



Published in final edited form as:

Annu Rev Psychol. 2022 January 04; 73: 103–129. doi:10.1146/annurev-psych-021021-103038.

Cognitive, Systems, and Computational Neurosciences of the Self in Motion

Jean-Paul Noel¹, Dora E. Angelaki^{1,2,*}

¹Center for Neural Science, New York University, New York, USA

²Tandon School of Engineering, New York University, New York, USA

Abstract

Navigating by path integration requires continuously estimating one's self-motion and heading. These estimates may be derived from visual velocity and/or from vestibular acceleration signals. Importantly, these senses in isolation are ill-equipped to provide accurate estimates, and thus visuo-vestibular integration is an imperative. After briefly sketching the visual and vestibular pathways involved, the crux of this review focuses on the human and theoretical approaches that have outlined a normative account of cue combination in behavior and neurons, as well as on the systems neuroscience efforts that are searching for its neural implementation. We highlight understanding how cues with time-varying reliabilities, and how prolonged velocity signals are integrated into a position estimate, as important contemporary frontiers. Further, we discuss how the brain builds internal models inferring when cues ought to be integrated vs. segregated – a process of causal inference. Lastly, we suggest that the study of spatial navigation has not yet addressed its initial condition: self-location.

Keywords

Bayesian Inference; Population Probabilistic Coding; Multisensory; Body; Peri-Personal; Space; Navigation

1. Introduction

Successful navigation is central to adaptive behavior as it underlines our ability to trade-off between exploiting our current location in the environment with exploring novel ones. Traditionally, navigation has been divided into two broad classes; landmark-based and path integration. The former relies on fixed environmental anchors for visual homing, re-orientation, and way finding. The latter, instead, involves integration of evolving estimates of heading, angular, and linear velocity derived from visual, vestibular, proprioceptive, and motor-efference signals into a best guess of position. In a sense, landmark-based navigation may be allocentric (e.g., turn left at the fridge), while path integration cannot – it relies on self-motion information derived from an egocentric perspective (e.g., an optical flow field radiating from a focus of expansion). As such, the study of path integration and

*Corresponding Author: Dora E. Angelaki, Professor, Center for Neural Science, NYU, Da93@nyu.edu.

self-motion may allow us to further understand not only navigation, but also our subjective and egocentric sense of self-location.

Here, we attempt to contextualize recent findings of path integration and self-motion while highlighting novel and interesting developments in neighboring and interdependent fields of study. First, we sketch the visual and vestibular neural pathways involved. We start with the vestibular pathways, as these are commonly less known to the general audience. Second, we highlight that the visual and vestibular systems are in isolation incapable of accurate self-motion perception. Thus, much of our focus is in outlining the computational and neural principles that underpin visuo-vestibular Bayes near-optimal integration. Further, we review initial findings and suggestions regarding the mechanism behind Bayesian causal inference. In the last section we highlight an area of study that is seldom incorporated into the study of navigation yet constitutes its initial condition; the sense of self-location.

2. Neural Pathways for Self-Motion

A multitude of sensory systems contribute to our subjective sense of self-motion. The strongest of these are likely vision and the vestibular system, and thus in what follows we briefly outline these neural pathways.

2.1. The Vestibular System in Self-Motion

The vestibular peripheral organ is located in the inner ear and comprises two components; the otolith organs and the semi-circular canals (Fig. 1, bottom-most). The former detect linear head acceleration, both horizontally and vertically (i.e., gravity). The latter sense head rotation in three orthogonal planes. In turn, the afferent fibers of the vestibular nerve project to central vestibular areas, in particular the vestibular nuclei. This area is composed of many cell types, some are involved in gaze-stabilization, while others (e.g., “vestibular-only”, VO) are thought to be involved in posture, self-motion, and likely navigation (see Cullen, 2019 for a recent review). VO neurons respond to passive (i.e., externally applied) head motion, but responses are suppressed during active head motion, translation (Carriot et al., 2013) or rotation (Roy & Cullen, 2001, 2004).

The suppression of VO neurons during active self-motion is predicted by a Kalman filter-based model of self-motion (Laurens & Angelaki, 2017). More specifically, the cerebellum is generally thought to form a forward internal model that predicts the sensory consequences of self-generated movement (Krakauer & Mazzoni, 2011). Hence, theoretical models (Laurens & Angelaki, 2017) of the vestibular system have similarly suggested that during active movement the cerebellum may compute an internal model of the expected sensory consequences of a motor command. This estimate is then compared with the observed sensory inflow to generate sensory prediction errors. When expectations are violated, as is the case during passive head movements, vestibular reafference cancellation signals from the cerebellum to the vestibular nuclei are not suppressed, and thus activity in VO neurons is enhanced (Fig. 1). In addition to VO neurons in the vestibular nuclei, recordings from the rostral fastigial nucleus of the primate cerebellum confirm the computation of sensory predictions that enable the distinction between self-generated and externally applied self-motion (Brooks et al., 2015). Remarkably, therefore, already at this early stage, signals

mediating self-motion estimation are multisensory (i.e., vestibular, motor efference copy, and likely proprioception from the neck).

From the vestibular nuclei information is sent to the cortex via two ascending thalamocortical pathways. The anterior vestibulo-thalamic pathway projects first to the prepositus and supragenual nucleus, then to the dorsal tegmental nucleus, and finally the lateral mammillary nucleus – all within the brainstem (Fig. 1). The association between the latter two areas is postulated to encode a ring attractor (reviewed in Knierim & Zhang, 2012) that leads to head-direction (HD) cells in their downstream area, the anterior dorsal thalamus (ADN; see Hulse & Jayaraman, 2020, for a recent review). These HD cells are egocentric in nature, in that they encode the direction of heading. The ADN outputs to the retrosplenial cortex and the dorsal presubiculum before this anterior pathway converges onto the well-known spatial codes of the entorhinal cortex and hippocampus (Moser et al., 2008). The entorhinal cortex is heterogeneously composed of and multiplexes (Hardcastle et al., 2017) head-direction, place (O'Keefe & Nadel, 1978), speed (Kropff et al., 2015), border (Solstad et al., 2008), and grid cells. It is likely best known for this latter cell-type, tiling space in a hexagonal pattern (Hafting et al., 2005). The hippocampus possesses place cells, neurons that fire when the animal is within a particular location of space (O'Keefe, 1976). Thus, interestingly, while supported by the vestibular system – an idiothetic sense - the anterior vestibulo-thalamic pathway ultimately is involved in building an allocentric map in limbic areas (Fig. 1).

The second ascending thalamocortical pathway is the posterior one. This pathway projects from the vestibular nuclei and the cerebellum to the ventral posterior lateral thalamus (VPL). The VPL is also a hub for somatosensory information (Jones, 1985), and thus it is not surprising that this area is highly multisensory, encoding for vestibular, somatosensory, proprioceptive, visual, and motor signals. From here, the posterior vestibulo-thalamic pathway projects directly to the parieto-insular vestibular cortex (PIVC) and the ventral intraparietal area (area VIP), among many others (see Lopez & Blanke, 2011 for an extensive review). This vast proliferation of vestibular signals from posterior thalamus to numerous cortical areas, and the fact that these target areas are multisensory in nature, is why it is said that there is no primary vestibular cortex.

In addition to the above-mentioned areas receiving vestibular input directly from VPL, the medial superior temporal area (MST), particularly the dorsal subdivision (MSTd; Duffy, 1998), but also the lateral one (Sasaki et al., 2019), and area 7a (Avila et al., 2019) also respond to vestibular stimulation. Thus, seemingly much of the dorsal stream (e.g., MSTd, VIP, 7a) is generally responsive to vestibular self-motion stimuli. Further, until recently it was thought that the response patterns in these areas showed a progressively stronger correlation with heading discrimination behavior (e.g., MSTd, Gu et al., 2008 vs. VIP, Chen et al., 2013), at least insofar as measured by choice probabilities (Britten et al., 1996). However, recent causal experiments performing chemical inactivation have questioned the causal role of MSTd and VIP in vestibular heading perception, as there is no or little impairment in heading discrimination when these areas are “shutdown” (Gu et al., 2012; Yu & Gu, 2018).

Interestingly, while the posterior parietal cortex is widely considered to be a hub for egocentric spatial navigation, and its tuning to allocentric variables (e.g., route information) is weak (e.g., Chen et al., 1994), area 7a – being downstream from most of posterior parietal cortex – shows properties that may suggest a putative gradual transformation toward cues amenable for allocentric encoding. That is, 7a seems to show weak visuo-vestibular convergence and distinct subpopulations of neurons either code for linear or angular velocity (Avila et al., 2019). Given that the distinctive characteristic of 7a relative to its parietal neighbors is its anatomical connection to the retrosplenial cortex and indirectly to hippocampal formation (Pandya & Seltzer, 1982; Kobayashi & Amaral, 2000), we may speculate that the neural codes in 7a and retrosplenial cortex (showing e.g. progress divergence as opposed to convergence of linear and angular velocity signals) may be best suited for readout in the hippocampus (see Kravitz et al., 2011, for a similar argument linking the caudal inferior posterior lobule with hippocampal formation spatial codes). Similarly supporting this conjecture, Avila and colleagues (2019) recently reported that 7a is most readily driven by vestibular and not visual optic flow information, and this former sensory modality is tuned to acceleration. In this line, Kropff et al., 2021, have recently demonstrated that contrarily to popular belief, theta rhythms organizing neural activity across hippocampus and entorhinal cortex are modulated by acceleration, and not speed, of running rats.

Together, we may speculate that while two ascending thalamocortical vestibular pathways exist (anterior and posterior), these in fact form a loop, being separate and egocentric at their outset (in thalamus and cortical areas), and converging in the hippocampal formation where they employ an allocentric code (Fig. 1, see Herweg & Kahana, 2018 for a similar argument, and see Andersen et al., 1985; Bicanski & Burgess, 2016, for arguments regarding the involvement of the retrosplenial cortex in allo-/ego-centric transformation).

2.2. The Visual System in Self-Motion

When a stationary observer views clouds move past her, a river flow by her, or a train departing on the adjacent track, she may experience an illusory sense of self-motion. This phenomenon is calledvection (Tschermak, 1931) and as the above examples illustrate, it is a sensation that may occur in nature. As such, it has long been appreciated that visual cues alone can generate self-motion perception (Mach, 1875). In particular, it is well established that large field, coherent, and global motion mimicking the pattern of flow that occurs on our retinæ as we move relative to the environment is capable of eliciting vivid sensations of self-motion (see Dichgans & Brandt, 1978). This pattern of motion was denominated “optic flow” (Gibson, 1950), and has served as the backbone for much of the modern-day study of self-motion.

What are the neural pathways involved in the processing of optic flow? The striate and extrastriate cortices are well studied, in particular for their motion responses (Maunsell & van Essen, 1983) and hence natural contenders for the processing of optic flow emerged rapidly. A subset of cells in primary visual cortex (V1) are highly selective for direction, but these cells have small spatiotemporal receptive fields and encode motion of local features (Hubel & Wiesel, 1968) – thus likely not ideally suited for self-motion

processing (but see Véléz-Fort et al., 2018). The middle temporal area (MT) likely integrates motion cues inherited from V1 (Adelson & Movshon, 1982) and cells in this area can encode two-dimensional motion, such as patterned motion (i.e., vertical bars moving north and south east yielding a rightward percept). MT is also thought to estimate velocity (Adelson & Movshon, 1982). However, this area does not seem tailored for complex and whole-field flow processing. Instead, the subsequent stages of the visual dorsal stream – dorsal subdivision of the medial superior temporal area (MSTd), ventral intraparietal area (VIP), and area 7a (Fig. 1) – all seemed to show properties well suited for optic flow processing: (1) large and often bilateral receptive fields, (2) selectivity for complex visual motion patterns, and (3) often partial remapping of reference frames allowing for heading representation independent of eye-position (Tanaka et al., 1986; Duffy & Wurtz, 1991; Siegel & Read, 1997; Bremmer et al., 2002; Avillac et al., 2005). These latter areas have therefore been those most extensively studied in the processing of optic flow and self-motion (see Britten, 2008, for an earlier review focusing on MSTd and VIP).

