were from permutation test (n = 1,000).

(C) Timing of significant choice signals (yellow patches; $p < 0.05$, two-tailed t test) for all LIP neurons (vertical axis) in the three cue conditions (left, middle, and right panels). Cells were sorted according to their grand CDs (CD computed from all spikes in 0–1,500 ms) under the combined condition. Small colored bars next to each panel indicate cells with significant grand CDs ($p < 0.05$, permutation test) under each condition (see D).

(D) Venn diagram showing the distribution of overall choice signals. The number within each colored area indicates the number of neurons that have significant grand CDs under the corresponding combination of cue conditions.

See also Figures S1–S3.

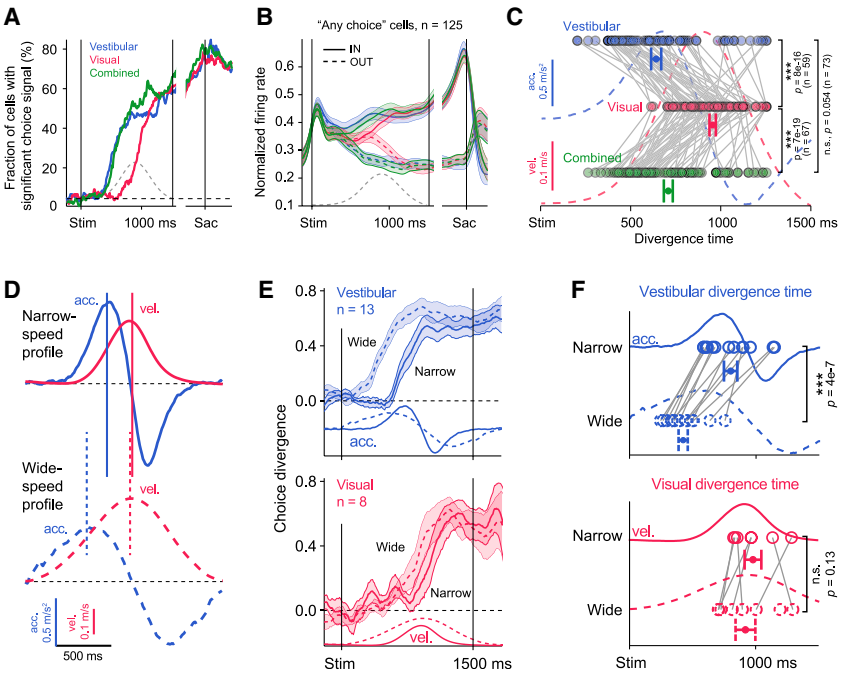

Figure 3. LIP Integrates the Visual Speed and Absolute Value of Vestibular Acceleration

(A) Fraction of cells with significant choice coding as a function of time (see Figure 2C). Horizontal dashed line represents the type I error rate of 5%.

(B) Population average of normalized PSTHs from 125 "any choice" cells. Shaded error bands represent SEM.

(C) Divergence time of cells with significant grand CD for each condition (STAR Methods). Gray lines connect data from the same cells. p values are from a two-tailed paired t test; note that the numbers of pairs were different for each comparison. Data points with horizontal error bars represent mean ± SEM of the population divergence time; acceleration and speed profiles are shown in the background.

(D) Two motion profiles used to isolate contributions of acceleration and speed to LIP ramping. Note that by widening the speed profile, we shifted the time of acceleration peak forward (blue vertical lines) while keeping the speed peak unchanged (red vertical lines).

(E) Vestibular and visual CDs under the two motion profiles.

(F) Comparison of divergence time between narrow and wide profiles. p values are from a two-tailed t test.

two-tailed paired t test) (Figure 3C). Notably, the averaged divergence time under the vestibular and combined conditions aligned well to the acceleration peak of the Gaussian-shape motion profile, whereas the divergence time under the visual condition lagged behind and aligned better to the velocity peak (Figure 3C, dashed curves). This suggests that the physical quantities being integrated over time are speed for the visual stimulus and the absolute value of acceleration for the vestibular stimulus.

An alternative explanation, however, might be that the apparent ~400-ms interval between the vestibular and visual ramping was caused purely by a difference in their sensory latencies. For example, LIP activity could have been driven by an ultrafast vestibular signal but a slow visual signal, both of which followed the velocity of the motion. To test this, we designed an experiment in which we used two distinct velocity profiles: a wide one and a narrow one (Figure 3D). These profiles were designed to have temporally aligned velocity peaks but misaligned acceleration peaks. If our original physical-quantity hypothesis was correct, then we would expect the visual ramping to remain nearly the same, while the vestibular ramping should start earlier for the wide profile than for the narrow one. In contrast, if the sensory-latency hypothesis was correct, then there should be no shift in either the vestibular or visual ramping across the two profiles. Our data matched the first prediction (Figures 3E and 3F), which supports the hypothesis that LIP accumulates different momentary physical quantities over time. This physiological finding echoes a recent psychophysical study showing that at the behavioral level, human subjects optimally integrate vestibular and visual momentary evidence with reliability following the amplitude of acceleration and velocity, respectively (Drugowitsch et al., 2014) (see Discussion).

Network Model Implementing iIPPC for Multisensory Decision Making

Next, we developed a neural model of multisensory decision making (refer to as M1 thereafter), which takes as input vestibular neurons tuned to acceleration and visual neurons tuned to velocity as observed *in vivo* (Equations 2 and 3 in STAR Methods; Figures 4A and 4B, bottom). These inputs converge onto an integrator layer, which takes the sum of the visual and vestibular inputs and integrates this summed input over time (Figure 4A, middle). This layer projects in turn to an output layer, labeled LIP, which sums the integrated visuo-vestibular inputs with the activity from another input layer encoding the two possible targets to which the animal can eventually saccade (Figure 4A, top, and Figure 4B, middle and top). As long as the input layers encode the sensory inputs with iIPPC, this simple network can be shown analytically to implement the Bayes optimal solution, even when the reliability of the sensory inputs varies over time, as is the case in our experiment (Ma et al., 2006; Beck et al., 2008).

In an iIPPC, the gain, or amplitude, of the tuning curves of the neurons should be proportional to the reliability of the encoded variable. For instance, in the case of vestibular neurons, the amplitude of the tuning curves to heading should scale with the absolute value of acceleration. *In vivo*, however, the responses of visual neurons are not fully consistent with the assumption of iIPPC, because while the amplitude does increase with reliability, in some neurons, the baseline activity decreases with reliability (Figure S4A; see also Equations 2 and 3 in STAR Methods). This violation of the iIPPC assumption implies that a simple sum of activity could incur an information loss. Instead, one would have to posit a model with time-varying synaptic weights in order to achieve optimal integration (Fetsch et al., 2011), which would be less biologically plausible. Fortunately,

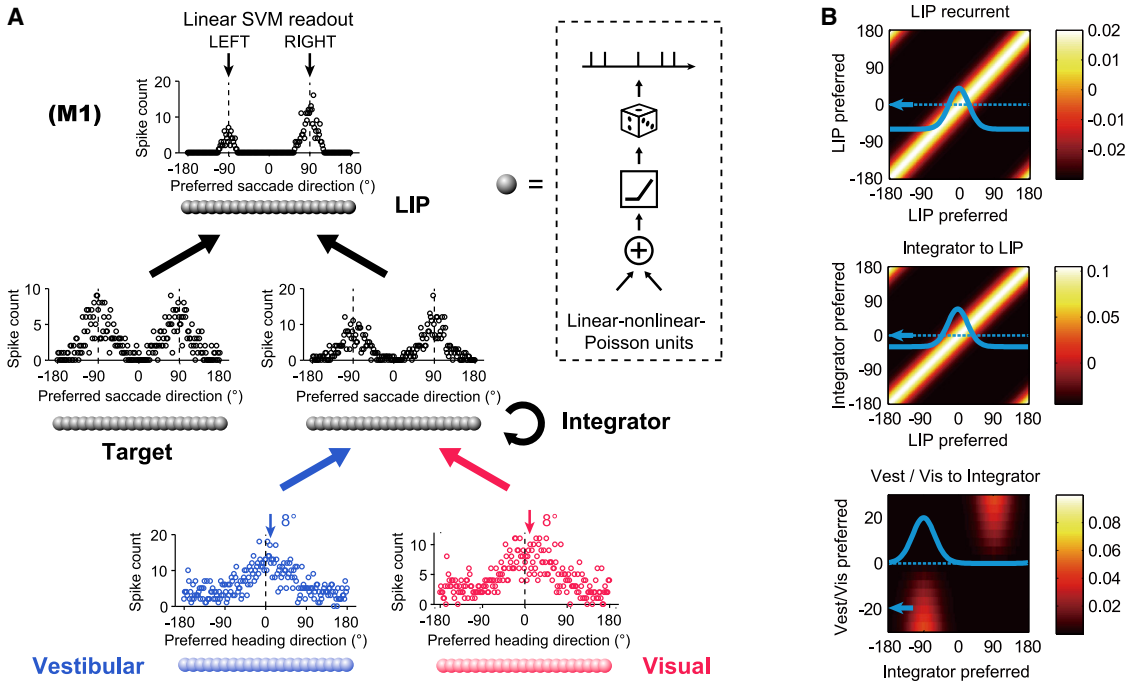

Figure 4. Neural Network Model with Invariant Linear Probabilistic Population Codes (ilPPCs) for Multisensory Decision Making

(A) Network architecture of model M1. The model consists of three interconnected layers of linear-nonlinear-Poisson units (inset). Units in visual and vestibular layers have bell-shaped ilPPC-compatible tuning curves for heading direction. The intermediate integrator layer simply sums the incoming spikes from the two sensory layers over time and transforms the tuning curves for heading direction to that for saccade direction (−90°, leftward choice; +90°, rightward choice). The LIP layer receives the integrated heading inputs from the integrator layer, together with visual responses triggered by the two saccade targets. Once a decision boundary is hit, or when the end of the trial is reached (1.5 s), LIP activity is decoded by a linear support vector machine for action selection. Circles indicate representative patterns of activity for each layer; spike counts from 800 to 1,000 ms; combined condition, 8° heading.

(B) Weight matrices used in the network. Bottom: feedforward connections from the vestibular and visual layers to the integrator layer; middle: feedforward connections from the integrator layer to the LIP layer; top: LIP recurrent connections. Cyan curve indicates synaptic weights from one representative presynaptic neuron (cyan arrow) to postsynaptic neural population (horizontal axis) (dashed line, zero).

See also Figures S4 and S7.

however, this information loss is small for a population of neurons with information limiting correlations (Moreno-Bote et al., 2014), which are thought to be present in MSTd (Pitkow et al., 2015; Lakshminarasimhan et al., 2018). Indeed, we found numerically that the information loss was ~5% over a wide range of parameters values (Fano factor, mean correlation, baseline changes, and so on) (STAR Methods; Figures S4B–S4E).

Importantly, we also endowed the network with a stopping mechanism that terminates sensory integration whenever a function of the LIP population activity reaches a preset threshold (STAR Methods). Our experiment is not a reaction time experiment and may not require, in principle, such a stopping bound. However, as can be seen in Figure 3B, LIP population response saturates at ~1 s, suggesting that evidence integration stops prematurely. This is indeed consistent with the previous results suggesting that animals and humans use a stopping bound even in fixed duration experiments (Kiani et al., 2008). This could also explain the paradoxical supra-optimal behavior in monkey M (Figure 1D). Specifically, the monkey's performance can be better than optimal if the stopping bound in the combined condition is well above the stopping bound in the visual and vestibular conditions (Drugowitsch et al., 2014; Drugowitsch et al., 2015). In this case, the animal would integrate evidence for a longer time in the combined condition than in single cue conditions, resulting in supra-optimal performance (Raposo et al., 2012; but see Nikbakht et al., 2018).

LIP Data Are Compatible with the ilPPC Framework

In the first set of simulations on M1, we adjusted the height of the stopping bounds and found that the model can replicate the animals' near-optimal performance (Figure 5A). We then plotted the population activity of neurons in the LIP layer (Figures 5B and 5C). As expected, and also similar to our monkey LIP data (Figure 3B), the average LIP activity of model M1 (Figure 5C) in the combined condition follows at first the vestibular condition and then diverges due to the rise of the visual input. In the second half of the trial, the activity in all conditions tend to saturate due to the fact that the population activity has hit the stopping bound on most trials and therefore stops rising until the end of the trial.

Neurons in M1 are homogeneous in the sense that they all take a perfect sum of their vestibular and visual inputs. Importantly, however, optimal integration does not require such a perfect sum; it can also be achieved with random linear combinations

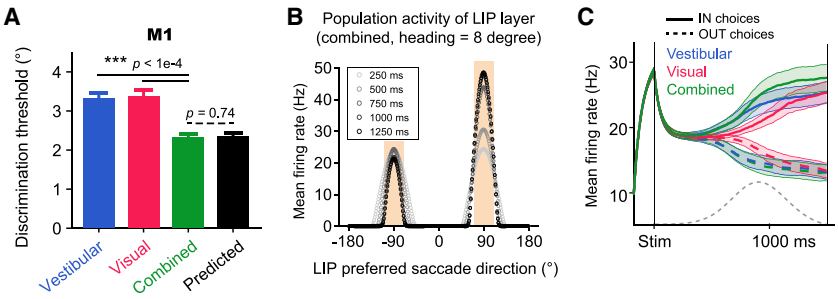

Figure 5. Homogeneous iIPPC Model M1 Performs the Task Near Optimally

(A) Model M1 exhibited near-optimal behavior as the monkey. The psychophysical threshold under the combined condition (green) was indistinguishable from the Bayesian optimal prediction (black).

(B) Population firing rate in the LIP layer at five different time points (the same stimulus as in A, averaged over 100 repetitions). Yellow shaded areas indicate the range of cells that were used to calculate the averaged PSTHs in (C).

(C) Average PSTHs across LIP population. Notations are the same as in Figure 3B.

See also Figure S5.

of vestibular and visual inputs (Ma et al., 2006). Accordingly, we simulated a second model, referred to as M2, in which the visual and vestibular weights of each neuron were drawn from lognormal distributions (Figures S5A and S5B; STAR Methods). Like M1, model M2 can be tuned to reproduce the Bayes optimal discrimination thresholds (Figures S5C and S5E). However, in contrast to model M1, the neurons showed a wide range of response profiles similar to what we observed *in vivo* (Figure S5D). In particular, we found that the distribution of visual and vestibular weights was similar in M2 and LIP data (Figure S5F; STAR Methods).

Since model M2 is a linear combination away from model M1, we tested whether the response of M1 neurons (Figure 6A) could be estimated by linearly combining the response of M2 neurons (Figure 6B). Multivariate linear regression confirmed that M1 response profiles could indeed be perfectly reproduced by linearly combining M2 responses (Figure 6C). Likewise, since LIP neurons also appear to be computing random linear combinations of visual and vestibular inputs, the same result should hold for LIP responses. This is indeed what we found; the response of M1 neurons can be closely approximated by linearly combining the response of LIP neurons (Figures 6D and S6A).

This last result is key; it suggests that LIP neurons behave quite similarly to the neurons in M2. The two sets of neurons, however, differ quite significantly in how they integrate their inputs over time. LIP neurons display a wide variety of temporal profiles (see Figures 2C and S1D–S1G), suggesting that very few neurons act like perfect temporal integrators, in contrast to M2 neurons. Nonetheless, the fact that linear combinations of LIP neurons could reproduce the response of M1 neurons indicates that LIP responses provide a basis set sufficiently varied to allow perfect integration at the population level, a result consistent with what has been recently reported in the posterior parietal cortex of rats engaged in a perceptual decision-making task (Scott et al., 2017).

In addition to this second model, we simulated a third model (M3) in which the time constant of the integrator layer was reduced to 100 ms. Interestingly, we found that it was not possible to linearly combine the responses of M3 output neurons to reproduce the traces of the optimal model M1 (Figures 6E and 6G), thus emphasizing the importance of long integration time constant for fitting the optimal model. We also wondered whether M1 could be fitted by the response of MSTd neurons, which are known to combine visual and vestibular responses and whose time constant are believed to be of the same order as model M3. We found that the fit to M1 from MSTd neurons was markedly worse than those obtained from M2 and LIP but was close to that from M3 (Figures 6F and 6G). Moreover, only a small fraction of cells contributed significantly to this fit, in sharp contrast to what we observed in M2 and LIP (Figures 6H and 6I). In fact, the late phase of M1 responses was captured mostly by MSTd cells with short time constants that seemed sensitive to deceleration rather than integrating cells (Figure S6B).

Finally, we computed the shuffled Fisher information over time for the models and the experimental data (Figure 7). The Fisher information in a neuronal population is a measure inversely proportional to the square of the discrimination threshold of an ideal observer (Seung and Sompolinsky, 1993; Beck et al., 2011) (STAR Methods). Our network simulations revealed that the shuffled Fisher information should increase over time in all conditions, reflecting the temporal accumulation of evidence (Figure 7A). In addition, we observed that this rise in information starts earlier in the vestibular condition than in the visual one because of the temporal offset between acceleration and velocity. In the combined condition, the Fisher information follows at first the vestibular condition, and then exceeds the vestibular trace even slightly earlier than the visual information becomes available. This seemingly nonlinear effect is due to the higher stopping bound used in the combined condition (see STAR Methods and Figures S6C–S6G). Remarkably, the shuffled Fisher information estimated from the LIP responses follows qualitatively the same trend as the ones observed in the model (Figure 7B). In contrast to M2 and LIP neurons, shuffled Fisher information in M3 and MSTd followed the profile expected for neurons with short time constant; it simply reflected the velocity profile of the stimulus and did not exhibit the plateau expected from a decision area (Figures 7C and 7D).

DISCUSSION

Integrating ever-changing sensory inputs from different sources across time is crucial for animals to optimize their decisions in a complex environment, yet little is known about the underlying mechanisms. This study presents, to the best of our knowledge, the first investigation of the neural basis of optimal multisensory decision making in non-human primates. We found that LIP neurons in the macaque posterior parietal cortex encode ramping

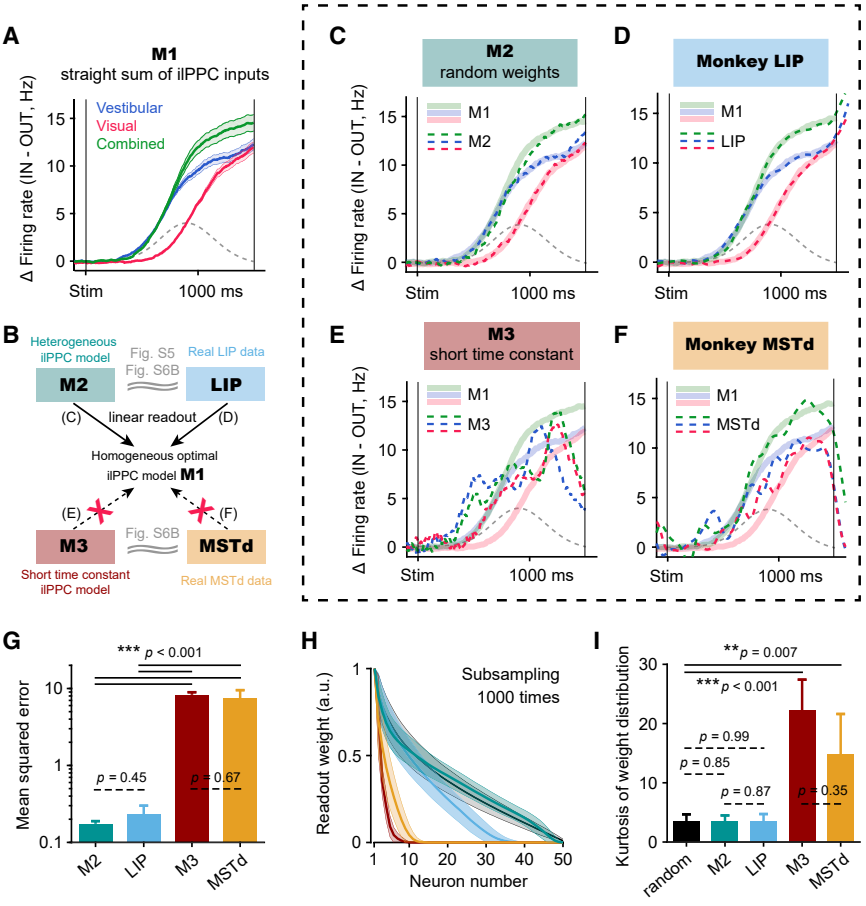

Figure 6. Optimal ilPPC Model M1 Can Be Linearly Approximated by M2 and LIP, but Not by M3 or MSTd

(A) Ramping activity of M1 computed as the difference of PSTHs for IN and OUT trials. Since M1 is optimal and homogeneous, we refer to M1's activities as "optimal traces" (see the main text). Notations are the same as before.

(B) Schematic summary of results in (C)–(F).

(C) Optimal traces from M1 (thick shaded bands) can be linearly reconstructed by population activities obtained from a heterogenous model M2 (dashed curves). Model M2 had the same network architecture as M1, except that it relies on random combinations of ilPPC inputs in the integrator layer.

(D) Optimal traces can also be linearly reconstructed by heterogenous single neuron activities from the LIP data.

(E and F) In contrast, the optimal traces cannot be reconstructed from activities of a suboptimal model M3 (E) or from the MSTd data (F).

(G) Mean squared error of the fits in (C)–(F). Error bars and p values are from a subsampling test (n = 50 neurons, 1,000 times; STAR Methods).

(H) Normalized readout weights ordered by magnitude. Shaded error bands indicate SDs of the subsampling distributions.

(I) The kurtosis of the distributions of weights. The black curve in (H) and black bar in (E) were from random readout weights.

See also Figures S5 and S6.

decision signals not only for the visual condition, as widely shown in the literature, but also for the vestibular and combined conditions, except with distinct temporal dynamics. Importantly, these data are compatible with an ilPPC framework where optimal multisensory evidence accumulation is achieved by simply summing sensory inputs across both modalities and time, even with mismatched temporal profiles of cue reliabilities and with heterogeneous sensory-motor representation. Therefore, our results provide the first neural correlate of near-optimal multisensory decision making and a crucial test of the ilPPC framework.

Distinct Visual and Vestibular Temporal Dynamics in LIP

By comparing the temporal dynamics of LIP population under different modalities, we found that LIP neurons accumulate the visual speed and the absolute value of vestibular acceleration, which serve as momentary evidence for their respective modalities. These findings may seem confusing at first glance, since it is more intuitive to assume that neural circuits would combine evidence with the same temporal dynamics across cues, namely, either visual and vestibular speed or visual and vestibular acceleration (Gu et al., 2006; Fetsch et al., 2010; Chen et al., 2011a; Smith et al., 2017). In support of this idea, although vestibular peripheral organs, such as the otoliths, encode inertial motion predominantly by acceleration (Fernández and Goldberg, 1976), recent studies have found a transformation from acceleration-dominated to speed-dominated vestibular signal along the subcortical vestibular pathway, i.e., from otolith organs to the CNS (Laurens et al., 2017), followed by a moderate but noticeable further transformation along several sensory cortices (Gu et al., 2006; Fetsch et al., 2010; Chen et al., 2011a; Laurens et al., 2017). Given that visual motion responses are typically dominated by speed (Rodman and Albright, 1987; Lisberger and Movshon, 1999; Gu et al., 2006), one would think that the brain may deliberately turn the vestibular signal from acceleration- to speed-sensitive to facilitate the combination with the visual signal.