Early studies suggested a weak but consistent correlation between spiking activity in MSTd and trial-to-trial fluctuations in heading perception derived from optic flow (Britten & van Wezel, 1998; Gu et al., 2008). This small correlation has recently also been shown in MT (Yu et al., 2018; Yu & Gu, 2018). In contrast, the downstream area VIP shows substantially larger correlations between brain activity and heading perception (Chen et al., 2013). To the best of our knowledge the choice probability for heading judgments has not been reported in 7a. While it would be tempting to suggest that higher levels of the visual hierarchy (e.g., from MT to VIP) have a stronger role in guiding heading perception, recent experiments do not support this simple view. Causal experiments bilaterally suppressing MSTd showed a three-fold increase in the psychophysical threshold for visual heading perception (Gu et al., 2012). Remarkably, however, bilateral suppression of VIP had no effect on heading perception as derived from optic flow (Chen et al., 2016). Supporting the conclusion linking MSTd but not VIP to heading perception, Zaidel and colleagues (2017) dissociated bottom-up sensory and top-down choice driven components to choice probabilities. This analysis suggested a preponderance of heading signals in MSTd and of choice signals in VIP.

Recordings in the vestibular and deep cerebellar nuclei (Bryan & Angelaki, 2008), as well as PIVC (Chen et al., 2010) showed a lack of responsiveness to optic flow.

In general, therefore, there is a convergence of visual and vestibular signals for self-motion in the parietal dorsal stream (e.g., MSTd, VIP, 7a). However, their exact functional roles are not yet fully understood. Speculatively, it seems as if the strongest signals relating to the *encoding* of self-motion from vision are in MSTd, but after this stage information may be most strongly related to decision-making processes and visuo-vestibular cue combination, while being highly distributed and redundantly coded (see Zhang et al., 2016a for a similar argument and modeling effort, and see Bizley et al., 2016, for an argument that distributed networks underlie multisensory decision-making).

3. Visuo-Vestibular Integration; computation, algorithm and implementation

Despite their clear contribution to self-motion processing, in isolation the visual and vestibular systems are ill equipped to guide spatial navigation. As mentioned above, given that self-motion is relative, during full field visual motion observers may misinterpret global world motion for self-motion (i.e.,vection; see Dichgans & Brandt, 1978 for an early and extensive review). Similarly, optic flow may be caused by true translation or rotation of the head in the environment, but may also be caused by rotation of the eyes in orbit, causing a confluence of signals that have to be parsed (i.e., the “rotation problem”). The vestibular system also has inherent limitations. For instance, given that the inner ear detects acceleration, in the absence of visual cues we cannot sense movement after a prolonged period of constant velocity (e.g., closing one’s eyes on a moving train). Likewise, given that otolith afferents encode linear acceleration and changes in head orientation relative to gravity in an identical manner, this system in isolation cannot distinguish between these (i.e., Einstein’s equivalence principle, Einstein, 1907). Thus, the integration of visual and vestibular signals does not only suppose a redundancy of encoding that via multisensory integration is likely to ameliorate perceptual sensitivity (see Fetsch et al., 2013 for a review), but also overcomes fundamental deficits in each of these systems.

In this section we first summarize behavioral and computational evidence specifying how signals ought to be combined, from a principled perspective. Then, we highlight probabilistic population codes as a theoretical framework detailing how optimal cue combination may occur in the brain and review the evidence for this sort of neural code in visuo-vestibular integration for heading perception (see Box 1 for a broader discussion on the neural instantiation of statistical inference). We attempt to highlight important advances that have 1) developed ideal observers who integrate signals over an undetermined period of time and with time-varying reliabilities, and 2) have sketched the putative neural implementation of this computation. Lastly, we discuss causal inference as a more general computation toggling between different internal models (e.g., dictating the integration or not of visuo-vestibular cues), and point to theoretical proposals, as well as recent findings from cognitive and systems neuroscience that together promise to ultimately elucidate the neural underpinning of this fundamental and ubiquitous computation.

3.1. Bayesian Observers

Our brains are locked inside dark and silent skulls. They understand the language of spikes and not that of visual objects and vestibular events. Thus, as Helmholtz (1867) most famously stated, perception is a process of (unconscious) inference. We do not have direct access to the external world, and instead we must make our best guesses based on available sensory evidence and prior knowledge.

More formally, and taking the example of heading discrimination, on a particular trial, t , an observer is presented with a specific heading, θ_t . This stimulus is encoded by noisy and stochastic biological elements, and thus, our measurements or observations, m , of the environment may change on a trial-by-trial basis, even for a fixed θ_t (Tolhurst et al. 1982).

The resultant distribution, $p(m | \theta_t)$ is called a measurement distribution and is defined for a fixed stimulus. It is typically considered to be Gaussian and centered on the true θ_t . Together, the relation between i) the different headings that we may experience, $p(\theta)$ and ii) the measurement distribution, specifies a generative model. This model is an explanation of how sensory data was generated by the world and our sensory systems, and is the schema the brain is tasked with ‘inverting’ to perceive. That is, we make a hypothesis to explain the observed data. The process of translating external stimuli to internal measurements, $\theta_t \rightarrow m$, is referred to as neural encoding and has a rich computational history (i.e., efficient coding; Barlow, 1961), yet is unfortunately typically considered separately from decoding processes, such as Bayesian inference (but see Wei & Stocker, 2015 for an exception).

In a first step of inference, observers generate a degree of belief (Ramsey, 1926) about θ based on their measurements. This belief is characterized by likelihood functions, $L(\theta_t)$ that effectively take the same shape as $p(m | \theta_t)$, but in this case are functions of θ_t and not m (see Ma, 2019, for details). Next, an observer may incorporate the belief about different headings being generally present in the environment, $p(\theta)$, the prior distribution. In different contexts this distribution is typically considered to be Gaussian, uniform, or broad enough to be negligible. According to Bayes’ Rule (Eq. 1), by combining the likelihood and prior we can compute posterior distributions; the probability of θ given m .

$$p(\theta | m) \propto L(\theta_t)p(\theta) \quad (\text{Eq. 1})$$

In a last step, the observer must make a decision or action. This requires a cost function (i.e., penalties and rewards for hits, misses, etc.) and a mapping from posterior distributions to a concrete action. In the general case where priors and likelihood distributions are Gaussian, the posterior will be so as well, and thus the mean, median, and mode of the posterior specify the same value, the same action. However, this is not always true (see Section 3.5) and thus loss functions and action-selection must be carefully considered (Rahnev & Denison, 2018). What characterizes an observer as optimal is the use of the correct generative model and computations that minimize cost or maximize reward (see Daptardar et al., 2019 for a description of rational observers as those making optimal decisions within an incorrect generative model).

3.2. Bayes Optimal Cue Fusion

Borrowing from insights in computer vision (Knill & Richards, 1996) and within the Bayesian framework detailed above, Ernst & Banks (2002; among others), specified an ideal observer for multisensory cue combination. They assumed a flat prior, Gaussian likelihoods, and that measurements are conditionally independent across modalities (i.e., signals may co-vary, but their noise does not). Under these assumptions, it can be shown that the likelihood function of a combined (e.g., visuo-vestibular) condition, $L^{\text{comb}}(\theta)$ is the product of the unisensory likelihoods, $L^{\text{vis}}(\theta) L^{\text{vest}}(\theta)$, and the maximum-likelihood estimate (MLE) will be:

$$\hat{\theta}_{\text{comb}} = w_{\text{vis}} \hat{\theta}_{\text{vis}} + w_{\text{vest}} \hat{\theta}_{\text{vest}} \quad (\text{Eq. 2})$$

with $\hat{\theta}_{vis}$ and $\hat{\theta}_{vest}$ being the unisensory estimates, and w_{vis} and w_{vest} being weights that are proportional to inverse variances (σ_{vis}^2) and (σ_{vest}^2):

$$W_{vis} = \frac{\frac{1}{\sigma_{vis}^2}}{\frac{1}{\sigma_{vis}^2} + \frac{1}{\sigma_{vest}^2}} \quad (\text{Eq. 3})$$

and analogously for w_{vest} . The variance of the combined estimate is:

$$\sigma_{comb}^2 = \frac{\sigma_{vis}^2 \sigma_{vest}^2}{\sigma_{vis}^2 + \sigma_{vest}^2} \quad (\text{Eq. 4})$$

Thus, if individuals are combining information across cues, their combined estimate will intuitively fall in between the unisensory estimates, weighted by the relative reliability of each cue. More importantly, given that a weighted average estimate across trials could also emanate from following a given estimate at times and the other estimate in the remaining of trials, the true hallmark of optimal cue combination is a reduction in uncertainty (predicted by Eq. 4). Humans have been shown to combine cues optimally or near optimally within senses (Hillis et al, 2004) and across visuo-tactile (Ernst & Banks, 2002), audio-visual (Alais & Burr, 2004), visuo-proprioceptive (van Beers et al., 1996), and visuo-vestibular (Fetsch et al., 2009; Prsa et al., 2012) pairings, among others.

3.3. Neural Instantiation of Bayes Optimal Cue Fusion

Armed with a principled account of multisensory behavior, the next step was to derive how such a computation could be instantiated algorithmically – what operations and set of rules may neurons follow in accomplishing the computation at hand?

In an influential theoretical contribution Ma and colleagues (2006) highlighted that neural populations had to represent the reliability associated with environmental cues in order to perform an inference of the type $p(S|r)$, where S is the cue and r is a vector of neural responses for a given presentation. In analogy to the Bayesian observers described above, $p(r|S)$ is proportional to $p(S|r)$, and the former is something we can measure. In fact, we know that cortical neurons tend to show Poisson-like variability (Tolhurst et al. 1982), meaning that their average activity is monotonically related to their variance. Taking this property into account it can be shown that the posterior distribution $p(S|r)$ approximates a Gaussian function, its mean closely corresponds to the peak of population activity, and importantly, its variance is implicitly encoded in the amplitude of the population response, or gain, g , such that $Kg = 1/\sigma^2$, where K is a constant.

In turn, regarding cue combination and again with the example of visuo-vestibular integration, this probabilistic population code (PPC; Ma et al., 2006) makes the hypothesis that if unisensory populations have the same number of neurons, identical tuning functions, and independent Poisson-like variability, optimal conservation of information (e.g., $I_{comb} = I_{vis} + I_{vest}$ Clark & Yuille, 1990) equates to a simple sum of neural activities, $r_{comb} =$

$r_{vis} + r_{vest}$. Since the unisensory areas are characterized by Poisson-like variability, so will the multisensory, and $\frac{1}{\sigma_{comb}^2} = \frac{1}{\sigma_{vis}^2} + \frac{1}{\sigma_{vest}^2}$ which is equivalent to the uncertainty reduction outlined in Eq. 3. Many of the assumptions outlined (e.g., equal number of neurons) can be relaxed in more general formulations (e.g., $w_{comb} r_{comb} = w_{vist} r_{vis} + w_{vest} r_{vest}$), but the important take home is that by incorporating the known distribution of single-unit variability, PPCs are able to accomplish a multiplication required at the computational level, $L^{vis}(\theta) L^{vest}(\theta)$, by simple summation – convergence of unisensory populations onto a multisensory one (see Ma et al., 2006 for mathematical details).

Angelaki, DeAngelis, and colleagues performed a series of experiments to detail the neural code underlying visuo-vestibular integration, and to specifically ascertain whether PPCs were indeed biologically implemented. First, Gu et al., 2008, demonstrated that non-human primates perform a discrimination task where they are required to indicate their direction of heading relative to straight-head in line with optimal cue combination, their sensitivity during visuo-vestibular conditions improving consistently with theoretical predictions (Ernst & Banks, 2002). Second, Fetsch et al., (2009, 2012) showed that these animals also took into account the relative uncertainty between cues in generating estimates when visual and vestibular cues were incongruent. In addition to the behavioral observations, these authors performed single-neuron recordings in MSTd, and observed two classes of neurons; those with congruent visual and vestibular tuning functions, and those with opposite preferences. Neurometric curves constructed from receiver operating characteristic (ROC) analysis of spiking activity of congruent cells had visuo-vestibular discrimination thresholds in line with predictions from optimal cue combination (Gu et al., 2008, see Chen et al., 2013 for a similar result in VIP). We return to the “opposite” cells below.