However, if the vestibular momentary evidence is proportional to acceleration corrupted by white noise across time, integrating this evidence to obtain a velocity signal would not simplify decision making. On the contrary, this step would introduce temporal correlations (Churchland et al., 2011), in which case, even with ilPPC, a simple sum of the momentary evidence would no longer be optimal (Bogacz et al., 2006). Instead, downstream circuits would have to compute a weighted sum of the sensory evidence, which would effectively differentiate the momentary evidence before summing them. In other words, optimal integration would effectively recover the original acceleration signals. Our results, along with previous psychophysical results (Drugowitsch et al., 2014), strongly suggest that the brain does not go through this extra step but uses the acceleration signals as momentary evidence instead, which makes sense given the nature of vestibular peripheral sensors (Fernández and Goldberg, 1976).

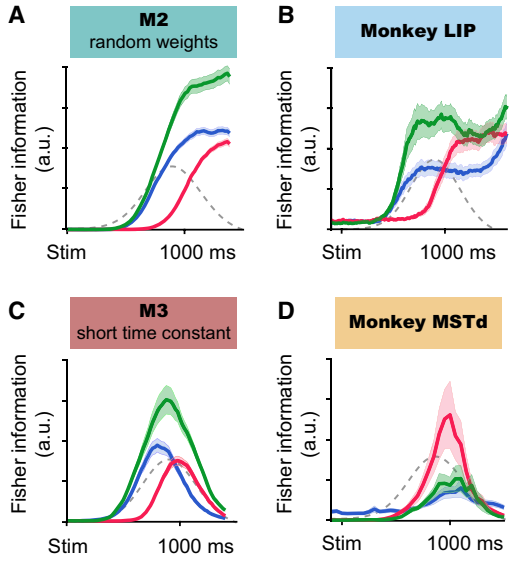

Figure 7. Shuffled Fisher Information for the Model and the Experimental Data

(A) Shuffled Fisher information of M2 (STAR Methods). Shaded error bands represent SEM estimated from bootstrap. Note that the absolute value of shuffled Fisher information is arbitrary.

(B–D) Same as in (A) but for the monkey LIP data (B), the M3 responses (C), and the monkey MSTd data (D), respectively. Note that the Fisher information of MSTd neurons was lower in the vestibular and the combined conditions than in the visual condition (compare curves in D). This is due to weaker vestibular responses and the existence of "opposite neurons" in MSTd (see discussion in STAR Methods).

See also Figure S6.

It should also be noted that accumulating vestibular evidence with reliability proportional to the absolute value of acceleration in decision areas such as LIP is a process fundamentally different from transforming vestibular acceleration to velocity in sensory areas such as MSTd and VIP (Laurens et al., 2017). First, LIP neurons do not encode passive vestibular heading information when the animal is translated without making any decision, in sharp contrast to neurons in the sensory cortices, including MSTd, VIP, and visual posterior sylvian area (VPS) (mean direction discrimination index [DDI] ± SEM: LIP, 0.42 ± 0.01; MSTd, 0.59 ± 0.01; VIP, 0.61 ± 0.01; VPS, 0.69 ± 0.01; $DDI_{LIP} < DDI_{MSTd, VIP, VPS}$, $p < 1e\text{-}10$, two-tailed t test; Y.G., unpublished data; Gu et al., 2006; Chen et al., 2011b). Second, LIP appears to integrate the absolute value of acceleration (not acceleration), because it provides the momentary evidence for deciding between rightward and leftward self-motion.

Multisensory Convergence in the Brain for Heading Decision

One of the long-standing questions about multisensory integration is whether integration takes place early or late along the sensory streams (Bizley et al., 2016). There are clear signs of multisensory responses in the relatively early or mid-stage of sensory areas, thus supporting the early theory (Gu, 2018). Our results are more consistent with the late-convergence theory in which multisensory momentary evidence are combined across modalities and time in decision areas such as LIP. However, this dichotomy between early and late theories does not necessarily make sense given the recurrent nature of the cortical circuitry. In a highly recurrent network, it is notoriously difficult to identify a node as a primary site of integration. Thus, integration might take place simultaneously across multiple sites but in such a way that the output of the computation is consistent across sites. For example, Deneve et al. (2001) demonstrated how this could take place in a large recurrent network performing optimal multisensory integration, though their work did not consider the problem of temporal integration.

It might be possible to gain further insight into the distributed nature of multisensory decision making by combining the previous models with the one we have presented here. Such an extended model might explain why vestibular momentary evidence is tuned to velocity by the time they appear in MSTd (Laurens et al., 2017; Gu, 2018) and why this velocity tuned vestibular input does not appear to be integrated in LIP. It could also shed light on recent physiological experiments in which electrical microstimulation and chemical inactivation in MSTd could dramatically affect heading discrimination based on optic flow while this effect was largely negligible in the vestibular condition (Gu et al., 2012). By contrast, and in accord with our finding that LIP integrates the absolute value of vestibular acceleration, inactivating the vestibular cortex parieto-insular vestibular cortex (PIVC), where vestibular momentary evidence is dominated by acceleration (Chen et al., 2011a; Laurens et al., 2017), substantially diminished the macaque's heading ability based on vestibular cue (Chen et al., 2016). A detailed construction of such a model will eventually be required for a multi-area theory of multisensory decision making.

The Role of LIP in Multisensory Decision Making

The causal role of macaque LIP in perceptual decision making remains controversial (Huk et al., 2017). While inactivation of macaque LIP has been found to have no effect on decisions in the classical random dots task (Katz et al., 2016), a recent study came to a different conclusion with a modified task design (Zhou and Freedman, 2019). In addition, causal experiments on rodent posterior parietal cortex (an area that is supposed to be analogous to macaque LIP) have also yielded varying results (Najafi and Churchland, 2018). The apparent role of rodent posterior parietal cortex turns out to depend on confounding factors such as sensory modality (Guo et al., 2014; Raposo et al., 2014; Erlich et al., 2015; Hanks et al., 2015; Licata et al., 2017), learning process (Zhong et al., 2019), and interactions among different brain regions (Li et al., 2016).

This controversial issue, however, is orthogonal to the conclusions of the current study. As long as LIP responses reflect the integration process of multisensory evidence, even if the actual integration takes place somewhere else, we can still use these responses to test the prediction of our ilPPC model (Pisupati et al., 2016; O'Connell et al., 2018).

Computational Models for Multisensory Decision Making

Our results indicate that, at the population level, LIP implements a near-optimal solution for multisensory decision making under

the assumption that the sensory inputs are encoded with ilPPC. At first sight, this assumption is not perfectly satisfied in our model system, but we saw that the deviation from ilPPC introduces only a minor information loss (see Figure S4), as long as there are strong differential correlations in MSTd and vestibular areas, which seems to be the case *in vivo* (Pitkow et al., 2015). As to the LIP layer, our negative control with model M3 showed that if LIP neurons responded like the output neurons in model M3 (the time constant of the integrator layer was 100 ms), we would not be able to reconstruct linearly the optimal responses from model M1. This result, again, lends support to the notion that LIP implements a solution close to ilPPC.

We showed in model M2 that the ilPPC solution for multisensory decision making can be implemented by heterogeneous population, just like what we found in LIP. In other words, heterogeneity is not a problem for optimal integration, and thus there is no incentive for neural circuits to get rid of it. On the contrary, there might be some positive reasons for keeping it, such as causal inference (Körding et al., 2007; Rohe and Noppeney, 2015). To decide whether to fuse or not to fuse sensory inputs from multiple modalities, the brain must have access to the unimodal responses that have been integrated only over time but not across modalities. Such responses can be retrieved from a heterogeneous population like M2 or LIP (e.g., vestibular or visual only neurons), but not from homogeneous population like M1.

While these results provide the strongest test of the ilPPC theory to date, it will be important to test in future experiments other predictions of the ilPPC framework. In particular, the ilPPC theory predicts that LIP activity encodes a full probability distribution over choices given the evidence so far (Beck et al., 2008). Testing this prediction thoroughly requires simultaneous recording of LIP ensemble, manipulating the cue reliability (motion profile or visual coherence) on a trial-by-trial basis, and preferably engaging the animals in a reaction-time task, all of which should be addressed in future studies.

There are of course other models of decision making that could potentially account for the responses we have observed in LIP (Chandrasekaran, 2017). In particular, it has been argued that LIP is part of a network of areas implementing point attractor networks (Wang, 2002; Wong and Wang, 2006). However, it is not immediately clear how this approach can be generalized to the type of decision we have considered here. Indeed, as we have seen, the optimal solution depends critically on the code that is used to encode the momentary evidence. To the extent that this code is close to an ilPPC, the optimal solution is to sum the inputs spikes, in which case one needs a line attractor network, which is effectively what our network approximates. Therefore, as long as these previous models of decision making are fine-tuned to approximate line attractor networks, and as long as they are fed ilPPCs as inputs, the two classes of models would be equivalent.

Training recurrent neural network (RNNs) on our task (Mante et al., 2013; Song et al., 2017) provides a third alternative for modeling multisensory decision making. We also tried this approach and found that the resulting network was capable of reproducing the behavioral thresholds of the animal while exhibiting a wide variety of single neuron responses similar to what we saw in LIP (Figure S7). Nonetheless, this approach has one major drawback: it makes it very difficult to understand how the network solves the task. We could try to reverse engineer the network, but given that an analytical solution can be derived from first principles for our task, and given that this solution is close to what we observed in LIP, it is unclear what insight could be gained from the recurrent network. In contrast, our ilPPC model provides a close approximation to the optimal solution, consistent with the experimental results, along with a clear understanding as to why this approach is optimal.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Apparatus
  - Behavioral Tasks
  - Electrophysiology
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Psychophysics
  - Choice-related neural activities
  - Linear Fitting of Mean Firing Rates
  - Fisher Information Analysis
  - Network Simulation of ilPPC Framework
  - Estimating the Information Loss of ilPPC Solution with MSTd-like Neural Population
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.neuron.2019.08.038.

ACKNOWLEDGMENTS

We thank Jianyu Lu, Zhao Zeng, and Xuefei Yu for collecting part of the MSTd data, Wenyao Chen for monkey care and training, and Ying Liu for C++ software programming. We thank Jan Drugowitsch and Tianming Yang for helpful discussions. We thank the National Supercomputing Center of Tianjin and the Bioinformatics Facility of ION, CAS for computational resources. This work was supported by grants from the National Natural Science Foundation of China Project (31761133014), the Strategic Priority Research Program of CAS (XDB32070000), and the Shanghai Municipal Science and Technology Major Project (2018SHZDZX05) to Y.G.; and by grants from the Simons Collaboration for the Global Brain and the Swiss National Science Foundation (31003A_165831) to A.P.

AUTHOR CONTRIBUTIONS

H.H. and Y.G. conceived the project and designed the experiments. H.H., Q.Z., and Y.Z. performed the experiments. H.H. analyzed the data. H.H. and A.P. developed the models and implemented the simulations. H.H., A.P., and Y.G. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 19, 2019
Revised: July 21, 2019
Accepted: August 22, 2019
Published: October 10, 2019

REFERENCES

Abbott, L.F., and Dayan, P. (1999). The effect of correlated variability on the accuracy of a population code. Neural Comput. *11*, 91–101.

Barash, S., Bracewell, R.M., Fogassi, L., Gnadt, J.W., and Andersen, R.A. (1991). Saccade-related activity in the lateral intraparietal area. I. Temporal properties; comparison with area 7a. J. Neurophysiol. *66*, 1095–1108.

Beck, J.M., Ma, W.J., Kiani, R., Hanks, T., Churchland, A.K., Roitman, J., Shadlen, M.N., Latham, P.E., and Pouget, A. (2008). Probabilistic population codes for Bayesian decision making. Neuron *60*, 1142–1152.

Beck, J., Bejjanki, V.R., and Pouget, A. (2011). Insights from a simple expression for linear fisher information in a recurrently connected population of spiking neurons. Neural Comput. *23*, 1484–1502.

Bizley, J.K., Jones, G.P., and Town, S.M. (2016). Where are multisensory signals combined for perceptual decision-making? Curr. Opin. Neurobiol. *40*, 31–37.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., and Cohen, J.D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. Psychol. Rev. *113*, 700–765.

Boussaoud, D., Ungerleider, L.G., and Desimone, R. (1990). Pathways for motion analysis: cortical connections of the medial superior temporal and fundus of the superior temporal visual areas in the macaque. J. Comp. Neurol. *296*, 462–495.

Chandrasekaran, C. (2017). Computational principles and models of multisensory integration. Curr. Opin. Neurobiol. *43*, 25–34.

Chen, A., DeAngelis, G.C., and Angelaki, D.E. (2011a). A comparison of vestibular spatiotemporal tuning in macaque parietoinsular vestibular cortex, ventral intraparietal area, and medial superior temporal area. J. Neurosci. *31*, 3082–3094.

Chen, A., DeAngelis, G.C., and Angelaki, D.E. (2011b). Representation of vestibular and visual cues to self-motion in ventral intraparietal cortex. J. Neurosci. *31*, 12036–12052.

Chen, A., Deangelis, G.C., and Angelaki, D.E. (2013). Functional specializations of the ventral intraparietal area for multisensory heading discrimination. J. Neurosci. *33*, 3567–3581.

Chen, A., Gu, Y., Liu, S., DeAngelis, G.C., and Angelaki, D.E. (2016). Evidence for a causal contribution of macaque vestibular, but not intraparietal, cortex to heading perception. J. Neurosci. *36*, 3789–3798.

Churchland, A.K., Kiani, R., Chaudhuri, R., Wang, X.-J., Pouget, A., and Shadlen, M.N. (2011). Variance as a signature of neural computations during decision making. Neuron *69*, 818–831.

Deneve, S., Latham, P.E., and Pouget, A. (1999). Reading population codes: a neural implementation of ideal observers. Nat. Neurosci. *2*, 740–745.

Deneve, S., Latham, P.E., and Pouget, A. (2001). Efficient computation and cue integration with noisy population codes. Nat. Neurosci. *4*, 826–831.

Drugowitsch, J., DeAngelis, G.C., Klier, E.M., Angelaki, D.E., and Pouget, A. (2014). Optimal multisensory decision-making in a reaction-time task. eLife *3*, e03005.

Drugowitsch, J., DeAngelis, G.C., Angelaki, D.E., and Pouget, A. (2015). Tuning the speed-accuracy trade-off to maximize reward rate in multisensory decision-making. eLife *4*, e06678.

Ecker, A.S., Berens, P., Tolias, A.S., and Bethge, M. (2011). The effect of noise correlations in populations of diversely tuned neurons. J. Neurosci. *31*, 14272–14283.

Erlich, J.C., Brunton, B.W., Duan, C.A., Hanks, T.D., and Brody, C.D. (2015). Distinct effects of prefrontal and parietal cortex inactivations on an accumulation of evidence task in the rat. eLife *4*, e05457.

Fernández, C., and Goldberg, J.M. (1976). Physiology of peripheral neurons innervating otolith organs of the squirrel monkey. III. Response dynamics. J. Neurophysiol. *39*, 996–1008.

Fetsch, C.R., Rajguru, S.M., Karunaratne, A., Gu, Y., Angelaki, D.E., and Deangelis, G.C. (2010). Spatiotemporal properties of vestibular responses in area MSTd. J. Neurophysiol. *104*, 1506–1522.

Fetsch, C.R., Pouget, A., DeAngelis, G.C., and Angelaki, D.E. (2011). Neural correlates of reliability-based cue weighting during multisensory integration. Nat. Neurosci. *15*, 146–154.

Gold, J.I., and Shadlen, M.N. (2007). The neural basis of decision making. Annu. Rev. Neurosci. *30*, 535–574.

Goldman, M.S. (2009). Memory without feedback in a neural network. Neuron *61*, 621–634.

Gu, Y. (2018). Vestibular signals in primate cortex for self-motion perception. Curr. Opin. Neurobiol. *52*, 10–17.

Gu, Y., Watkins, P.V., Angelaki, D.E., and DeAngelis, G.C. (2006). Visual and nonvisual contributions to three-dimensional heading selectivity in the medial superior temporal area. J. Neurosci. *26*, 73–85.

Gu, Y., Angelaki, D.E., and Deangelis, G.C. (2008). Neural correlates of multisensory cue integration in macaque MSTd. Nat. Neurosci. *11*, 1201–1210.

Gu, Y., Fetsch, C.R., Adeyemo, B., Deangelis, G.C., and Angelaki, D.E. (2010). Decoding of MSTd population activity accounts for variations in the precision of heading perception. Neuron *66*, 596–609.

Gu, Y., Liu, S., Fetsch, C.R., Yang, Y., Fok, S., Sunkara, A., DeAngelis, G.C., and Angelaki, D.E. (2011). Perceptual learning reduces interneuronal correlations in macaque visual cortex. Neuron *71*, 750–761.

Gu, Y., Deangelis, G.C., and Angelaki, D.E. (2012). Causal links between dorsal medial superior temporal area neurons and multisensory heading perception. J. Neurosci. *32*, 2299–2313.

Guo, Z.V., Li, N., Huber, D., Ophir, E., Gutnisky, D., Ting, J.T., Feng, G., and Svoboda, K. (2014). Flow of cortical activity underlying a tactile decision in mice. Neuron *81*, 179–194.

Hanks, T.D., Kopec, C.D., Brunton, B.W., Duan, C.A., Erlich, J.C., and Brody, C.D. (2015). Distinct relationships of parietal and prefrontal cortices to evidence accumulation. Nature *520*, 220–223.

Huk, A.C., Katz, L.N., and Yates, J.L. (2017). The role of the lateral intraparietal area in (the study of) decision making. Annu. Rev. Neurosci. *40*, 349–372.

Katz, L.N., Yates, J.L., Pillow, J.W., and Huk, A.C. (2016). Dissociated functional significance of decision-related activity in the primate dorsal stream. Nature *535*, 285–288, advance online publication.

Kiani, R., Hanks, T.D., and Shadlen, M.N. (2008). Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. J. Neurosci. *28*, 3017–3029.

Knill, D.C., and Richards, W. (1996). Perception as Bayesian Inference (Cambridge University Press).

Kobak, D., Brendel, W., Constantinidis, C., Feierstein, C.E., Kepecs, A., Mainen, Z.F., Qi, X.-L., Romo, R., Uchida, N., and Machens, C.K. (2016). Demixed principal component analysis of neural population data. eLife *5*, e10989.

Körding, K.P., Beierholm, U., Ma, W.J., Quartz, S., Tenenbaum, J.B., and Shams, L. (2007). Causal inference in multisensory perception. PLoS ONE *2*, e943.

Lakshminarasimhan, K.J., Pouget, A., DeAngelis, G.C., Angelaki, D.E., and Pitkow, X. (2018). Inferring decoding strategies for multiple correlated neural populations. PLoS Comput. Biol. *14*, e1006371.

Laming, D.R.J. (1968). Information Theory of Choice-Reaction Times (Academic Press).

Laurens, J., Liu, S., Yu, X.-J., Chan, R., Dickman, D., DeAngelis, G.C., and Angelaki, D.E. (2017). Transformation of spatiotemporal dynamics in the macaque vestibular system from otolith afferents to cortex. eLife *6*, e20787.

Li, N., Daie, K., Svoboda, K., and Druckmann, S. (2016). Robust neuronal dynamics in premotor cortex during motor planning. Nature *532*, 459–464.

Licata, A.M., Kaufman, M.T., Raposo, D., Ryan, M.B., Sheppard, J.P., and Churchland, A.K. (2017). Posterior Parietal Cortex Guides Visual Decisions in Rats. J. Neurosci. *37*, 4954–4966.

Lisberger, S.G., and Movshon, J.A. (1999). Visual motion analysis for pursuit eye movements in area MT of macaque monkeys. J. Neurosci. *19*, 2224–2246.

Ma, W.J., Beck, J.M., Latham, P.E., and Pouget, A. (2006). Bayesian inference with probabilistic population codes. Nat. Neurosci. *9*, 1432–1438.

Mante, V., Sussillo, D., Shenoy, K.V., and Newsome, W.T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature *503*, 78–84.

Meister, M.L., Hennig, J.A., and Huk, A.C. (2013). Signal multiplexing and single-neuron computations in lateral intraparietal area during decision-making. J. Neurosci. *33*, 2254–2267.

Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., and Pouget, A. (2014). Information-limiting correlations. Nat. Neurosci. *17*, 1410–1417.

Najafi, F., and Churchland, A.K. (2018). Perceptual decision-making: a field in the midst of a transformation. Neuron *100*, 453–462.

Nikbakht, N., Tafreshiha, A., Zoccolan, D., and Diamond, M.E. (2018). Supralinear and supramodal integration of visual and tactile signals in rats: psychophysics and neuronal mechanisms. Neuron *97*, 626–639.e8.

O'Connell, R.G., Shadlen, M.N., Wong-Lin, K., and Kelly, S.P. (2018). Bridging neural and computational viewpoints on perceptual decision-making. Trends Neurosci. *41*, 838–852.

Park, I.M., Meister, M.L., Huk, A.C., and Pillow, J.W. (2014). Encoding and decoding in parietal cortex during sensorimotor decision-making. Nat. Neurosci. *17*, 1395–1403.

Pisupati, S., Chartarifsky, L., and Churchland, A.K. (2016). Decision activity in parietal cortex: leader or follower? Trends Cogn. Sci. *20*, 788–789.

Pitkow, X., Liu, S., Angelaki, D.E., DeAngelis, G.C., and Pouget, A. (2015). How can single sensory neurons predict behavior? Neuron *87*, 411–423.

Raposo, D., Sheppard, J.P., Schrater, P.R., and Churchland, A.K. (2012). Multisensory decision-making in rats and humans. J. Neurosci. *32*, 3726–3735.

Raposo, D., Kaufman, M.T., and Churchland, A.K. (2014). A category-free neural population supports evolving demands during decision-making. Nat. Neurosci. *17*, 1784–1792.

Ratcliff, R. (1978). A theory of memory retrieval. Psychol. Rev. *85*, 59.

Ratcliff, R., and McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. Neural Comput. *20*, 873–922.

Ratcliff, R., and Rouder, J.N. (1998). Modeling response times for two-choice decisions. Psychol. Sci. *9*, 347–356.

Ratcliff, R., and Smith, P.L. (2004). A comparison of sequential sampling models for two-choice reaction time. Psychol. Rev. *111*, 333–367.

Rodman, H.R., and Albright, T.D. (1987). Coding of visual stimulus velocity in area MT of the macaque. Vision Res. *27*, 2035–2048.

Rohe, T., and Noppeney, U. (2015). Cortical hierarchies perform Bayesian causal inference in multisensory perception. PLoS Biol. *13*, e1002073.

Roitman, J.D., and Shadlen, M.N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. J. Neurosci. *22*, 9475–9489.

Rossi-Pool, R., Zainos, A., Alvarez, M., Zizumbo, J., Vergara, J., and Romo, R. (2017). Decoding a decision process in the neuronal population of dorsal premotor cortex. Neuron *96*, 1432–1446.e1437.

Scott, B.B., Constantinople, C.M., Akrami, A., Hanks, T.D., Brody, C.D., and Tank, D.W. (2017). Fronto-parietal cortical circuits encode accumulated evidence with a diversity of timescales. Neuron *95*, 385–398.e5.

Seriès, P., Latham, P.E., and Pouget, A. (2004). Tuning curve sharpening for orientation selectivity: coding efficiency and the impact of correlations. Nat. Neurosci. *7*, 1129–1135.

Seung, H.S., and Sompolinsky, H. (1993). Simple models for reading neuronal population codes. Proc. Natl. Acad. Sci. USA *90*, 10749–10753.

Shadlen, M.N., and Newsome, W.T. (1996). Motion perception: seeing and deciding. Proc. Natl. Acad. Sci. USA *93*, 628–633.

Shadlen, M.N., and Newsome, W.T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. J. Neurophysiol. *86*, 1916–1936.