These behavioral and physiological studies set the stage for questioning whether optimal cue combination in fact occurs in the brain as predicted by PPCs. However, the headings probed in the early reports were fairly restricted and thus did not allow for sketching a neural combination rule – the set of weights A , such that $R_{comb} = A_{vis} R_{vis} + A_{vest} R_{vest} + C$, where R are neural responses and C is a constant. To remedy this situation, Morgan et al., 2008 recorded from MSTd while presenting non-human primates with the full gamut of visual, vestibular, and visuo-vestibular headings – including incongruent presentations. Results demonstrated that a linear combination of unisensory visual and vestibular responses was indeed well able to predict multisensory responses (i.e., the addition of non-linear components did not significantly improve fits). However, somewhat surprisingly weights A_{vis} and A_{vest} were smaller than ‘1’ (i.e., sub-additive as opposed to the additive), and perhaps most vexingly, varied with changes in cue reliability. This latter observation was at face value incongruent with PPC, as this theory in essence suggests that the weighting of likelihood functions by their reliability is accomplished at the unisensory level, where, for example, visual responses are modulated by coherence, and thus Poisson statistics imply that there is no need in updating neural weights, A , with changes in stimulus coherence.

Two subsequent reports proposed why, and putatively how, neural weights may change as a function of visual coherence, and hence re-instated PPCs as a putative neural mechanism of optimal cue integration. Fetsch and colleagues (2012) first showed that Poisson statistics

do not entirely account for how neural responses in MSTd change with visual motion coherence. Instead, with increasing coherence in visual stimuli there is both a multiplicative scaling of neural responses and a change in baseline firing. Taking these properties into account, the researchers derived the optimal neural weights for visuo-vestibular integration in MSTd and showed a correlation between mathematically-derived optimal and measured neural weights (see Hou et al., 2019, for a suggestion that incorporating neural correlations could have strengthened the agreement between PPC theory and empirical observations). Secondly, Ohshiro and colleagues (2011) suggested that both the sub-additivity in neural weights (i.e., $A < 1$) and rapid changes of these weights on a trial-by-trial basis could be accounted by divisive normalization acting at the stage of multisensory integration. Divisive normalization is a ubiquitous neural computation wherein the output of each neuron is divided by the summed activity of a “normalization pool” (Carandini & Heeger, 1994, 2011). Thus, the strength of the normalization pool depends on unisensory firing rates, and hence, as firing rates co-vary with stimuli coherence, so will the neural weights. In fact, divisive normalization can give rise to a neural combination rule similar to that measured in Morgan et al., 2008 and Fetsch et al., 2012. Further, this property is likely critical for appropriate function of the nervous system as a whole in that it prevents neural saturation – a “ceiling effect” in firing rates – and hence potentially why neural weights in fact need to be sub-additive. Lastly, in a beautiful convergence of evidence, divisive normalization at a multisensory layer is not only able to account for population level properties in cortex (Ohshiro et al., 2011), but is equally able to account for properties of individual multisensory neurons in subcortex, such as their supra-additive responses during weak stimuli presentations (“inverse effectiveness”), or presentations within co-localized receptive fields (“spatial principle”; see Stein & Stanford, 2008, for a review summarizing early work detailing the properties of multisensory neurons in superior colliculus). The divisive normalization conjecture makes a strong and testable prediction: non-preferred sensory input from one modality should suppress the response to a preferred input in another modality. Recent recordings have confirmed the presence of this form of cross-modal suppression in MSTd and not in MT (Ohshiro et al., 2017).

Broadly, therefore, a multitude of visuo-vestibular phenomena (e.g., vection) and the basic peripheral properties of the visual and vestibular system implied that appropriate self-motion perception requires the integration of visual and vestibular information. Normative approaches to modeling behavior then suggested how these signals ought to be integrated, and landmark theoretical studies bridged the gap between behavioral cue combination and neural integration. Physiological recordings in MSTd then largely confirmed predictions from theory, while iteratively adding caveats – e.g., MSTd responses are further from Poisson-like than initially suggested, neural weights may vary trial-to-trial with changes in stimuli reliability. These empirical observations lead to the conjecture of a network-level operation that may account for the inconsistencies between theory and empirical observations. And in turn, this circuit property (i.e., divisive normalization) was not only able to account for population level responses in cortex, but also to incorporate traditional properties of single neurons in subcortex (see Fetsch et al., 2013 and Hou & Gu, 2020, for insightful reviews).

3.4. Time and Time-Varying Reliabilities

An important and more recent extension to the study of optimal cue combination is the incorporation of time as a critical variable. Indeed, the human literature on multisensory integration appears to be divided between those employing estimation tasks and optimal cue combination as a theoretical framework on one side (e.g., Gu et al., 2008; Fetsch et al., 2009), and those using reaction time tasks and race-models (Raab, 1962, Miller, 1982) or principles derived from early single-unit electrophysiology (Stein & Stanford, 2008) on the other hand. The former have so far ignored a critical dimension present in all perceptual and decisional processes, time, while the latter ignore the perceptual sensitivity benefits derived from multisensory integration and have not been able to connect behavior with neurons, nor can they establish whether multisensory inputs are combined optimally.

Drugowitsch et al. (2014) closed the gap between the study of multisensory precision and speed by deriving an extension to the traditional drift diffusion model (DDM). The conventional DDM (Ratcliff & Rouder, 1998) is based on particle dynamics accumulating evidence until hitting a decision bound. These models can account well for stereotypical distributions of reaction times, and changes in the speed of evidence accumulation (i.e., drift-rate) and/or the initial distance of particles to the decisional boundary can accommodate speed-accuracy trade-offs during decision making. Additionally, these diffusion models are known to optimally integrate evidence over time given that the reliability of the evidence is time-invariant (Bogacz et al., 2006). However, in their standard implementation DDMs are not optimal when the speed of evidence accumulation changes over time (within or across trials), nor are they designed to integrate disparate sources of information. In Drugowitsch and colleagues' (2014) extension, a multisensory DDM's drift-rate is determined by a weighted combination of unisensory drift-rates, each weighted in proportion to their relative and momentary (i.e., time-evolving) sensitivities.

This version of DDMs is optimal despite time-varying reliability of cues (see Drugowitsch et al., 2014 for mathematical detail). In applying this model, and within the context of a speeded version of the visuo-vestibular heading discrimination task, the multisensory DDM can account for apparent sub-optimal behavior as indexed by standard analyses not incorporating time as a factor. It also suggests a near optimal speed-accuracy trade-off in maximizing reward rate across trials (Drugowitsch et al., 2015). Further, in analogy, this framework specifying both accumulation of evidence across time and across the senses may be able to account not only for apparent sub-optimal behavior (Drugowitsch et al., 2014), but also for recent reports of "supra-optimal" behavior, most common in the rodent literature (Raposo et al., 2012; Nikbakht et al., 2018, but see Shalom & Zaidel, 2018, for an alternative explanation). Finally, and perhaps most interestingly from a neural implementation standpoint, the extended DDM suggests that in natural self-motion visual and vestibular signals may each play an outweighed role during different time periods. As alluded to above, vestibular signals are most sensitive to acceleration while visual signals are tuned to velocity, and thus their weight during visuo-vestibular motion may vary accordingly. This conjecture would also suggest that there is no need to integrate vestibular acceleration into a velocity signal, a process that could in principle be costly in terms of signal-to-noise (Bogacz et al., 2006; Churchland et al., 2011; but see Laurens et al., 2017 for

evidence that vestibular acceleration seems to indeed be transformed into velocity estimates as it climbs the neuraxis).

In the most recent physiological recordings attempting to further delineate the neural underpinning of optimal visuo-vestibular integration, Hou and colleagues (2019) took on the challenge of determining if in fact cue combination is dependent on momentary evidence (e.g., visual velocity and vestibular acceleration), and whether such a code is compatible with PPC (Ma et al., 2006). These authors presented non-human primates with translations of a Gaussian velocity profile, naturally dissociating moments of maximal vestibular information (early and late in the motion profile, due to its encoding of acceleration) vs. visual information (peaking with maximal velocity). Single unit recordings were performed in the lateral intraparietal (LIP) cortex. This area receives anatomical input from MSTd and VIP (Boussaoud et al., 1990), two areas heavily implicated in the coding of self-motion (Gu et al., 2006, 2008; Chen et al., 2011, 2013). However, while a majority of neurons in LIP are in fact tuned to visual motion direction, this selectivity is very broad ($>120^\circ$; Fanini & Assad, 2009). Thus, in keeping with the general thought of LIP as an area reflecting evidence accumulation (but see Huk et al., 2017; Katz et al., 2016; Zhou & Freedman, 2019 for recent controversy), recording in LIP (as opposed to earlier areas) likely allowed Hou et al., 2019, to examine a neural node that is a good candidate for one performing a computation akin to integration in the multisensory DDM (Drugowitsch et al., 2014). Further, recording from LIP implicitly supports the hypothesis that multisensory integration occurs at a decisional stage (see Bizley et al., 2016). Hou and colleagues, 2019, demonstrated that LIP indeed harbors heading discrimination choice signals that peak in accordance with vestibular acceleration and visual speed. Moreover, the authors demonstrated that a network performing decisions by summing spikes across time and cues via an invariant linear PPC (Beck et al., 2008) was able to perform optimal multisensory decisions. Finally, a linear approximation of the optimal model showed responses similar to LIP, while decreasing its time-constant of integration did not. In other words, this report suggests that i) PPCs is the algorithm supporting optimal cue combination even for time-varying reliabilities, ii) this algorithm is housed (at least partially) in LIP, and iii) a defining characteristic of LIP vs. its neighbors is its time constant of integration.

Novel path integration studies employing optic flow alone (Lakshminarasimhan et al., 2018, 2020; Noel et al., 2020, 2021) or visuo-vestibular signals (Stavropoulos et al., 2020) during protracted timelines ($\sim 2\text{--}4$ seconds) will be ideally suited to further examine the circuit motifs sustaining long vs. short integration time constants. Initial results within this domain suggest there is no “leak” in the integration of self-motion information into a position estimate, and instead errors in path integration may be due to initial mis-estimations of velocity (Lakshminarasimhan et al., 2018; Noel et al., 2020).