Shamir, M., and Sompolinsky, H. (2006). Implications of neuronal diversity on population coding. Neural Comput. *18*, 1951–1986.

Smith, A.T., Greenlee, M.W., DeAngelis, G.C., and Angelaki, D.E. (2017). Distributed visual–vestibular processing in the cerebral cortex of man and macaque. Multisens. Res. *30*, 91–120.

Song, S., Sjöström, P.J., Reigl, M., Nelson, S., and Chklovskii, D.B. (2005). Highly nonrandom features of synaptic connectivity in local cortical circuits. PLoS Biol. *3*, e68.

Song, H.F., Yang, G.R., and Wang, X.J. (2017). Reward-based training of recurrent neural networks for cognitive and value-based tasks. eLife *6*, 6.

Wang, X.J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. Neuron *36*, 955–968.

Wong, K.F., and Wang, X.J. (2006). A recurrent network mechanism of time integration in perceptual decisions. J. Neurosci. *26*, 1314–1328.

Zhong, L., Zhang, Y., Duan, C.A., Deng, J., Pan, J., and Xu, N.L. (2019). Causal contributions of parietal cortex to perceptual decision-making during stimulus categorization. Nat. Neurosci. *22*, 963–973.

Zhou, Y., and Freedman, D.J. (2019). Posterior parietal cortex plays a causal role in perceptual and categorical decisions. Science *365*, 180–185.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
All data	This paper	https://data.mendeley.com/datasets/b8ybw4shv3/1
Experimental Models: Organisms/Strains		
Rhesus macaque (Macaca mulatta)	Beijing Institute of Xieerxin Biology Resource, China	http://www.xexbio.com/
Software and Algorithms		
TEMPO Software	Reflective Computing	http://www.greatislandsoftware.com/ReflectiveComputing/Tempo.html
CED Spike2	Cambridge Electronic Design	http://ced.co.uk/products/spkovin
MATLAB	MathWorks	https://www.mathworks.com/
Other		
Tungsten electrodes	FHC	https://www.fh-co.com/product/metal-microelectrodes/
Single channel amplifier	BAK Electronics	https://www.bakelectronicsinc.com/
Riverbend Eye Tracking system	Riverbend Instruments	http://www.riverbendinst.com/
Motion platform MOOG MB-E-6DOF/12/1000KG	MOOG	https://www.moog.com

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Yong Gu (guyong@ion.ac.cn).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All animal procedures were approved by the Animal Care Committee of Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences and have been described previously in detail (Gu et al., 2006, 2008). Briefly, two male adult rhesus monkeys, Monkey P and Monkey M, weighing ~8 kg, were chronically implanted with a lightweight plastic ring for head restraint and a scleral coil for monitoring eye movements (Riverbend Instruments). During experiments, the monkey sat comfortably, with the head restrained, in a primate chair mounted on top of a custom-built virtual reality system.

METHOD DETAILS

Apparatus

The vestibular-visual virtual reality system consisted of a motion platform (MOOG MB-E-6DOF/12/1000KG) and an LCD screen (~30 cm of view distance and ~90° × 90° of visual angle; HP LD4201), presenting vestibular and visual motion stimuli to the monkey, respectively. The stimuli were controlled by customized C++ software and synchronized with the electrophysiological recording system by TEMPO (Reflective Computing, U.S.A).

To tune the synchronization between vestibular and visual stimuli, we rendered a virtual world-fixed crosshair on the screen while projected a second crosshair at the same place on the screen using a real world-fixed laser pen. When the platform was moving, we carefully adjusted a delay parameter in the C++ software (with 1 ms resolution) until the two crosshairs aligned precisely together all the time, as verified by a high-speed camera (Meizu Pro 5) and/or a pair of back-to-back mounted photodiodes. This synchronization procedure was repeated occasionally over the whole period of data collection.

Behavioral Tasks

Memory-guided Saccade Task

We used the standard memory-guided saccade task (Barash et al., 1991) to characterize and select LIP cells for recording in the main decision-making experiments. The monkey fixated at a central fixation point for 100 ms and then a saccade target flashed briefly

(500 ms) in the periphery. The monkey was required to maintain fixation during the delay period (1000 ms) until the fixation point extinguished and then saccade to the remembered target location within 1000 ms for a liquid reward. For all tasks in the present study, at any time when there existed a fixation point, trials were aborted immediately if the monkey's gaze deviated from a 2° × 2° electronic window around the fixation point.

Multisensory Heading Discrimination Task

In the main experiments, we trained the monkeys to report their direction of self-motion in a two-alternative forced-choice heading discrimination task (Gu et al., 2008) (Figure 1). The monkey initiated a trial by fixating on a central, head-fixed fixation point, and two choice targets then appeared. The locations of the two targets were determined case-by-case for each recording session (see below). After fixating for a short delay (100 ms), the monkey then began to experience a fixed-duration (1.5 s) forward motion in the horizontal plane with a small leftward or rightward component relative to straight ahead. The animals were required to maintain fixation during the presentation of the motion stimuli. At the end of the trial, the motion ended, and the monkey was required to maintain fixation for another 300–600 ms random delay (uniformly distributed) until the fixation point disappeared, at which point the monkey was allowed to make a saccade choice toward one of the two targets to report his perceived heading direction (left or right).

Across trials, nine heading angles (±8°, ± 4°, ± 2°, ± 1°, and 0°) and three cue conditions (vestibular, visual, and combined) were jointly interleaved, resulting in 27 unique stimulus conditions, each of which was repeated 15 ± 3 (median ± m.a.d.) times per one session. In a vestibular or a visual trial, heading information was solely provided by inertial motion (real movement of the motion platform) or optic flow (simulated movement through a star field on the display), respectively, whereas in a combined trial, congruent vestibular and visual cues were provided synchronously. To maximize the behavioral benefit of cue integration, we balanced the monkey's performance under the vestibular and the visual conditions by manipulating the motion coherence of the optic flow (the percentage of dots that moved coherently). The visual coherence was 12% and 8% for monkey P and M, respectively.

To ensure that the reliabilities of sensory cues varied throughout each trial, we used Gaussian-shape, rather than constant, velocity profiles for all motion stimuli. In the main experiments, the Gaussian profile had a displacement $d = 0.2$ m and a standard deviation $\sigma = 210$ ms (half duration at about 60% of the peak velocity), resulting in a peak velocity $v_{max} = 0.37 \mathrm{m/s}$ and a peak acceleration $a_{max} = 1.1\ \mathrm{m/s^2}$. In the experiment where we sought to independently vary the peak times of velocity and acceleration (Figure 3), two additional sets of motion parameters were used. For the narrow-speed profile, $d = 0.10$ m, $\sigma = 150$ ms, $v_{max} = 0.26 \mathrm{m/s}$, and $a_{max} = 1.1\ \mathrm{m/s^2}$; for the wide-speed profile, $d = 0.25$ m, $\sigma = 330$ ms, $v_{max} = 0.31 \mathrm{m/s}$, and $a_{max} = 0.6 \mathrm{m/s^2}$.

Electrophysiology

We carried out extracellular single-unit recordings as described previously (Gu et al., 2008) from four hemispheres in two monkeys. For each hemisphere, reliable area mapping was first achieved through cross-validation between structural MRI data and electrophysiological properties, including transition patterns of gray/white matter along each penetration, sizes of visual receptive/response field, strengths of spatial tuning to visual and vestibular heading stimuli, and activities in the memory-guided saccade task. Based on the mapping results, Area LIP was registered by its spatial relationships with other adjacent areas (VIP, Area 5, MSTd, etc.), its weak sensory encoding of heading information, and its overall strong saccade-related activity (Figures S1A–S1C). Our recording sites located in the ventral division of LIP, extending from 7–13 mm lateral to the midline and −5 mm (posterior) to +3 mm (anterior) relative to the interaural plane.

Once we encountered a well-isolated single unit in LIP, we first explored its response field (RF) by hand (using a flashing patch) and then examined its electrophysiological properties using the memory-guided saccade task. The saccade target in each trial was randomly positioned at one of the 8 locations 45° apart on a circle centered on the fixation point (5°–25° radius, optimized according to the cell's RF location). We calculated online the memory-saccade spatial tuning for three response epochs: (1) visual response period, 75–400 ms from target onset; (2) delay period, 25–900 ms from target offset; and (3) presaccadic period, 200–50 ms before the saccade onset (Figures S1D–S1G). The cell's spatiotemporal tunings were used to refine its RF location (via vector sum) and to determine its inclusion in the subsequent decision-making task. Since the decision-related activity of LIP neurons cannot be strongly predicted by the persistent activity during the delay period alone (Meister et al., 2013) (Figure S3C), we adopted a wider cell selection criterion than conventionally used, in which we included cells that have significant spatial selectivity for *any* of the three response epochs (Meister et al., 2013) (one-way ANOVA, $p < 0.05$, 3–5 repetitions). If the cell met this criterion, then we recorded its decision-related activity while engaging the monkey in the main multisensory decision-making task, with the two choice targets being positioned in its RF and 180° opposite to its RF, respectively.

Although we collected data from a relatively broad sample of LIP neurons, we nevertheless had two sampling biases during this process. First, we were biased toward cells with strong persistent activity so that our multisensory data could be better compared with previous unisensory data in the decision-making literature, where in most cases only these cells were recorded. Second, we were biased toward cells with RF close to the horizontal line through the fixation point. Unlike the classical random dot stimuli whose motion direction on the fronto-parallel plane can be aligned with the cell's RF (and the choice targets) session by session, our self-motion stimuli were always on the horizontal plane and thus were not adjustable according to the cell's RF on the fronto-parallel plane. As a result, the subjects had to make an additional mapping from their perceived heading directions (always left or right) to the choice targets (often inclined, and in extreme cases, up or down). Therefore, to make the task more intuitive to the monkeys

and to minimize the potential influence of this mapping step on neural activity, we discarded a cell if the angle between the horizontal line and the line connecting the fixation point to its RF exceeded 60°, although we observed little change in monkeys' behavior even when this angle approached 80°.

QUANTIFICATION AND STATISTICAL ANALYSIS

Psychophysics

To quantify the behavioral performance for both the monkeys and the model in the multisensory decision-making task, we constructed psychometric curves by plotting the proportion of "rightward" choices as a function of heading (Figure 1C) and fitted them with cumulative Gaussian functions (Gu et al., 2008). The psychophysical threshold for each cue condition was defined as the standard deviation of their respective Gaussian fit. The Bayesian optimal prediction of psychophysical threshold under the combined condition $\sigma_{\text{prediction}}$ was solved from the inverse variance rule (Knill and Richards, 1996)

$$\frac{1}{\sigma^2_{\text{prediction}}} = \frac{1}{\sigma^2_{\text{vestibular}}} + \frac{1}{\sigma^2_{\text{visual}}}$$

where $\sigma_{\text{vestibular}}$ and σ_{visual} represent psychophysical thresholds under the vestibular and visual conditions, respectively.

Choice-related neural activities

We constructed peri-stimulus time histograms (PSTHs) for two epochs of interest in a trial, the decision formation epoch and the saccade epoch, by aligning raw spike trains to the stimulus onset and the saccade onset, respectively. Firing rates were computed in non-overlapping 10-ms bins and smoothed over time by convolving with a Gaussian kernel ($\sigma = 50$ ms). Unless otherwise noted, only correct trials were used in the following analyses, except for the ambiguous 0° heading where we included all complete trials.