3.5 Causal Inference

In addition to the reliability of different sensory signals, the world around us and the objects and events in our surrounding change dynamically over time. That is, the approach of optimal cue combination (Ernst & Banks, 2002) outlined above is sometimes referred to as a “forced fusion” model, given that its main limitation is that it can only consider one

alternative; the signals must be combined. However, in the real world there are instances when multiple signals refer to the same source (e.g., auditory and visual signals conveying speech and mouthing of an interlocutor) and thus should be combined, and instances when these signals relate to different sources (e.g., an unskillful ventriloquist) and should be separated. To appropriately perceive and act in the world, therefore, we must first use the samples we draw from our environment, observations, to build an internal model specifying the likely causal structure of the environment (“building” the generative model, Fig. 2). Then, we can use this deduced generative model in perceiving. This process is referred to as Bayesian Causal Inference (Fig. 2, Kording et al., 2007), and is again based on Bayes Rule;

$$p(C | x_{vis}, x_{vest}) = \frac{p(x_{vis}, x_{vest} | C)p(c)}{p(x_{vis}, x_{vest})} \quad (\text{Eq.5})$$

where x_{vis} and x_{vest} refer to sensory measurements, and C is categorical variable whose value depends on the state of the world. In an example where visual and vestibular signals either index the same ($C=1$) or separate causes ($C=2$),

$$\begin{aligned} p(C = 1 | x_{vis}, x_{vest}) \\ = \frac{p(x_{vis}, x_{vest} | C = 1)p(c = 1)}{p(x_{vis}, x_{vest} | C = 1)p(c = 1) + p(x_{vis}, x_{vest} | C = 2)(1 - p(c = 1))} \end{aligned} \quad (\text{Eq.6})$$

Solving $p(x_{vis}, x_{vest} | C = 1)$ and $p(x_{vis}, x_{vest} | C = 2)$ allows establishing the likelihood of signals emanating from a single cause, $p(C = 1 | x_{vis}, x_{vest})$, and these have closed form analytical solutions assuming measurement distributions and priors are Gaussian or uninformative (see Kording et al., 2007 for mathematical detail). In turn, the maximum-a-posteriori estimate of the different signals $\hat{\theta}_{vis}$ and $\hat{\theta}_{vest}$ can be computed under the different hypotheses, $C = 1$ or $C = 2$ (“inverting” the generative model, Fig. 2). Now, exactly how these estimates and the inferred causal structure are used in generating actions depends on the loss function, and this one is largely dependent on the specific task. The three decision strategies that are routinely considered are; model averaging, probability matching, and model selection (e.g., Wozny et al., 2010; Cao et al, 2019). The former linearly combines estimates derived from integration and segregation, each weighted by the inferred posterior probability over the respective causal structure. On the other hand, probability matching and model selection commit to a certain world-view for a given trial. The final estimate is sampled from, say, $\hat{\theta}_{vis, c = 1}$ or $\hat{\theta}_{vis, c = 2}$, with a proportion that is either stochastic (probability matching) or fixed (model selection; Fig. 2)

Causal inference has been shown to account well for a number of empirical observations, from low-level audio-visual localization (Odegaard et al., 2015) to speech perception (Magnotti et al., 2017; Noel et al., 2018), and heading discrimination (Acerbi et al., 2018; Dokka et al., 2019), among many others (see French & DeAngelis, 2020, for a recent review, and below for further examples). However, the precise neural underpinning of this computation is less well established.

To tackle this gap in our knowledge, recent human neuroimaging studies based on functional Magnetic Resonance Imaging (Rohe & Noppeney, 2015, 2016) or time-resolved

M/EEG (Rohe et al., 2019; Aller & Noppeney, 2019; Cao et al., 2019) are starting to delineate the general principles putatively guiding the neural implementation of causal inference. By and large these reports all agree in describing causal inference as a hierarchical process, where early sensory areas (e.g., V1 or A1; Rohe & Noppeney, 2015) and early neural latencies (e.g., <100ms; Aller & Noppeney, 2019) encode their preferred sensory modality independently. Intermediate areas (e.g., posterior parietal cortex) and latencies (<250ms) show patterns most consistent with “forced-fusion” (Ernst & Banks, 2002), and finally more anterior regions (anterior parietal cortex; Rohe & Noppeney, 2015) and latter neural latencies (>250ms) flexibly vary their response patterns in accordance with causal inference. An existing discrepancy is whether groups emphasize anterior aspects of the parietal cortex (Rohe & Noppeney, 2015) or the inferior frontal lobe (Cao et al., 2019) as the primary seat of causal inference.

Broadly, these neuroimaging studies are conceptually consistent with initial efforts to implement causal inference in biologically plausible neural networks. Early work in this area suggested that a decentralized and interconnected network (e.g., MSTd and VIP) receiving input from pools encoding unisensory stimuli (e.g., MT and PIVC, respectively) could perform optimal cue combination (Zhang et al., 2016). This type of redundant encoding, not only of stimuli but also across neural areas, is robust to local failure (vs. e.g., postulating LIP as the sole region of multisensory integration), and seemingly consistent with the observation that while disrupting one area may alter unisensory encoding, optimal cue combination is preserved (see Hou & Gu, 2020 for a review). More recently, this architecture has been updated to include known properties of areas computing visuo-vestibular heading. That is, in addition to cells with overlapping visual and vestibular tuning functions (“congruent cells”), there is a large fraction of neurons showing opposite tuning functions (offset by ~180°; opposite cells; Gu et al., 2008). Opposite cells have been postulated to be involved in dissociating object motion during self-motion (e.g., Sasaki et al., 2017, 2019, see also Sasaki et al., 2020) and recent neural network models suggest they may more generally compute Bayes factors (Zhang et al., 2019a; ratio between a segregation and integration model of the world) and retaining access to unisensory likelihoods even after fusion (Zhang et al., 2019b, also see Hillis et al., 2002).

Ultimately, the precise detailing of the neural implementation of Bayesian causal inference will depend on invasive neurophysiology, and thus on the development of behavioral paradigms capable of indexing causal inference in animal models. In this line, Dokka and colleagues (2019) recently demonstrated that non-human primates perform causal inference in determining heading direction in the presence of independent object motion (i.e., object motion must be parsed from optic flow caused by self-location for appropriate heading perception). Mohl and colleagues (2019) similarly showed that both humans and rhesus monkeys make either a single or multiple saccade(s) to audio-visual targets depending on their disparity and in line with causal inference. Interestingly, however, while human behavior was best explained by model averaging, non-human primate behavior was most consistent with model selection. Whether this latter effect is a true difference between humans and non-human primates, or whether it is a corollary of the fact that the monkeys were trained on the specific task – and during training animals are rewarded for committing to a single (and correct) word-view – will be an interesting area for future study (see Noel

et al., 2021, for an example experimental ecosystem that should allow the study of causal inference without explicit training, and thus without putatively shaping task strategies).

Lastly, Fang and colleagues (2019) had non-human primates reach toward a target during different levels of visuo-proprioceptive disparities (i.e., real hand position vs. visual rendering of a dummy hand). Results suggested no bias during congruent visuo-proprioceptive presentations, and a saturating level of reaching end-point error as the visuo-proprioceptive conflict grew. Moreover, these researchers recorded single units in the premotor cortex and neural activity in this area was similarly modulated by visuo-proprioceptive conflict. Overall, therefore, Fang et al., 2019, showed behavior and neural activity consistent with causal inference. Further, these results imply that not only visuo-vestibular self-motion perception may be rooted in causal inference – the example par excellence in the study of cue combination and probabilistic coding (Gu et al., 2006, 2008; Fetsch et al., 2009, 2012, 2013; Dokka et al., 2019; Ma et al., 2006; Hou et al., 2019) – but also aspects more personal to the self, such as body ownership, may be rooted in this computation. Below we further explore the “self” in “self-motion”.

4. Self-Location as an Initial Condition

As described so far, successful navigation via path integration depends on both the visual and vestibular sense, and on the integration of these to generate accurate and precise self-motion and heading estimates. In turn, it is thought that the continual integration of self-motion velocity estimates generates a dynamic sense of self-location (although this process is generally less studied, particularly within a computational framework; see Lakshminarasimhan et al., 2018; Noel et al., 2020a for recent exceptions). These processes are routinely considered to be central in the study of spatial navigation. However, there is another critical condition that is seldom considered within the spatial navigation literature; an initial condition. Our initial sense of self-location must be correct to enable successful navigation.

Note, where “I” am, and where my body is, are typically one and the same, but need not be, as demonstrated by neurological conditions such as heautoscopy, autoscopic hallucinations, and out-of-body experiences (see Blanke & Metzinger, 2009). The study of static (i.e., prior to movement) and egocentric self-location is typically considered within the broader study of bodily self-consciousness (Blanke, 2012) and in conjunction with our subjective experience of body ownership and having a first-person perspective on the environment (Blanke & Metzinger, 2009; the study of the location of the body is also widely considered in the rodent literature, but mostly from an allocentric encoding point of view, see Barry & Burgess 2014 for a review, and the dissociation between body and self-position is hard in rodents). Philosophically, it has been argued that these three together – a sense of being encapsulated within a body that belongs to ‘me’ (body ownership), that is located at a specific location within the external environment (self-location), and from where ‘I’ perceive (first-person perspective) – constitute the minimal requirement for a pre-reflective phenomenal selfhood (Damasio, 2000; Legrand, 2006; Carruthers, 2008; Blanke & Metzinger, 2009).

Empirically, this area of investigation was jumpstarted by a seminal contribution from Botvinick & Cohen (1998) who demonstrated that by providing touch on participants' real hand while synchronously showing touch on a dummy hand, they could elicit hand ownership over a rubber hand (i.e., the "rubber-hand illusion"). Further, when subjects were asked to close their eyes and indicate the location of their real hand, they were systematically biased toward the rubber hand (i.e., "proprioceptive drift"). Over the following 20 years a number of similar illusions have been developed (e.g., faces, legs, tongue, and even tails, in rodents; Wada et al., 2016, 2019) and most importantly, the computational and neural correlates of the rubber-hand illusion are being established. Interestingly, recent models have casted the process of limb-ownership as a process of Bayesian causal inference (Samad et al., 2015) and have postulated that neural networks dedicated to encoding the space near our bodies (see below) act as a coupling prior between our body and what is near us (Noel et al., 2018a). As such, the computational principles (e.g., Bayesian observer with particular priors and performing causal inference) underlying inferences of the world around us and of ourselves within it may largely overlap. The recent neurophysiology recordings from Fang and colleagues (2019) equally support this speculation by demonstrating that reaches during visuo-proprioceptive conflicts were in line with causal inference and that firing patterns in the premotor cortex reflected this computation.

Now, empirical results (Rohde et al., 2011) have shown that the subjective sense of embodiment over a limb and the sense of where it is located in external space do not necessarily co-vary. And more importantly, while studies derived from the rubber-hand illusion are interesting in and of themselves, a change in the subjective location of one's hand is still described from an unmoved egocentric location and perspective. That is, it does not involve a manipulation of our reference frame as a whole, a global translation in space. To tackle this more general question – and one that ought to impact the initial conditions during self-motion guided navigation – Blanke and colleagues, as well as Ehrsson and colleagues, devised a manipulation similar to the rubber hand illusion but applied to the whole body. These researchers administered touch either on the back (Lenggenhager et al., 2007) or chest (Ehrsson, 2007) of participants who equally viewed synchronous (or asynchronous, control) touch being applied far from their location (i.e., ~ 2 meters in front). In both studies participants reported subjective experiences somewhat akin to that of out-of-body experiences (Blanke et al., 2004; De Ridder et al, 2007). Further, in Lenggenhager and colleagues' (2007) protocol subjects were blindfolded and moved backward from their original location. When asked to return to their initial spatial location via path integration, they overshot their target, as if returning to a location in between their initial physical location and that of the avatar they felt body ownership over. That is, a visuo-tactile manipulation was able to induce a subjective sense of embodiment over a virtual avatar and to perturb the subjective sense of self-location.

To the best of our knowledge, single unit recordings during full-body illusions such as those described above have not been performed. This would be particularly interesting, not only for the full body (vs. body part) ownership piece, but more so for the corollary that these illusions have on what has been denominated the "spatial aspects" of bodily self-consciousness; self-location and first-person perspective (Blanke, 2012). In this vein,

there is a well-established neural circuit that largely overlaps with that for optic flow and self-motion processing that is widely considered to play a fundamental role in bodily self-consciousness generally, and in its spatial aspects in particular. Fang et al., 2019, recorded from ventral premotor and found correlates of arm reaching errors. Approximately 20 years early, Graziano et al., 2000, recorded single unit activity from parietal area 5 during a rubber hand illusion and found that the activity of these neurons was influenced by the location of the rubber hand after synchronous but not asynchronous (control) visuo-tactile stroking. Both these reports concern body-part (i.e., hand) ownership and not self-location. Remarkably, however, these areas house neurons encoding for peripersonal space (PPS), and there seems to be a strong association between PPS and self-location.