To illustrate the choice-related activity of a cell, we grouped the trials according to the monkey's choice, i.e., trials ending up with a saccade toward the cell's RF (IN trials) versus trials ending up with a saccade away from the cell's RF (OUT trials), and computed the averaged PSTHs of these two groups of trials for each cue condition (Figure 2A). When averaged across cells, each cell's PSTHs were normalized such that the cell's overall firing rate had a dynamic range of [0, 1] (Figure 3B). To quantify the strength of choice signals and better visualize ramping activities, we calculated choice divergence (Raposo et al., 2014) for each 10-ms time bin and for each cue condition using receiver operating curve (ROC) analysis (Figure 2B). Choice divergence ranged from −1 to 1 and was defined as $2 \times (\text{AUC} - 0.5)$, where AUC represents the area under the ROC curve derived from PSTHs of IN and OUT trials. To capture the onset of choice signals, we computed a divergence time defined as the time of the first occurrence of a 250-ms window (25 successive 10-ms bins) in which choice divergence was consistently and significantly larger than 0 (Figures 3C and 3F). Cells that did not have significant choice encoding in certain condition(s) will not contribute to the divergence time analysis under that condition(s). The statistical significance of choice divergence ($p < 0.05$, relative to the chance level of 0) was assessed by two-tailed permutation test (1000 permutations). We also calculated a grand choice divergence (grand CD) which ignored temporal information and used all the spikes in the decision formation epoch (0–1500 ms from the stimulus onset). The same permutation test was performed on the grand CD to determine whether a cell had overall significant choice signals for a certain cue condition (for example, in Figure 2D).

Linear Fitting of Mean Firing Rates

We fitted a linear weighted summation model to predict neural responses under the combined condition with those under the single cue conditions, using (Gu et al., 2008)

$$r_{\text{combined}} = w_{\text{vestibular}} r_{\text{vestibular}} + w_{\text{visual}} r_{\text{visual}} + C$$

where C is a constant, and r_{combined}, $r_{\text{vestibular}}$, and r_{visual} are mean firing rates across a trial (0–1500 ms from stimulus onset) for the three cue conditions, respectively. The weights for single cue conditions, $w_{\text{vestibular}}$ and w_{visual}, were determined by the least-squares method and plotted against each other to evaluate the heterogeneity of choice signals in the population for both LIP data and the model (Figure S5F).

Fisher Information Analysis

To compute Fisher information (Seung and Sompolinsky, 1993), the full covariance matrix of the population responses is needed, but this requires simultaneously recording from hundreds of neurons, which is not accessible to us yet. Instead, we calculated the shuffled Fisher information, which corresponds to the information in a population of neurons in which correlations have been removed (typically via shuffling across trials, hence the name). Shuffled Fisher information is given by (Seriès et al., 2004; Gu et al., 2010):

$$I_{\text{shuffled}} = \sum_{i=1}^{N} \frac{f_i'^2}{\sigma_i^2} \quad (1)$$

where N is the number of neurons in the population; for the ith neuron, f_i' denotes the derivative of its local tuning curve, and σ_i^2 denotes the averaged variance of its responses around 0° heading. The tuning curve f_i was constructed from both correct and wrong trials grouped by heading angles, using spike counts in 250-ms sliding windows (advancing in 10-ms steps), and its derivative f_i' was obtained from the slope of a linear fit of f_i against headings. The variance σ_i^2 was computed for each heading angle and then averaged. To estimate the standard errors of I_{shuffled}, we used a bootstrap procedure in which random samples of neurons were drawn from the population by resampling with replacement (1000 iterations). To compare the experimental data with the model, we repeated all the above steps on artificial LIP neurons in the model M2 and M3 (see below), with the inter-neuronal noise correlation being ignored as well (Figure 7).

Some caveats are noteworthy when interpreting the Fisher information results. First, since the slope of tuning curve f' is squared in the right-hand side of Equation 1, the Fisher information will always be non-negative regardless of the sign of f'. As a result, even when the motion speed was zero at the beginning of a trial, the population Fisher information already had a positive value because of the noisy tuning curves during that period. Second, since we ignored inter-neuronal noise correlations, I_{shuffled} is most likely very different from the true Fisher information and thus its value is arbitrary (Seriès et al., 2004). Nonetheless, if we assume the noise correlation structure of LIP population is similar across cue conditions, we can still rely on the qualitative temporal evolution of I_{shuffled} to appreciate how multisensory signals are accumulated across time and cues in LIP. Finally, the Fisher information of MSTd neurons appeared much lower in the vestibular and the combined conditions than in the visual condition (Figure 7D). The lower vestibular information is because fewer MSTd neurons show spatial tuning in response to vestibular stimuli (Gu et al., 2006). The lower combined information, however, results from the fact that about half of the multisensory neurons in MSTd have opposite tuning curves to visual and vestibular inputs (“opposite cells”) (Gu et al., 2006, 2008). Therefore, the key point of Figure 7D is the temporal dynamics of I_{shuffled}, rather than its amplitude.

Network Simulation of iIPPC Framework

The responses of visual and vestibular neurons closely approximate iIPPC

As mentioned previously, an important assumption of iIPPC is that the amplitude of the sensory tuning curves be proportional to the nuisance parameters (in our case visual speed and the absolute value of vestibular acceleration) (Beck et al., 2008). To check whether this is the case for the visual neurons, we analyzed the spatiotemporal tuning curves of neurons in area MSTd (data from (Gu et al., 2006)). We noticed that, for some neurons, the average tuning curves are not fully consistent with the iIPPC assumption (Figure S4A). Briefly, the mean firing rate of an MSTd neuron at time t in response to a visual stimulus with heading θ can be well captured by

$$f(\theta,t) = v(t)(A\exp[K(\cos(\theta-\theta_i)-1)]-C)+B \quad (2)$$

where θ_i denotes the preferred heading of the neuron i and $v(t)$ is the velocity profile; A, K, C, and B correspond to the amplitude, the width, the null inhibition, and the baseline of its tuning curve, respectively. The iIPPC framework requires the $v(t)$ term to be separable, namely, $f(\theta,t) = h(\theta)g(v(t))$, where $h(\theta)$ is a pure spatial component and $g(v(t))$ is a multiplicative gain function (Ma et al., 2006; Beck et al., 2008). In Equation 2, this requirement is equivalent to $C = 0$ and $B = 0$, however, we found that some MSTd neurons often had non-zero baselines ($C > 0$ and $B > 0$). This will be harmful to the optimality of the iIPPC framework because, for example, when $v(t) = 0$ (and thus the sensory reliability is zero), MSTd neurons still tend to generate background spikes, which will bring nothing but noise into the simply summed population activity of downstream areas in an iIPPC network.

To estimate the information loss due to this deviation, we simulated a population of MSTd neurons with heterogeneous spatiotemporal tuning curves similar to what has been found experimentally (Gu et al., 2006). We calculated the information that can be decoded from the population by a series of optimal decoders I_{optimal} and that can be recovered by the iIPPC solution I_{iIPPC}. We assumed that the population responses in MSTd contains differential correlations (Moreno-Bote et al., 2014) such that the discrimination threshold of an ideal observer of MSTd activity was of the same order as the animal’s performance. Under such conditions, we found that the information loss $(I_{\text{optimal}} - I_{\text{iIPPC}})/I_{\text{optimal}}$ was around 5%. Detailed calculations of information loss are provided in a later section. Therefore, the population response of MSTd neurons provide a close approximation to an iIPPC, in the sense that simply summing the activity of MSTd neurons over time preserves 95% of the information conveyed by these neurons.

We also checked whether the iIPPC assumption holds in the case of vestibular neurons. Equation 2 above still provides a good approximation to vestibular tuning curves, except that C is close to zero for most neurons (Laurens et al., 2017), in which case the information loss is even less pronounced.

Network Model Implementing the iIPPC solution (Model M1)

We extended a previous iIPPC network model for unisensory decision making (Beck et al., 2008) to our multisensory decision-making task (Figure 4). Two sensory layers, the vestibular layer and the visual layer, contained 100 linear-nonlinear-Poisson (LNP) neurons with bell-shape tuning curves to the heading direction (Equation 2). For the ith neuron in the vestibular or visual layer, the probability of firing a spike at time step $[t_n - \delta t, t_n]$ was given by

$$p(r_i^{\cdot}(t_n) = 1) = [\delta t(g.(t)(A\exp[K(\cos(\theta-\theta_i)-1)]-C)+B)+n_i]^+ \quad (3)$$

where $\cdot \in \{\text{VEST}, \text{VIS}\}$, A, K, C, B, θ, and θ_i have the same meanings as in Equation 2, n_i is a correlated noise term, and $[\cdot]^+$ is the threshold-linear operator: $[x]^+ = \max(x, 0)$. The spatial tuning was gain-modulated by a time-dependent function $g_\cdot(t)$, which modeled the reliability of the sensory evidence and took the form

$$g_{\text{VEST}}(t) = c_{\text{VEST}}|\widehat{a}(t)|, \quad g_{\text{VIS}}(t) = c_{\text{VIS}}\widehat{v}(t) \tag{4}$$

in which $\widehat{a}(t)$ and $\widehat{v}(t)$ are the same acceleration and velocity profiles as the experiments but with the maximum values normalized to 1, respectively, whereas c_{VEST} and c_{VIS} are scaling parameters used to control the signal-to-noise ratio of sensory inputs and to balance the behavior performance between the two cue conditions like in the experiments. The noise n_i in Equation 3 was generated by convolving independent Gaussian noise with a circular Gaussian kernel,

$$n_i = \sum_j A_\eta \exp(K_\eta(\cos(\theta_i - \theta_j) - 1))\eta_j$$

where $\eta_j \sim i.i.d.N(0, 1)$, and A_η and K_η were set to 10^{-5} and 2, respectively. Other parameters we used were: $A = 60$ Hz, $K = 1.5$, $C = 10$ Hz, $B = 20$ Hz, $c_{\text{VEST}} = c_{\text{VIS}} = 2.4$, $\delta t = 1$ ms. Note that in Equation 3, the gain $g_\cdot(t)$ cannot be factored out because $B > 0$, which is the same case as in MSTd (Equation 2). Accordingly, the neural code of M1's sensory layers is not exact ilPPC (Beck et al., 2008). However, it is still a close approximation to ilPPC, since we have shown in the previous section that MSTd is 95% ilPPC-compatible.

The two sensory layers then projected to 100 LNP neurons in the integrator layer. The integrator layer summed the sensory responses across both cues and time,

$$m_i^{\text{INT}}(t_{n+1}) = m_i^{\text{INT}}(t_n) + g_{\text{stim}}(t_n)\left(\sum_j W_{ij}^{\text{INTVEST}} r_j^{\text{VEST}}(t_n) + \sum_j W_{ij}^{\text{INTVIS}} r_j^{\text{VIS}}(t_n)\right) \tag{5}$$

where m_i^{INT} denotes the membrane potential proxy of neuron i, W_{ij}^{INTVEST} and W_{ij}^{INTVIS} are matrices for the feedforward weights from the vestibular and visual layer to the integrator layer, respectively, and $g_{\text{stim}}(t_n)$ is an attentional gain factor (see below). Note that we ignored the issue of how neural circuits perform perfect integration and just assumed that they do. We could have simulated one of the known circuit solutions to this problem (Goldman, 2009), but this would not have affected our results, while making the simulation considerably more complicated. Also note that separating the integrator layer from the LIP layer is not critical either. We did so to reflect the fact that current experimental data suggest that LIP may not be the layer performing the integration *per se*, but may only reflect the results of this integration (Katz et al., 2016). Nonetheless, whether or not the integration of multisensory evidence actually takes place in LIP, we can still use our empirical LIP data to test against the network model.