Peripersonal space (PPS) is the space immediately adjacent to and surrounding one's body (Serino, 2019). This space is encoded by a fronto-parietal network composed of multisensory neurons in ventral pre-motor cortex (areas F4 and F5; Fogassi et al., 1996), VIP (Colby et al., 1993), and 7b (Hyvärinen, 1981) among other areas (see Clery et al., 2015a for a recent and extensive review). These neurons respond both to touch on the body, and to visual or auditory stimuli when these are presented near, but not far, from one's body. That is, they map the body and the space near it (~30cm in depth, but this is body-part specific and highly heterogeneous). The receptive fields of these neurons are anchored to the body, in that visual responses are largely independent of gaze position (particularly true in pre-motor areas) and instead follow the movement of specific body parts or the body as a whole (Graziano et al., 1997). These areas receive projections from earlier motion processing regions such as MSTd, and as a consequence it is no surprise that they are velocity (Fogassi et al., 1996; Noel et al., 2018b) and motion direction (Duhamel et al., 1998) selective (particularly sensitive to looming stimuli). Similarly, both the premotor areas and VIP are activated by large field optic flow stimulation and by vestibular input (Chen et al., 2011; Bremmer et al., 2001, 2002). Finally, the pre-motor neurons in this network seem to preferentially respond during voluntary as opposed to passive head rotation (Graziano et al., 1997). As a whole, therefore, there is a spatial code that specifically maps the body and the space near it (and seems involved in body ownership; Graziano et al., 2000; Fang et al., 2019), and this code is largely overlapping and interdependent with the areas highlighted earlier as encoding self-motion and heading perception (e.g., optic flow and vestibular translation responses, differentiating between active and passive movement).

Psychophysical methods have been developed to study PPS in humans, and many of these rely on indexing tactile detection facilitation when exteroceptive sensory signals (audition or vision) are presented near as opposed to far from the body (Serino et al., 2015, 2018). In addition to replicating in humans many of the earlier findings from the monkey electrophysiology literature, these methods have advanced our understanding of PPS and self-location in two aspects. First, Noel et al., 2015, behaviorally mapped peri-trunk encoding in the front and backspace during a full-body illusion. As expected, participants reported feeling ownership over a virtual avatar placed in front of them, and more importantly, their PPS shrank in the back while it expanded in the front – as if translating forward to encode not the location of their physical body, but their subjective self-location. This finding mimics that of neural responses during the rubber-hand illusion (Graziano et al., 2000) and has been replicated while rendering subliminal both the stimuli

eliciting the full-body illusion and the stimuli used for mapping PPS (Salomon et al., 2017). Second, a plethora of results have highlighted the incredible plasticity of PPS (Noel et al., 2020b), remapping due to personality traits and the perceived danger of our surrounding environment (e.g., Sambo & Iannetti, 2013), our social context (e.g., Teneggi et al., 2013; Noel et al., 2020c), and the state space of potential actions (see Bufacchi & Iannetti, 2018, Serino, 2019 for reviews). Given these observations, the general agreement is that PPS serves as an interface between self and environment, is involved in defensive behaviors (see Graziano & Cooke, 2006), and likely computes time-to-contact or impact prediction (Clery et al., 2015b).

The latest interpretation of PPS as involved in impact prediction may be colored by the fact that PPS is most sensitive to looming stimuli and often studied in static individuals. In a more active setting, however, we would attribute cause to the agent and not to the external environment, and thus we may re-phrase this interpretation as PPS predicting the future location of the body (and not the future location of objects in the environment). In fact, PPS remapping has been shown to anticipate arm movement (Brozzoli et al., 2010; Patane et al., 2018) and its size to enlarge during full-body actions such as walking (Noel et al., 2014). This emphasis on PPS as encoding i) subjective self-location, and ii) anticipating future self-locations may be particularly fruitful in imbedding the study of the bodily self within the study of self-motion, and conversely, in furthering our understanding of path integration. That is, incoming sensory evidence is by definition egocentric, and the parietal cortex seems outfitted to process this information; from edge detection to motion detection to a multisensory estimate of self-motion. However, eventually this egocentric information must converge with the spatial codes of the hippocampal formation (e.g., grid and place cells). Thus, just as clear spatial codes exist in the limbic system (e.g., place, grid, border, and speed cells), it is useful to define and identify spatial codes that exist in the parietal cortex. Via PPS we have one - we have an egocentric encoding of self-location and future potential locations (see Moon et al., 2020, for recent evidence suggesting that bodily self-consciousness impacts the tuning of spatial codes in the hippocampal formation). Relatedly, reinforcement learning models have emphasized that codes that represent future relations may be particularly useful in navigating state spaces (Dayan, 1993; Gershman, 2018) and within this framework some (Stachenfeld et al., 2017; Behrens et al., 2018) have reinterpreted place cells as encoding an animal's best estimate of where it will be in the immediate future, one step ahead, as opposed to its current location. Arguably, this desideratum is accomplished in parietal cortex by the PPS network. While subjective self-location may be encoded by a population of place cells (Robinson et al., 2020), it may be encoded in individual PPS neurons.

A last aspect worth briefly mentioning related to the study of self-location is that of first-person perspective. This can broadly be defined as one's outlook on the environment, an outlook that is *directed* at external components of the environment (see Blanke & Metzinger, 2009, for more detail). Most often one navigates in their heading direction, and thus first-person perspective is thought to be an important component to self-location (but see autoscopic hallucinations vs. Out-of-body experiences as two neurological conditions defined by a differential relationship between self-location and first-person perspective; Blanke & Metzinger, 2009). Importantly, however, first-person perspective is not exclusively

defined by one's visual viewpoint. To demonstrate this, Ionta and colleagues (2011) had participants experience a full-body illusion while lying in a supine position. Subjects viewed an avatar in virtual reality that provided conflicting information; while gravity on participant's real body suggested a vector pointing downward, the visual image presented suggested that gravity was pointing upward for the seen body. During the synchronous visuo-tactile condition participants reported feeling ownership over the virtual avatar. Most interestingly, approximately half the subjects perceived themselves to be lying under the seen avatar, and thus during the illusion moved upward. The other half of subjects were more influenced by observed rather than felt gravity, and thus in seeing the back of an avatar in front of them, felt as if they were viewing this subject from above. During the illusory condition, they felt their self-location to be closer to the ground than in the asynchronous control condition (Ionta et al., 2011). Together, these data show that subjective self-location can generally be fooled by visuo-tactile stimulation, and further that experienced direction of first-person perspective depends on a balance between visual and vestibular cues, and this outlook may affect the perceived direction of self-motion.

5. Outlook and Concluding Remarks

Admittedly, our review on the state of knowledge regarding the neural underpinning and computation of self-motion is broad in scope. Importantly, we find this to be an imperative toward building true knowledge, and consider that the ability to leverage implementation, algorithmic, and computational (maybe even philosophical!) insights to reciprocally inform one another is a true asset. Maybe even the envy of other fields of study.

We have detailed the known cortical circuit involved in visual optic flow processing, as well as the subcortical and cortical networks involved in vestibular processing. Perhaps more importantly, we have highlighted that visuo-vestibular integration is a necessity for accurate and precise self-motion guided navigation. Gratifyingly, studying how these senses are combined for the purpose of self-motion estimation has allowed to more generally sketch the common principles underlying cue combination as a whole, and has hopefully informed the study of probabilistic coding.

Of course, however, for as much as we have learned, there is so much more we do not yet understand. As underlined in previous sections, the exact role of different elements in the neural circuitry are not yet clear. Similarly, there seem to be an inherent tension between information converging in certain areas for or prior to integration (e.g., LIP) versus more distributed schemes. The neural underpinning of causal inference, a general computation for attributing likely causes to observations – particularly relevant for path integration during independent object motion, but applicable to all sorts of problems – is not understood. Further, our understanding of basic elements such as how self-motion velocity estimates get accumulated over protracted periods of time, or how are initial conditions set (i.e., self-location), are only in their infancy and not always considered. In fact, a recent psychophysical study in humans has suggested that *only* optic flow with time-varying velocity (i.e., evolving sequence of flow) is informative vis-à-vis heading direction (Burlingham & Heeger, 2020). This example highlights that we do not yet quite understand even what particular elements of visual signals guide self-motion. Thus, the challenges

moving forward are numerous and will be important in furthering our understanding of brain function. The next decades should see major advances, and we couldn't be more excited to go along for the ride.

Acknowledgements

J.P.N and D.E.A are supported by NIH U19NS118246.

Literature Cited

- Acerbi L, Dokka K, Angelaki DE, Ma WJ. (2018). Bayesian comparison of explicit and implicit causal inference strategies in multisensory heading perception. *PLoS Comput Biol*, 14:e1006110. [PubMed: 30052625]
- Adelson EH, and Movshon JA (1982). Phenomenal coherence of moving visual patterns. *Nature* 300, 523–525. doi: 10.1038/300523a0 [PubMed: 7144903]
- Andersen RA, Essick GK, Siegel RM. (1985). Encoding of Spatial Location by Posterior Parietal Neurons. *Science*; 230(4724):456–458. doi: 10.2307/1696086 [PubMed: 4048942]
- Avillac M, Deneve S, Olivier E, Pouget A, and Duhamel JR. (2005). Reference frames for representing visual and tactile locations in parietal cortex. *Nat Neurosci* 8: 941–949. [PubMed: 15951810]
- Barry C, Burgess N (2014). Neural Mechanisms of Self-Location. *Curr Biol*, 24 (8):R330–R339. [PubMed: 24735859]
- Beck JM, Ma WJ, Kiani R, Hanks T, Churchland AK, Roitman J, Shadlen MN, Latham PE and Pouget A (2008). Probabilistic population codes for Bayesian decision making. *Neuron*, 60, 1142–1152 [PubMed: 19109917]
- Behrens TE, Muller TH, Whittington JCR, Mark S Baram AB, Stachenfeld KL, Kurth-Nelson Z (2018). What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron* 100, 490–509 [PubMed: 30359611]
- Berkes P Orban G, Lengyel M, Fiser J (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science*, 331:83 [PubMed: 21212356]
- Bicanski A, Burgess N. (2016). Environmental Anchoring of Head Direction in a Computational Model of Retrosplenial Cortex. *The Journal of Neuroscience*; 36(46):11601–11618 [PubMed: 27852770]
- Bizley JK, Jones GP and Town SM (2016). Where are multisensory signals combined for perceptual decision-making? *Current Opinion in Neurobiology*, 40, 31–37 [PubMed: 27344253]
- Blanke O (2012). Multisensory brain mechanisms of bodily self-consciousness. *Nat. Rev. Neurosci* 13, 556–571. [PubMed: 22805909]
- Blanke O, and Metzinger T (2009). Full-body illusions and minimal phenomenal selfhood. *Trends Cogn. Sci* 13, 7–13. [PubMed: 19058991]
- Blanke O, Landis T, Spinelli L, & Seeck M (2004). Out-of-body experience and autoscopia of neurological origin. *Brain*, 127(2), 243–258. 10.1093/brain/awh040 [PubMed: 14662516]
- Bogacz R, Brown E, Moehlis J, Holmes P, Cohen JD. 2006. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review* 113:700–765. doi: 10.1037/0033-295X.113.4.700. [PubMed: 17014301]
- Botvinick M, & Cohen J (1998). Rubber hands “feel” touch that eyes see. *Nature*, 391(6669), 756. 10.1038/35784 [PubMed: 9486643]
- Boussaoud D, Ungerleider LG, 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* 296:462–95 [PubMed: 2358548]
- Bremmer F, Duhamel J-R, Ben Hamed S and Graf W (2002). Heading encoding in the macaque ventral intraparietal area (VIP). *European Journal of Neuroscience*, 16, 1554–1568
- Bremmer F, Schlack A, Duhamel JR, Graf W, Fink GR, (2001). Space coding in primate posterior parietal cortex. *Neuroimage* 14, S46–51. doi:10.1006/nimg.2001.0817 [PubMed: 11373132]

- Britten KH (2008). Mechanisms of self-motion perception. *Annual review of neuroscience* 31, 389–410
- Britten KH, and Van Wezel RJ. (1998). Electrical microstimulation of cortical area MST biases heading perception in monkeys. *Nat Neurosci* 1: 59–63. [PubMed: 10195110]
- Britten KH, Newsome WT, Shadlen MN, Celebrini S and Movshon JA (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Vis Neurosci*, 13, 87–100 [PubMed: 8730992]
- Brooks JX, Carriot J & Cullen KE Learning to expect the unexpected: rapid updating in primate cerebellum during voluntary self-motion. *Nat. Neurosci* 18, 1310–1317 (2015). [PubMed: 26237366]
- Brozzoli C, Cardinali L, Pavani F, Farnè A. Action-specific remapping of peripersonal space. *Neuropsychologia* 2010, 48(3):796–802 [PubMed: 19837102]
- Bryan AS, and Angelaki DE. (2008). Optokinetic and vestibular responsiveness in the macaque rostral vestibular and fastigial nuclei. *J Neurophysiol* 101: 714–720. [PubMed: 19073813]
- Bufacchi RJ, Iannetti GD (2018) An action field theory of peripersonal space. *Trends Cogn Sci* 22:1076–1090 [PubMed: 30337061]
- Burlingham CS, & Heeger DJ (2020). Heading Perception Depends on Time-Varying Evolution of Optic Flow. *bioRxiv* 356758; doi: 10.1101/356758
- Cao Y, Summerfield C, Park H, Giordano BL, Kayser C. (2019). Causal Inference in the Multisensory Brain. *Neuron* Jun 5;102(5):1076–1087.e8. doi:10.1016/j.neuron.2019.03.043 [PubMed: 31047778]
- Carandini M Heeger DJ (1994). Summation and division by neurons in primate visual cortex. *Science* 264, 1333–1336. [PubMed: 8191289]
- Carandini M, Heeger DJ (2011). Normalization as a canonical neural computation. *Nat Rev Neurosci* 13, 51–62. [PubMed: 22108672]
- Carriot J, Brooks JX & Cullen KE Multimodal integration of self-motion cues in the vestibular system: active versus passive translations. *J. Neurosci* 33, 19555–19566 (2013). [PubMed: 24336720]
- Chen LL, Lin LH, Green EJ, Barnes CA, McNaughton BL. 1994a. Head-direction cells in the rat posterior cortex—I. Anatomical distribution and behavioral modulation. *Exp Brain Res* 101:8–23. [PubMed: 7843305]
- Chen A, Gu Y, Takahashi K, Angelaki DE, Deangelis GC. 2008. Clustering of self-motion selectivity and visual response properties in macaque area MSTd. *J Neurophysiol* 100:2669 2683. [PubMed: 18753323]
- Chen A, DeAngelis GC and Angelaki DE (2011). Representation of vestibular and visual cues to self-motion in ventral intraparietal cortex. *Journal of Neuroscience*, 31, 12036–12052 [PubMed: 21849564]
- Chen A, DeAngelis GC and Angelaki DE (2013). Functional Specializations of the Ventral Intraparietal Area for Multisensory Heading Discrimination. *Journal of Neuroscience*, 33, 3567–3581. [PubMed: 23426684]
- Chen A, Deangelis GC, and Angelaki DE. (2010). Macaque parieto-insular vestibular cortex: Responses to self-motion and optic flow. *J Neurosci* 30: 3022–3042. [PubMed: 20181599]
- Churchland AK, Kiani R, Chaudhuri R, Wang X-J, Pouget A and Shadlen MN (2011). Variance as a signature of neural computations during decision making. *Neuron*, 69, 818–831. [PubMed: 21338889]
- Clark JJ, Yuille AL. 1990. *Data fusion for sensory information processing systems* Boston: Kluwer Academic.
- Cléry J, Guipponi O, Odouard S, Wardak C, and Ben Hamed S (2015b). Impact prediction by looming visual stimuli enhances tactile detection. *J. Neurosci* 35, 4179–4189. doi: 10.1523/JNEUROSCI.3031-14.2015 [PubMed: 25762665]
- Cléry J, Guipponi O, Wardak C, and Ben Hamed S (2015a). Neuronal bases of peripersonal and extrapersonal spaces, their plasticity and their dynamics: knowns and unknowns. *Neuropsychologia* 70, 313–326. doi: 10.1016/j.neuropsychologia.2014.10.022 [PubMed: 25447371]

- Colby CL, Duhamel JR, Goldberg ME, (1993). Ventral intraparietal area of the macaque: anatomic location and visual response properties. *Journal of Neurophysiology* 69, 902–914. [PubMed: 8385201]
- Cullen KE (2019). Vestibular processing during natural self-motion: implications for perception and action. *Nat. Rev. Neurosci* 20, 346–363 [PubMed: 30914780]
- Dayan P (1993). Improving Generalisation for Temporal Difference Learning: The Successor Representation. *Neural Comput* 5, 613–624.
- De Ridder D, Van Laere K, Dupont P, Menovsky T, & Van de Heyning P (2007). Visualizing outof-body experience in the brain. *New England Journal of Medicine*, 357(18), 1829–1833. 10.1056/NEJMoa070010
- Dichgans J, Brandt T. (1978). Visual-vestibular interaction. Effects on self-motion perception and postural control. In: Held R, Leibowitz HW, Teuber HL (eds) *Handbook of sensory physiology* Springer, Berlin Heidelberg New York 755–804
- Dokka K, Park H, Jansen M, DeAngelis GC, Angelaki DE. (2019). Causal inference accounts for heading perception in the presence of object motion. *Proc Natl Acad Sci Apr* 17;116(18):9060 5. doi:10.1073/pnas.1820373116 [PubMed: 30996126]
- Drugowitsch J, DeAngelis GC, Angelaki DE and Pouget A (2015). Tuning the speed accuracy trade-off to maximize reward rate in multisensory decision-making. *Elife*, 4.
- Drugowitsch J, DeAngelis GC, Klier EM, Angelaki DE and Pouget A (2014). Optimal multisensory decision-making in a reaction-time task. *Elife*, 3.
- Duffy CJ, and Wurtz RH. (1991). Sensitivity of MST neurons to optic flow stimuli: I. A continuum of response selectivity to large-field stimuli. *J Neurophysiol* 65: 1329–1345. [PubMed: 1875243]
- Duffy CJ (1998). MST neurons respond to optic flow and translational movement. *Journal of Neurophysiology*, 80, 1816–1827 [PubMed: 9772241]
- Duhamel J-R, Colby CL, Goldberg ME, (1998). Ventral Intraparietal Area of the Macaque: Congruent Visual and Somatic Response Properties. *Journal of Neurophysiology* 79, 126–136 [PubMed: 9425183]
- Ehrsson HH (2007). The experimental induction of out-of-body experiences. *Science* 317, 1048. [PubMed: 17717177]
- Einstein A (1907). Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen. *Jahrbuch der Radioaktivität und Elektronik*, 4, 411–462.
- Ernst MO & Banks MS (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429–433 [PubMed: 11807554]
- Fang W, Li J, Qi G, Li S, Sigman M, & Wang L (2019). Statistical inferences of body representation in the macaque brain. *Proc Natl Acad Sci U S A* 10.1073/pnas.1902334116
- Fanini A and Assad JA (2009). Direction selectivity of neurons in the macaque lateral intraparietal area. *Journal of neurophysiology*, 101, 289–305. [PubMed: 18987126]
- Fetsch CR, Turner AH, DeAngelis GC, Angelaki DE (2009). Dynamic reweighting of visual and vestibular cues during self-motion perception. *J Neurosci* 29: 15601–15612 [PubMed: 20007484]
- Fetsch CR, DeAngelis GC and Angelaki DE (2013). Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. *Nature Review Neuroscience*, 14, 429–442 [PubMed: 23686172]
- Fetsch CR, Pouget A, DeAngelis GC and Angelaki DE (2012). Neural correlates of reliability-based cue weighting during multisensory integration. *Nature Neuroscience*, 15, 146 154.
- Fogassi L, Gallese V, Fadiga L, Luppino G, Matelli M, and Rizzolatti G (1996). Coding of peripersonal space in inferior premotor cortex (area F4). *J. Neurophysiol* 76, 141–157. [PubMed: 8836215]
- Fiser J, Berkes P, Orban G, Lengyel M (2010). Statistically optimal perception and learning: from behavior to neural representations. *Trends Cogn Sci* 114: 119–130
- French RL, DeAngelis GC (2020) Multisensory neural processing: from cue integration to causal inference. *Current Opinion in Physiology*
- Gershman SJ (2018). The successor representation: Its computational logic and neural substrates. *Journal of Neuroscience*, 38, 7193–7200. [PubMed: 30006364]

- Gibson JJ (1950). *The perception of the visual world* Boston: Houghton Mifflin
- Graziano MS, Cooke DF, and Taylor CS (2000). Coding the location of the arm by sight. *Science* 290, 1782–1786. [PubMed: 11099420]
- Graziano MS, Hu XT, and Gross CG (1997). Visuospatial properties of ventral premotor cortex. *J. Neurophysiol* 77, 2268–2292. [PubMed: 9163357]
- Graziano MS, Cooke DF (2006) Parieto-frontal interactions, personal space, and defensive behavior. *Neuropsychologia* 44:845–859. [PubMed: 16277998]
- Gu Y, Angelaki DE and DeAngelis GC (2008). Neural correlates of multisensory cue integration in macaque MSTd. *Nature Neuroscience*, 11, 1201–1210 [PubMed: 18776893]
- Gu Y, DeAngelis GC and Angelaki DE (2012). Causal links between dorsal medial superior temporal area neurons and multisensory heading perception. *Journal of Neuroscience*, 32, 2299–2313 [PubMed: 22396405]
- Gu Y, Watkins PV, Angelaki DE and DeAngelis GC (2006). Visual and nonvisual contributions to threedimensional heading selectivity in the medial superior temporal area. *Journal of Neuroscience*, 26, 73–85 [PubMed: 16399674]
- Hafting T, Fyhn M, Molden S, Moser MB, and Moser EI (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature* 436, 801–806. [PubMed: 15965463]
- Hardcastle K, Maheswaranathan N, Ganguli S & Giocomo LM (2017). A multiplexed, heterogeneous, and adaptive code for navigation in medial entorhinal cortex. *Neuron* 94, 375–387 [PubMed: 28392071]
- Helmholtz H (1867). *Handbuch der Physiologischen Optik* Leipzig: Voss. (English translation. 1924 JPC Southall as *Treatise on Physiological Optics*)
- Herweg NA Kahana MJ (2018). Spatial representations in the human brain. *Front Hum Neurosci*, 12(297)
- Hillis JM, Ernst MO, Banks MS, Landy MS. (2002). Combining sensory information: mandatory fusion within, but not between, senses. *Science* 298: 1627–1630. [PubMed: 12446912]
- Hillis JM, Watt SJ, Landy MS, Banks MS. (2004). Slant from texture and disparity cues: optimal cue combination. *J Vis* 4: 967–992 [PubMed: 15669906]
- Hoyer P, Hyvarinen A (2003). Interpreting neural response variability as Monte Carlo sampling of the posterior NIPS.
- Hou H, Gu Y (2020). Multisensory integration for self-motion perception. Reference module in neuroscience and biobehavioral psychology
- Hou H, Zheng Q, Zhao Y, Pouget A and Gu Y (2019). Neural Correlates of Optimal Multisensory Decision Making under Time-Varying Reliabilities with an Invariant Linear Probabilistic Population Code. *Neuron*, 104, 1010–1021.e10. [PubMed: 31607423]
- Hubel DH, and Wiesel TN (1968). Receptive fields and functional architecture of monkey striate cortex. *J. Physiol* 195, 215–243. doi: 10.1113/jphysiol.1968.sp008455 [PubMed: 4966457]
- Huk AC, Katz LN, and Yates JL (2017). The role of the lateral intraparietal area in (the study of) decision making. *Annu. Rev. Neurosci* 40, 349–372. [PubMed: 28772104]
- Hulse BK & Jayaraman V (2020). Mechanism underlying the neural computation of head direction. *Annual Review of Neuroscience*
- Hyvärinen J, (1981). Regional distribution of functions in parietal association area 7 of the monkey. *Brain Res* 206, 287–303. [PubMed: 7214136]
- Ionta S, Heydrich L, Lenggenhager B, Mouthon M, Fornari E, Chapuis D, Gassert R, & Blanke O (2011). Multisensory mechanisms in temporo- parietal cortex support self-location and first-person perspective. *Neuron* 70, 363–374. [PubMed: 21521620]
- Jones EG. (1985). *The thalamus* (Plenum, New York)
- Katz LN, Yates JL, Pillow JW, and Huk AC (2016). Dissociated functional significance of decision-related activity in the primate dorsal stream. *Nature* 535, 285–288 [PubMed: 27376476]
- Keshavarzi S, Bracey EF, Faville RA, Campagner D, Tyson AL, Lenzi SC, Branco T, Margrie TW (2021). The retrosplenial cortex combines internal and external cues to encode head velocity during navigation. *bioRxiv* 2021.01.22.42778