The feedforward connections $W_{ij}^{\text{INT}\cdot}$ map the negative and positive heading directions onto the two saccade targets, i.e., neurons preferring -90° and $+90^\circ$ in the integrator layer, respectively, by

$$W_{ij}^{\text{INT}\cdot} = a \exp\left(k\left(\cos\left(\theta_i^{\text{INT}} - \widehat{\theta}\right) - 1\right)\right)\left|\sin\left(\theta_j^{\cdot}\right)\right|$$

in which a step function $\widehat{\theta}$ controls the mapping,

$$\widehat{\theta} = \begin{cases} -\pi/2, \text{ if } \theta_j^{\cdot} \le 0 \\ \pi/2, \text{ if } \theta_j^{\cdot} > 0 \end{cases}$$

(Figure 4B, bottom). We used $a = 20$ and $k = 4$ in our simulations. After the linear step, the membrane potential proxy was used to determine the probability of the ith integrator neuron firing a spike between times t_n and $t_n + \delta t$,

$$p\left(r_i^{\text{INT}}(t_n) = 1\right) = \left[m_i^{\text{INT}}(t_n)\right]^+.$$

Finally, the LIP layer received excitatory inputs from the integrator layer, together with visual inputs triggered by the two saccade targets (sent from the target layer). In addition, there were also lateral connections in LIP to prevent saturation. In the linear step, the membrane potential proxy of the ith LIP neuron followed

$$m_i^{\text{LIP}}(t_{n+1}) = \left(1 - \frac{\delta t}{\tau}\right)m_i^{\text{LIP}}(t_n) + \frac{1}{\tau}\left(\sum_j W_{ij}^{\text{LIPINT}} r_j^{\text{INT}}(t_n) + \sum_j W_{ij}^{\text{LIPTARG}} r_j^{\text{TARG}}(t_n) + \sum_j W_{ij}^{\text{LIP}} r_j^{\text{LIP}}(t_n)\right) \tag{6}$$

where the time constant, τ, was set to 100 ms; W_{ij}^{LIPINT} and W_{ij}^{LIPTARG} are weight matrices for the feedforward connections from the integrator layer (Figure 4B, middle) and the target layer to the LIP layer, respectively, and W_{ij}^{LIP} is the matrix for the recurrent connections within LIP (Figure 4B, top). We used translation-invariant weights for all these connections,

$$W_{ij} = a \exp(k(\cos(\theta_i - \theta_j) - 1)) + b.$$

For W_{ij}^{LIPINT}, we used $a = 15, k = 10, b = -3.6$; for W_{ij}^{LIPTARG}, we used $a = 8, k = 5, b = 0$; and for W_{ij}^{LIP}, we used $a = 5, k = 10, b = -3$. The term $r_j^{\text{TARG}}(t_n)$ in Equation 6 denotes the visual response of the jth neuron in the target layer induced by the two saccade targets,

$$p\left(r_j^{\mathrm{TARG}}(t_n)=1\right)=s_{\mathrm{targ}}(t_n)\sum_{m=1}^{2}p_{\mathrm{targ}}\exp\left(k_{\mathrm{targ}}\left(\cos\left(\theta_j^{\mathrm{TARG}}-\theta_m\right)-1\right)\right)$$

where $\theta_1=-\pi/2$ and $\theta_2=\pi/2$, $p_{\mathrm{targ}}=0.050$, and $k_{\mathrm{targ}}=4$. The term $s_{\mathrm{targ}}(t_n)$ modeled the saliency of the targets: $s_{\mathrm{targ}}(t_n)=1$ before stimulus onset and $s_{\mathrm{targ}}(t_n)=0.6$ afterward.

After the linear step done in Equation 6, the probability of observing a spike from the *i*th LIP neuron for the next time step was given by, again,

$$p\left(r_i^{\mathrm{LIP}}(t_{n+1})=1\right)=\left[m_i^{\mathrm{LIP}}(t_{n+1})\right]^{+}. \tag{7}$$

Decision Bound and Action Selection

To let the model make decisions, we endowed it with a stopping bound such that the evidence integration terminated when the peak activity in the LIP layer reached a threshold value. This mechanism generates premature decisions in our fixed duration task, which have been observed in the previous experiments (Kiani et al., 2008) as well as ours (see the main text). Specifically, once the firing rate of any neuron in the LIP layer (determined from Equation 7) exceeded $\Theta^{\cdot}=37$ Hz for a vestibular or a visual trial and $\Theta^{\mathrm{COMB}}=42$ Hz for a combined trial, we blocked the sensory inputs to the integrator layer by setting the gain factor in Equation 5 to zero:

$$g_{\mathrm{stim}}(t_n)=\begin{cases}1, if\ t_n<t_\Theta\\ 0, if\ t_n\geq t_\Theta\end{cases}$$

where t_Θ denotes the time of bound crossing. The instantaneous population activity at this time point $\mathbf{r}^{\mathrm{LIP}}(t_\Theta)$ was then used to determine the model's choice, while the network dynamics continued to evolve until the end of the 1.5 s trial.

To read out the model's choice, we trained a linear support vector machine (SVM) to classify the heading direction from $\mathbf{r}^{\mathrm{LIP}}(t_\Theta)$. We ran the network for 100 trials, used $\mathbf{r}^{\mathrm{LIP}}(t_\Theta)$ in 30 trials to train the SVM, and then applied the SVM on the remaining 70 trials to make decisions and generate psychometric functions of the model (with bootstrap 1000 times, Figure 5A and Figure S5C). The SVM acts like (or even outperforms) a local optimal linear estimator (LOLE) trained by gradient descent (Seriès et al., 2004). Importantly, such decoders could be implemented with population codes in a biologically realistic point attractor network tuned for optimal action selection in a discrimination task (Deneve et al., 1999), which could correspond to downstream areas such as the motor layer of the superior colliculus (Beck et al., 2008).

Heterogeneous ilPPC Network (M2)

In model M2, we generalized the homogeneous ilPPC network described above (model M1) to a heterogeneous one. Instead of taking perfect sums like in model M1, neurons in the integrator layer of the model computed random linear combinations of vestibular and visual inputs (see Figure S5B). It is indeed been widely shown that integration weights *in vivo* are heterogeneous and are well-captured by "long-tailed" lognormal distributions (see for example (Song et al., 2005)). To simulate this in M2, we drew each synaptic weight w_{M2} in M2 from a lognormal distribution

$$p(w_{\mathrm{M2}}=x)=\frac{1}{\sqrt{2\pi}\sigma x}\exp\left(-\frac{(\log x-\mu)^2}{2\sigma^2}\right) \tag{8}$$

where μ and σ were chosen such that the expectation $e(w_{\mathrm{M2}})$ and the standard deviation $s(w_{\mathrm{M2}})$ of w_{M2} were both equal to its counterpart synaptic weight w_{M1} in model M1:

$$e(w_{\mathrm{M2}})=s(w_{\mathrm{M2}})=w_{\mathrm{M1}}.$$

The parameters μ and σ in Equation 8 were related to e and s through

$$\mu=\log\left(e^2\Big/\sqrt{e^2+s^2}\right)$$
$$\sigma=\sqrt{\log(s^2/e^2+1)}.$$

If $w_{\mathrm{M1}}<0$, a negative sign was added to the resulting w_{M2}, since lognormal distributions are always non-negative.

Network with Short Integration Time Constant (M3)

We also simulated a sub-optimal model M3 in which the network does not integrate evidence over time. This was done by replacing Equation 5 with

$$m_i^{\mathrm{INT}}(t_{n+1})=\left(1-\frac{\delta t}{\tau}\right)m_i^{\mathrm{INT}}(t_n)+\frac{1}{\tau}g_{\mathrm{stim}}(t_n)\left(\sum_j W_{ij}^{\mathrm{INTVEST}}r_j^{\mathrm{VEST}}(t_n)+\sum_j W_{ij}^{\mathrm{INTVIS}}r_j^{\mathrm{VIS}}(t_n)\right)$$

where $\tau=100$ ms and other terms are the same as in Equation 5.

Linear Reproduction of M1 Response

To test whether the responses of the optimal and homogeneous model M1 can be linearly reproduced from responses of M2, M3, and the experimental data, we first calculated the "optimal traces" from M1 (Figure 6A), using

$$\Delta PSTH^{\bullet}_{\mathrm{M1}} = <PSTH^{\bullet,+}_{M1,i}> - <PSTH^{\bullet,-}_{M1,i}>$$

Where $\bullet$ denotes three cue conditions (vestibular, visual, and combined), $PSTH^{\bullet,+}_{M1,i}$ and $PSTH^{\bullet,-}_{M1,i}$ denote averaged PSTH for the ith LIP unit in the network M1 when the network makes correct choices toward the neuron's preferred direction and null direction, respectively, and $<\bullet>$ denotes averaging across cells. To mimic the experimental procedure, only cells whose preferred directions were close to $\pm 90^{\circ}$ (with deviations less than 20°) were used (yellow shaded areas in Figure 5B). Similarly, we extracted single cell activities from M2, M3, the LIP data, and the MSTd data (Gu et al., 2006)

$$\Delta PSTH^{\bullet}_{*,j} = PSTH^{\bullet,+}_{*,j} - PSTH^{\bullet,-}_{*,j}$$

where $*\in\{\mathrm{M2}, \mathrm{M3}, \mathrm{LIP\ data}, \mathrm{MSTd\ data}\}$. Then we optimized sets of linear weights $\mathbf{w}_*$ to minimize the cost function

$$E_* = \sum_{\bullet}\sum_{n}\left(\Delta PSTH^{\bullet}_{\mathrm{M1}}(t_n) - \sum_i w_{*,i}\Delta PSTH^{\bullet}_{*,i}(t_n)\right)^2 \tag{9}$$

where, for example, $w_{\mathrm{LIP},i}$ represents the weight of the neuron i in the LIP data when a downstream area reads out LIP dynamics linearly to reproduce the optimal traces. To reduce overfitting, we partitioned the data into two subsets along time by randomly assigning the time bins into two sets, one for fitting (T_{fit}) and the other for validating (T_{valid}). During fitting, when the validating error $E_{*,t_n\in T_{\mathrm{valid}}}$ started increasing, we stopped the iteration, a procedure known as early stopping. The fitting results are shown in Figure 6C–F. Note that the $\Delta PSTH$s in the cost function (Equation 9) grouped all the heading angles together. The results were qualitatively similar when the cost function included error terms calculated from each heading angle separately, i.e.,

$$E_* = \sum_{\bullet}\sum_{n}\sum_{|h|}\left(\Delta PSTH^{\bullet,|h|}_{\mathrm{M1}}(t_n) - \sum_i w_{*,i}\Delta PSTH^{\bullet,|h|}_{*,i}(t_n)\right)^2 \tag{10}$$

where $|h|$ denotes the absolute value of heading angle (0°, 1°, 2°, 4°, 8°). The reconstructions of M1 traces with LIP activities using Equation 10 are shown in Figure S6A.

To assess the robustness of the linear reconstruction, we randomly subsampled the same number of neurons (n = 50, without replacement) from the four datasets, performed the linear fitting, and repeated this procedure for 1000 times. The mean squared error and the distribution of readout weights of the fittings are shown in Figure 6G and H. To examine whether only a small fraction of cells contributed heavily to the fittings or whether the majority of cells did, we compared the distributions of weights from the four datasets with the distribution of weights from a random linear decoder. To do so, for each subsampling, we also generated a set of random readout weights from a rectified Gaussian distribution (Figure 6H, black curve) and computed the kurtosis of the distribution of weights from the random decoder as well as those from the four datasets (Figure 6I). The p values were derived from the empirical subsampling distributions (two-tailed).

Estimating the Information Loss of ilPPC Solution with MSTd-like Neural Population

As discussed in the main text, a prerequisite of the ilPPC solution being optimal is the spatiotemporal separability of sensory tuning curves. In practice, this requirement can be loosened by allowing a time-independent baseline to exist in the right-hand side of Equation 2, i.e., $B>0$ but $C=0$, since this baseline term can be readily removed without loss of information by, for example, a simple rectified linear unit (ReLU) layer. A recent study shows that the tuning curves of most vestibular neurons fall into this category (Laurens et al., 2017). However, our analysis of the visual responses of MSTd neurons (n = 195, data from Gu et al., 2006) suggests that this is not the case in MSTd (Figure S4A, upper panel). The downward shift of firing rates around the non-preferred directions points to a time-dependent baseline component in the visual tunings, i.e., $B>0$ and $C>0$ in Equation 2, which cannot be simply eliminated with a layer of ReLU units. This implies that summing the activity of MSTd neurons over time will necessarily result in an information loss though the amplitude of this loss is unclear.