- Kim SS, Hermundstad AM, Romani S, Abbott LF, and Jayaraman V (2019). Generation of stable heading representations in diverse visual scenes. *Nature* 576, 126–131. [PubMed: 31748750]
- Knierim JJ, Zhang K. 2012. Attractor dynamics of spatially correlated neural activity in the limbic system. *Annu. Rev. Neurosci* 35:267–85 [PubMed: 22462545]
- Knill DC and Richards W (1996). *Perception as Bayesian inference* Cambridge, UK: Cambridge University Press.
- Kobayashi Y, Amaral DG. (2000). Macaque monkey retrosplenial cortex: I. Three-dimensional and cytoarchitectonic organization. *J Comp Neurol* 426:339–365. [PubMed: 10992242]
- Kording KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB and Shams L (2007). Causal inference in multisensory perception. *PLoS One*, 2, e943. [PubMed: 17895984]
- Krakauer JW & Mazzoni P Human sensorimotor learning: adaptation, skill, and beyond. *Curr. Opin. Neurobiol* 21, 636–644 (2011) [PubMed: 21764294]
- Kropff E, Carmichael JE, Moser MB, Moser EI. (2015). Speed cells in the medial entorhinal cortex. *Nature* 523:419–424. doi:10.1038/nature14622 [PubMed: 26176924]
- Lakshminarasimhan KJ, Avila E, Neyhart E, DeAngelis GC, Pitkow X, Angelaki D (2020). Tracking the Mind's Eye: Primate Gaze Behavior during Virtual Visuomotor Navigation Reflects Belief Dynamics. *Neuron*, 106, 1–13 [PubMed: 32272061]
- Lakshminarasimhan KJ, Petsalis M, Park H, DeAngelis GC, Pitkow X, Angelaki DE. (2018). A Dynamic Bayesian Observer Model Reveals Origins of Bias in Visual Path Integration. *Neuron* 14:99:194–206.e5.
- Laurens J and Angelaki DE (2017). A unified internal model theory to resolve the paradox of active versus passive self-motion sensation. *eLife*, 6.
- Laurens J, Liu S, Yu X-J, Chan R, Dickman D, DeAngelis GC and Angelaki DE (2017). Transformation of spatiotemporal dynamics in the macaque vestibular system from otolith afferents to cortex. *Elife*, 6, e20787 [PubMed: 28075326]
- Lenggenhager B, Tadi T, Metzinger T, and Blanke O (2007). Video ergo sum: manipulating bodily self-consciousness. *Science* 317, 1096–1099. [PubMed: 17717189]
- Lopez C & Blanke O The thalamocortical vestibular system in animals and humans. *Brai Res. Rev* 67, 119–146 (2011).
- Ma WJ (2019). Bayesian decision models: A primer. *Neuron*, 104 (1), 164–175 [PubMed: 31600512]
- Ma WJ, Beck JM, Latham PE & Pouget A (2006). Bayesian inference with probabilistic population codes. *Nat. Neurosci* 9, 1432–1438 [PubMed: 17057707]
- Mach E (1875). *Grundlinien der Lehre von den Bewegungsempfindungen* Leipzig: Engelmann
- Magnotti JF, Beauchamp MS: A Causal Inference Model Explains Perception of the McGurk Effect and Other Incongruent Audiovisual Speech. *PLoS Comput Biol* 2017, 13:e1005229. [PubMed: 28207734]
- Maunsell JH, Van Essen DC. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *J Neurophysiol* 49:1127–1147 [PubMed: 6864242]
- Miller J 1982. Divided attention: evidence for coactivation with redundant signals. *Cognitive Psychology* 14:247–279. doi: 10.1016/0010-0285(82)90010-X. [PubMed: 7083803]
- Mohl JT, Pearson JM, Groh JM (2019). Monkeys and humans implement causal inference to simultaneously localize auditory and visual stimuli. *bioRxiv* 10.1101/823385.
- Morgan ML, DeAngelis GC and Angelaki DE (2008). Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron*, 59, 662–673. [PubMed: 18760701]
- Moser EI, Kropff E, and Moser M-B (2008). Place cells, grid cells, and the brain's spatial representation system. *Annu. Rev. Neurosci* 31, 69–89. [PubMed: 18284371]
- Moon H-J, Gauthier B, Park H-D, Faivre N, Blanke O (2020). Sense of self impacts spatial navigation and hexadirectional coding in human entorhinal cortex. *BioRxiv*, 10.1101/2020.09.13.295246
- Nikbakht N, Tafreshiha A, Zoccolan D and Diamond ME (2018). Supralinear and Supramodal Integration of Visual and Tactile Signals in Rats. *Psychophysics and Neuronal Mechanisms. Neuron*, 97, 626–639.e8. [PubMed: 29395913]

- Noel JP, Samad M, Doxon A, Clark J, Keller S, Di Luca M (2018). Peri-personal space as a prior in coupling visual and proprioceptive signals. *Sci Rep*, 8:15819. [PubMed: 30361477]
- Noel JP, Grivaz P, Marmaroli P, Lissek H, Blanke O, and Serino A (2014). Full body action remapping of peripersonal space: the case of walking. *Neuropsychologia* 70, 375–384. [PubMed: 25193502]
- Noel JP, Lakshminarasimhan KJ, Park H, Angelaki DE (2020). Increased variability but intact integration during visual navigation in Autism Spectrum Disorder *PNAS*, 117 (20) 11158–11166
- Noel JP, Caziot B, Burni S, Fitzgerald N, Avila E, Angelaki D (2021). Supporting generalization in non-human primate behavior by tapping into structural knowledge: Examples from sensorimotor mappings, inference, and decision-making. *Progress in Neurobiology*
- Noel JP, Stevenson R, Wallace M (2018). Atypical audiovisual temporal function in autism and schizophrenia; similar phenotype, different cause. *European Journal of Neuroscience*, 47 (10), 1230–1241
- Noel JP, Blanke O, Magosso E Serino A (2018). Neural Adaptation Accounts for the Resizing of Peri-Personal Space Representation: Evidence from a Psychophysical Computational Approach. *Journal of Neurophysiology*
- Noel JP, Bertoni T, Terrebonne E, Pellencin E, Herbelin B, Cascio C, ... Serino A (2020). Rapid Recalibration of Peri-Personal Space: Psychophysical, Electrophysiological, and Neural Network Modeling Evidence. *Cereb Cortex*
- Noel JP, Pfeiffer C, Blanke O, and Serino A (2015). Peripersonal space as the space of the bodily self. *Cognition* 144, 49–57. [PubMed: 26231086]
- O’Keefe J & Nadel L (1978). *The Hippocampus as a Cognitive Map* (Clarendon Press, Oxford)
- O’Keefe J (1976). Place units in the hippocampus of the freely moving rat. *Exp Neurol* 51: 78–109 [PubMed: 1261644]
- Odegaard B, Wozny DR, Shams L: Biases in Visual, Auditory, and Audiovisual Perception of Space. *PLoS Comput Biol* 2015, 11:e1004649. [PubMed: 26646312]
- Ohshiro T, Angelaki DE, Deangelis GC (2011). A normalization model of multisensory integration. *Nature neuroscience* 14, 775–782. [PubMed: 21552274]
- Ohshiro T, Angelaki DE, Deangelis GC (2017). A Neural Signature of Divisive Normalization at the Level of Multisensory Integration in Primate Cortex. *Neuron* 95, 399–411 e398 [PubMed: 28728025]
- Pandya DN, Seltzer B (1982). Intrinsic connections and architectonics of posterior parietal cortex in the rhesus monkeys. *J. Comp. Neurol* 204: 196–210. [PubMed: 6276450]
- Patane I, Cardinali L, Salemme R, Pavani F, Farne A, Brozzoli C (2018). Action planning modulates peri-personal space. *Journal of Cognitive Neuroscience*, doi: 10.1162/jocn_a_01349
- Prsa M, Gale S, and Blanke O (2012). Self-motion leads to mandatory cue fusion across sensory modalities. *J. Neurophysiol* 108, 2282–2291. [PubMed: 22832567]
- Raab DH. 1962. Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences* 24:574–590. doi: 10.1111/j.2164-0947.1962.tb01433.x. [PubMed: 14489538]
- Rahnev D, & Denison RN (2018). Suboptimality in perceptual decision making. *Behavioral and Brain Sciences*, e223. doi: 10.1017/S0140525X18000936
- Ramsey F (1926). Truth and probability. In Braithwaite RB (Ed.), *The foundations of mathematics and other logical essays* (pp. 156–198). London: Kegan, Paul, Trench, Trubner & Co.
- Raposo D, Sheppard JP, Schrater PR and Churchland AK (2012). Multisensory decision making in rats and humans. *Journal of Neuroscience*, 32, 3726–3735. [PubMed: 22423093]
- Ratcliff R and Rouder JN (1998). Modeling response times for two-choice decisions. *Psychological science*, 9, 347–356.
- Robinson NTM, Descamps LAL, Russell LE, Buchholz MO, Bicknell BA, Antonov GK, Lau JYN, Nutbrown R, Schmidt-Hieber C, Hausser M. 2020. Targeted activation of hippocampal place cells drives Memory-Guided spatial behavior. *Cell* 183:1586–1599. [PubMed: 33159859]
- Rohde M, Di Luca M, & Ernst MO (2011). The rubber hand illusion: feeling of ownership and proprioceptive drift do not go hand in hand. *PLoS One*, 6(6), e21659 [PubMed: 21738756]