To estimate this information loss, we simulated a heterogeneous MSTd population with baseline-changing tuning curves and computed the proportion of encoded information that can be recovered by the ilPPC solution (i.e., a simple temporal integration). Specifically, we modeled the mean firing rate of neuron i in response to a stimulus with heading θ at time t by

$$f_i(\theta,t) = cv(t)(A_i\exp[K_i(\cos(\theta-\theta_{pref,i})-1)] - \beta_i B_i) + B_i,\quad i=1,2,\ldots,N \tag{11}$$

where $c\in[0,1]$ represents the motion coherence of the visual stimulus, $v(t)$ is the Gaussian velocity profile normalized to $[0,1]$, A_i, K_i, and B_i controls the peak, the width, and the baseline of the tuning curve, respectively, and $\beta_i\in[0,1]$ controls how strong the baseline varies with time. The preferred headings, $\theta_{pref,i}$, were drawn uniformly from $(-\pi,\pi]$. To account for the heterogeneous tuning properties of real MSTd neurons, we sampled A_i, B_i, β_i, and the tuning width (full width at the half maximum, or $FWHM_i$) from Gamma distributions with the following mean and standard deviation values (β_i larger than 1 were truncated to 1):

	A	B	$FWHM$	β
Mean	50 Hz	20 Hz	125°	0.6
Std.	30 Hz	20 Hz	50°	0.4

Then we found K_i by $K_i = (\log(0.5)/\cos(FWHM_i/2 - 1))$. These tuning parameters were chosen so that both the averaged and the single cell spatiotemporal tuning curves were similar to the MSTd data at the 100% visual coherence (Figure S4A, lower panel). Note that we used 10% visual coherence ($c = 0.1$ in Equation 11) for the following calculations, a value close to what we used in the present study.

Next, we introduced correlated noise into this heterogeneous population via a stimulus- and time-dependent covariance matrix $\mathbf{\Sigma}(\theta, t)$,

$$\Sigma_{ij}(\theta, t) = c_{ij}\sqrt{f_i(\theta, t)f_j(\theta, t)} + \frac{\epsilon}{v(t)}\frac{\partial}{\partial\theta}f_i(\theta, t)\frac{\partial}{\partial\theta}f_j(\theta, t). \tag{12}$$

The first term of the right-hand side is exponentially decaying pairwise correlations with a Fano factor = 1 and correlation coefficients

$$c_{ij} = (1-\rho)\delta_{ij} + \rho\exp[\kappa_c(\cos(\theta_{pref,i} - \theta_{pref,j}) - 1)] \tag{13}$$

where we used $\rho = 0.1$ and $\kappa_c = 2$, and δ_{ij} is the Kronecker delta; the second term represents correlations proportional to the product of the derivatives of the tuning curves, referred to as differential correlation (Moreno-Bote et al., 2014). The strength of the differential correlation is controlled by ϵ, and the factor $1/v(t)$ is used to make the two terms of the right-hand side of Equation 12 have the same dependence on $v(t)$.

Having modeled both the tuning curves $\mathbf{f}(\theta, t)$ and the covariance matrix $\mathbf{\Sigma}(\theta, t)$, we set out to estimate the information. Without loss of generality, we considered information around $\theta = 0$ and we dropped the term θ hereafter for clarity. We focused on two kinds of information: 1), the amount of total information that can be decoded by a series of locally optimal linear estimators (LOLEs) optimized for each time t, denoted I_{optimal}, and 2), the amount of information that can be recovered by the ilPPC solution, that is, by a single LOLE acting on the linear summation of spikes over the entire trial, denoted I_{ilPPC}. Thus for I_{optimal}, by definition,

$$I_{\text{optimal}} = \int_0^T \dot{I}_{\text{optimal}}(t)dt \tag{14}$$

where T is the trial duration (2 s) and $\dot{I}_{\text{optimal}}(t)$ is the optimal information *rate* at time t. For LOLEs, $\dot{I}_{\text{optimal}}(t)$ is equal to the linear Fisher information (Abbott and Dayan, 1999; Beck et al., 2011; Moreno-Bote et al., 2014),

$$\dot{I}_{\text{optimal}}(t) = \mathbf{f}'(t)^{\mathrm{T}}\mathbf{\Sigma}^{-1}(t)\mathbf{f}'(t). \tag{15}$$

To obtain $\mathbf{f}'(t)$, we first computed the derivatives of tuning curves with respect to θ around $\theta = 0$ at time t, using Equation 1, and then multiplied the derivatives by a factor of 2. We did the second step because it has been shown that, in MSTd, the slopes of the local tuning curves are approximately two times larger than those derived from the global tuning curves (Gu et al., 2010). Note, however, that this local sharpening of tuning curves will not affect the covariance $\mathbf{\Sigma}$, since the mean firing rate at $\theta = 0$ does not change. Combing $\mathbf{f}'(t)$ and $\mathbf{\Sigma}(t)$ in Equations 12 and 13, we can thus calculate I_{optimal} using Equations 14 and 15.

Similarly, for I_{ilPPC}, we have

$$I_{\text{ilPPC}} = \mathbf{f}_{\text{sum}}'^{\mathrm{T}}\,\mathbf{\Sigma}_{\text{sum}}^{-1}\mathbf{f}_{\text{sum}}', \tag{16}$$

except that here $\mathbf{f}_{\text{sum}}$ and $\mathbf{\Sigma}_{\text{sum}}$ are tuning curves and covariance matrix of another N hypothetical LIP cells that implement the ilPPC solution via linear summation of spikes from their corresponding N MSTd neurons. Therefore, by definition,

$$\mathbf{f}_{\text{sum}}' = \frac{\partial}{\partial\theta}\int_0^T \mathbf{f}(t)dt = \int_0^T \mathbf{f}'(t)dt. \tag{17}$$

To calculate $\mathbf{\Sigma}_{\text{sum}}$, using the property of covariance

$$\Sigma_{\text{sum},ij} = \text{Cov}\left(\int_0^T r_i(t)dt, \int_0^T r_j(t)dt\right) = \int_0^T\int_0^T dt_1\, dt_2 \text{Cov}(r_i(t_1), r_j(t_2)) \tag{18}$$

and noting that the firing rate of MSTd neurons $r(t)$ is assumed to be independent across time (see Discussion in the main text), namely,

$$\mathrm{Cov}(r_i(t_1), r_j(t_2)) = \delta(t_1 - t_2)\mathrm{Cov}(r_i(t_1), r_j(t_1)) = \delta(t_1 - t_2)\Sigma_{ij}(t_1), \tag{19}$$

we have

$$\mathbf{\Sigma}_{\mathrm{sum}} = \int_0^T \int_0^T \delta(t_1 - t_2)\mathbf{\Sigma}(t_1)dt_1 dt_2 = \int_0^T \mathbf{\Sigma}(t_1)dt_1 \int_0^T \delta(t_1 - t_2)dt_2 = \int_0^T \mathbf{\Sigma}(t)dt. \tag{20}$$

Now we can calculate I_{ilPPC} by plugging Equations 17 and 20 into Equation 16. In our simulations, we discretized the time into small steps of $\Delta t = 50$ ms, and the choice of Δt did not influence our results much.

Figure S4B shows I_{optimal} and I_{ilPPC} as a function of the number of neurons N for two different ϵs. When $\epsilon = 0$, the total optimal information I_{optimal} did not saturate as N went to infinity, and I_{ilPPC}, although smaller than I_{optimal}, also scaled almost linearly with N (black solid and dashed curves). This is because for a heterogenous population, the first part of correlation in Equation 12 alone is not information-limiting (Shamir and Sompolinsky, 2006; Ecker et al., 2011; Moreno-Bote et al., 2014). This regime however is unrealistic as it predicts that the information in MSTd would be considerably larger than the information available in the behavior of the animal. Indeed, single neurons often performs only slightly worse than the animal, suggesting that the information they convey is only slightly less than the information in the behavior. Since information is proportional to the number of neurons for $\epsilon = 0$, it follows that the information in a large neuronal population would vastly exceed the information in the behavior. Of course, downstream neurons may read out MSTd suboptimally, but Pitkow et al. (2015) have shown that choice correlations in MSTd are consistent with near optimal read out of MSTd activity.

Therefore, we assumed that there are significant information-limiting correlations, or equivalently, differential correlations (Moreno-Bote et al., 2014) in MSTd. When differential correlations were present $(\epsilon = 0.0015)$, both I_{optimal} and I_{ilPPC} saturated rapidly with increasing N (red solid and dashed lines in Figure S4B). The upper limit at which they saturated (blue dashed curves in Figure S4B and D) were predicted by

$$I_{\mathrm{optimal},\infty}(\epsilon) = \int_0^T \dot{I}_{\mathrm{optimal},\infty}(\epsilon)dt = \int_0^T \left(\frac{\epsilon}{v(t)}\right)^{-1} dt = \frac{1}{\epsilon}\int_0^T v(t)dt. \tag{21}$$

Here we used the fact that for any total covariance matrix of the form

$$\mathbf{\Sigma} = \mathbf{\Sigma}_0 + \epsilon \mathbf{f}' \mathbf{f}'^{\mathrm{T}} \tag{22}$$

where $\mathbf{\Sigma}_0$ does not limit information, the information when N goes to infinity will be (Moreno-Bote et al., 2014)

$$\dot{I}_\infty = \lim_{\dot{I}_0 \to \infty} \left(\frac{\dot{I}_0}{1 + \epsilon \dot{I}_0}\right) = \frac{1}{\epsilon}, \tag{23}$$

which depends only on ϵ, where $\dot{I}_0$ is the non-saturating information corresponding to $\epsilon = 0$.

To assess how the baseline problem affects the optimality of ilPPC, we calculated the percentage of information preserved by the ilPPC solution,

$$\mathrm{Optimality\ ratio} = 1 - \mathrm{InformationLoss\%} = \frac{I_{\mathrm{ilPPC}}}{I_{\mathrm{optimal}}} \times 100\%. \tag{24}$$

As shown in Figure S4C, for $\epsilon = 0.0015$, the optimality ratio of ilPPC solution gradually increased with N and reached 95% when $N =$ 10000 (red curve). In contrast, for biologically implausible value of $\epsilon = 0$, the ilPPC solution is much more sensitive to the time-shifting baseline of MSTd tuning curves (> 40% information lost at $N = 10000$; black curve).

Obviously, the choice of ϵ is critical to our calculation, and the reason we have chosen $\epsilon = 0.0015$ is illustrated in Figure S4D. We plotted the optimality and the psychophysical threshold σ_{psy} together as a function of ϵ. The threshold σ_{psy} is defined as a small heading deviation around $\theta = 0$ that could be discriminated at 84% correct by an ideal observer (Gu et al., 2011), which is directly comparable with the animal's psychophysical threshold reported in our experiments (around 4°). Thus, we have

$$\sigma_{\mathrm{psy}} = \sqrt{2}\sigma_{\mathrm{LOLE}} = \frac{\sqrt{2}}{\sqrt{I_{\mathrm{ilPPC}}}}, \tag{25}$$

where σ_{LOLE} is the standard deviation of a LOLE, which is equal to the inverse of the linear Fisher information (Gu et al., 2011; Moreno-Bote et al., 2014). As predicted, when ϵ decreased, both σ_{psy} and the optimality decreased, and we found that for a value around $\sigma_{\mathrm{psy}} = 4°$, ϵ should be around 0.0015, and corresponding to an information loss around 5%.

Furthermore, we checked whether our results were sensitive to the distributions of the baselines B_i. As expected, the optimality of ilPPC decreased gradually from 100% to 84% as the averaged baseline increased from 0 Hz to 40 Hz (Figure S4E). However, the gradient of the optimality around the biologically realistic regime (red star) was small enough for us to be confident about the stability of our estimates.

Taken together, we have shown that a moderate deviation from the ilPPC's prerequisite ($B_i = 0, \forall i$) will only result in ~5% information loss for the ilPPC solution.

DATA AND CODE AVAILABILITY

All experimental data and MATLAB code for data analysis have been deposited in the Mendeley Data: https://data.mendeley.com/datasets/b8ybw4shv3/1. MATLAB code for the network model and the information loss calculation is available at the following public repository: https://github.com/hanhou/Multisensory-PPC.