- Rohe T, Ehlis AC, Noppeney U (2019). The neural dynamics of hierarchical Bayesian causal inference in multisensory perception. *Nat Commun*, 10:1907. [PubMed: 31015423]
- Rohe T, Noppeney U. (2015). Cortical Hierarchies Perform Bayesian Causal Inference in Multisensory Perception. *PLoS Biol* Feb 24;13(2):e1002073. doi:10.1371/journal.pbio.1002073 [PubMed: 25710328]
- Rohe T, Noppeney U. (2016). Distinct computational principles govern multisensory integration 684 in primary sensory and association cortices. *Curr Biol*. 2016;26(4):509–14. 685 doi:10.1016/j.cub.2015.12.056
- Roy JE & Cullen KE Dissociating self-generated from passively applied head motion: neural mechanisms in the vestibular nuclei. *J. Neurosci* 24, 2102–2111 (2004) [PubMed: 14999061]
- Roy JE & Cullen KE Selective processing of vestibular reafference during self-generated head motion. *J. Neurosci* 21, 2131–2142 (2001). [PubMed: 11245697]
- Salomon R, Noel JP, Lukowska M, et al. (2017). Unconscious integration of multisensory bodily inputs in the peripersonal space shapes bodily self-consciousness. *Cognition* 166: 174–183. [PubMed: 28577447]
- Sambo CF, & Iannetti GD (2013). Better safe than sorry? The safety margin surrounding the body is increased by anxiety. *J Neurosci*, 33(35), 14225–14230. [PubMed: 23986256]
- Samad M, Chung AJ, & Shams L (2015). Perception of body ownership is driven by Bayesian sensory inference. *PloS One*, 10(2), e0117178. 10.1371/journal.pone.0117178 [PubMed: 25658822]
- Sasaki R, Anzai A, Angelaki DE, DeAngelis GC (2020). Flexible coding of object motion in multiple reference frames by parietal cortex neurons. *Nature Neuroscience*,
- Sasaki R, Angelaki DE, DeAngelis GC (2019). Processing of object motion and self-motion in the lateral subdivision of the medial superior temporal area in macaques. *J Neurophysiol*, 121:1207–1221. [PubMed: 30699042]
- Savin C and Deneve S (2014). Spatio-temporal representations of uncertainty in spiking neural networks. In *Advances in Neural Information Processing Systems 2024–2032*
- Serino A (2019). Peripersonal space (PPS) as a multisensory interface between the individual and the environment, defining the space of the self. *Neuroscience and Biobehavioral Reviews*, 99, 138–159. 10.1016/j.neubiorev.2019.01.016 [PubMed: 30685486]
- Serino A, Noel J-P, Mange R, et al. (2018). Peripersonal space: an index of multisensory body–environment interactions in real, virtual, and mixed realities. *Front. ICT* 4. 10.3389/fict.2017.00031
- Serino A, Noel JP, Galli G, Canzoneri E, Marmoroli P, Lissek H, Blanke O (2015b). Body part-centered and full body-centered peripersonal space representations. *Sci Rep* 5:18603, DOI: 10.1038/srep18603 [PubMed: 26690698]
- Shalom S, Zaidel A. Better than optimal. *Neuron* 2018; 97: 484–7. [PubMed: 29420930]
- Siegel RM, and Read HL. (1997). Analysis of optic flow in the monkey parietal area 7a. *Cereb Cortex* 7: 327–346. [PubMed: 9177764]
- Sohn H, Narain D, Meirhaeghe N & Jazayeri M (2019). Bayesian Computation through Cortical Latent Dynamics. *Neuron* 103, 934–947.e5 [PubMed: 31320220]
- Solstad T, Boccara CN, Kropff E, Moser MB, Moser EI. Representation of geometric borders in the entorhinal cortex. *Science* 2008;322:1865–1868. [PubMed: 19095945]
- Stachenfeld KL, Botvinick MM, and Gershman SJ (2017). The hippocampus as a predictive map. *Nat. Neurosci* 20, 1643–1653. [PubMed: 28967910]
- Stavropoulos A, Lakshminarasimhan K, Laurens J, Pitkow X Angelaki D (2020). Influence of sensory modality and control dynamics on human path integration
- Tanaka K, Hikosaka K, Saito H, Yukie M, Fukada Y, and Iwai E. (1986). Analysis of local and wide-field movements in the superior temporal visual areas of the macaque monkey. *J Neurosci* 6: 134–144. [PubMed: 3944614]
- Teneggi C, Canzoneri E, di Pellegrino G, Serino A (2013) Social modulation of peripersonal space boundaries. *Curr Biol* 23:406–411. [PubMed: 23394831]
- Tolhurst D, Movshon J, & Dean A (1982). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research*, 23, 775–785.

- Tschermak A (1931). Optischer Raumsinn. In: Handbuch der normalen und pathologischen Physiologie Bethe A, Bergmann G, Embden G, Ellinger A (eds.). Berlin: Springer. Vo!. XII/2, pp. 834–1000.
- van Beers RJ, Sittig AC, Denier van der Gon JJ. (1996). How humans combine simultaneous proprioceptive and visual position information. *Exp Brain Res* 111: 253–261 [PubMed: 8891655]
- Vélez-Fort M, Bracey EF, Keshavarzi S, Rousseau CV, Cossell L, Lenzi SC, Strom M, and Margrie TW (2018) A circuit for integration of head- and visual-motion signals in layer 6 of mouse primary visual cortex. *Neuron*, 98: 179–191.e6. [PubMed: 29551490]
- Wada M, Takano K, Ora H, et al. (2016). The rubber tail illusion as evidence of body ownership in mice. *J Neurosci*;36:11133–7. [PubMed: 27798192]
- Wada M, Ide M, Atsumi T, et al. (2019). Rubber tail illusion is weakened in Ca(2p)-dependent activator protein for secretion 2 (Caps2)- knockout mice. *Sci Rep*;9:7552. [PubMed: 31101876]
- Walker EY, Cotton RJ, Ma WJ & Tolias AS (2020). A neural basis of probabilistic computation in visual cortex. *Nat. Neurosci* 23, 122–129. [PubMed: 31873286]
- Wei XX, & Stocker AA. (2015). A Bayesian observer model constrained by efficient coding can explain ‘anti-Bayesian’ percepts. *Nature Neuroscience*, 18(10), 1509–1517. DOI:10.1038/nn.4105 [PubMed: 26343249]
- Whitlock JR, Sutherland RJ, Witter MP, Moser MB, and Moser EI (2008). Navigating from hippocampus to parietal cortex. *Proc. Natl. Acad. Sci. U.S.A* 105, 14755–14762. doi: 10.1073/pnas.0804216105 [PubMed: 18812502]
- Wozny DR, Beierholm UR, and Shams L (2010). Probability matching as a computational strategy used in perception. *PLoS Comput. Biol* 6:e1000871. doi: 10.1371/journal.pcbi.1000871 [PubMed: 20700493]
- Yu X, and Gu Y (2018). Probing sensory readout via combined choice-correlation measures and microstimulation perturbation. *Neuron* 100, 715.e5–727.e5. doi: 10.1016/j.neuron.2018.08.034 [PubMed: 30244884]
- Zhang W, Wu S, Doiron B and Lee TS (2019). A Normative Theory for Causal Inference and Bayes Factor Computation in Neural Circuits, paper presented at Advances in Neural Information Processing Systems.
- Zhang WH, Chen A, Rasch MJ and Wu S (2016). Decentralized Multisensory Information Integration in Neural Systems. *Journal of Neuroscience*, 36, 532–547. [PubMed: 26758843]
- Zhang W-H, Wang H, Chen A, Gu Y, Lee TS, Wong KM and Wu S (2019). Complementary congruent and opposite neurons achieve concurrent multisensory integration and segregation. *eLife*, 8

BOX 1.**Neural Instantiation of Probabilistic Inference**

The central tenet of the Bayesian framework is that the brain represents uncertainty about the environment in the form of probability distributions. In the main text we have emphasized probabilistic population codes (PPCs; Ma et al., 2006; Beck et al., 2008; Hou et al., 2019) as a putative neural implementation of probabilistic inference, given that these have a strong history of accounting for cue combination, and in turn their application to this problem has supposed the bulk of their empirical support. Walker et al., 2020, have also recently shown strong support for this framework in demonstrating that trial-to-trial changes in the shape of likelihood functions as derived from a population of V1 neurons can account for fluctuations in behavior.

PPCs essentially suggest that the response of a neural population is proportional to parameters of probability distributions. Given this distributed format (i.e., a spatial code), this framework is thought to represent probabilities almost instantaneously, a great strength. However, this code has also been criticized, most commonly for only being able to represent a restrictive class of distributions and for its prohibitive computational cost in performing exact inference (e.g., Savin & Deneve, 2014). On the other end of the theoretical spectrum lie sampling models suggesting that the activity of each neuron within a population encodes a different random variable, and that neural activity represents samples drawn from a latent probability distribution (Hoyer & Hyvarinen, 2003). This second framework, relying on a temporal code, is slower than PPCs, but said to allow for easier marginalization, and accounts for trial-to-trial variability in single-unit neural variability (Fiser et al., 2010). Strong empirical support for the sampling framework comes from spontaneous and evoked V1 activity of developing ferrets showing a progressive adaptation of internal models (i.e., spontaneous activity) to the statistic of natural stimuli with age (Berkes et al., 2011). Relatedly, Sohn et al., 2019, recently argued that Bayesian computations depend on the shape (i.e., curvature) of cortical dynamics within a latent low dimensional space, thus also suggesting that neural activity defines a latent space where Bayesian computations occur.

More broadly, it must be noted that PPCs are a theory of statistical inference that occurs at the population level, while sampling puts the burden on single neurons. As such, these may not be mutually exclusive. In fact, Festa et al., 2020, recently suggested that sampling in V1 might account for poisson-like variability of single neurons. The outset of PPCs is exactly this form of variability, and thus we may conjecture that certain statistical inferences occur at the single neuron level and via sampling, while others occur at the population level via PPCs and poisson-like variability, the latter being inherited from the individual neurons and sampling.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

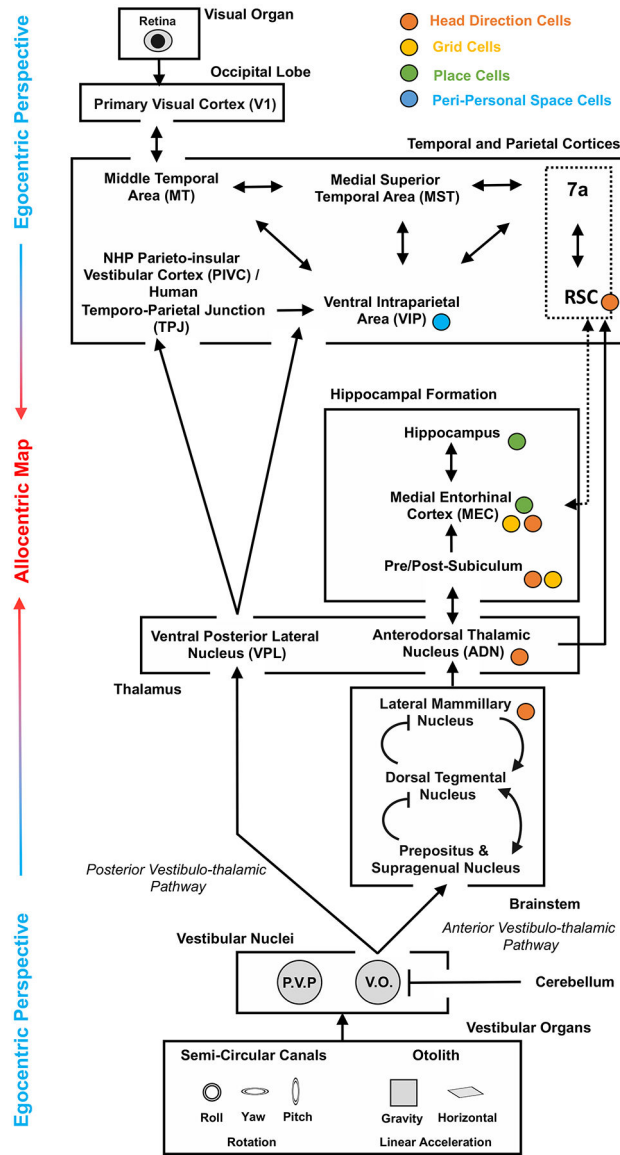


Figure 1. Visual and vestibular pathways leading to allocentric coding in parahippocampal formation.

Vestibular-only (VO) cells in the vestibular nuclei project via the anterior vestibulo-thalamic pathway to hippocampal formation, first reflecting an egocentric code – given the idiothetic nature of the vestibular system – and ending in an allocentric code (e.g., place fields). Via the posterior vestibulo-thalamic pathway, vestibular signals permeate much of the posterior parietal cortex. The exact nature and strength of the message-passing across much of this schematic network remain to be fully described, and this schematic coalesces evidence from a number of species; macaques, rodents, and fruit-flies. Thus, there are likely species-specific variations (e.g., head-direction cells exist in retrosplenial cortex (RSC) in rodents, Keshavarzi et al., 2021, yet this is unknown in macaque). Nonetheless, overall area 7a and RSC seem to be strong points of contact between egocentric coding in cortex and allocentric coding in the hippocampal formation (e.g., Whitlock et al., 2008; Keshavarzi et al., 2021).

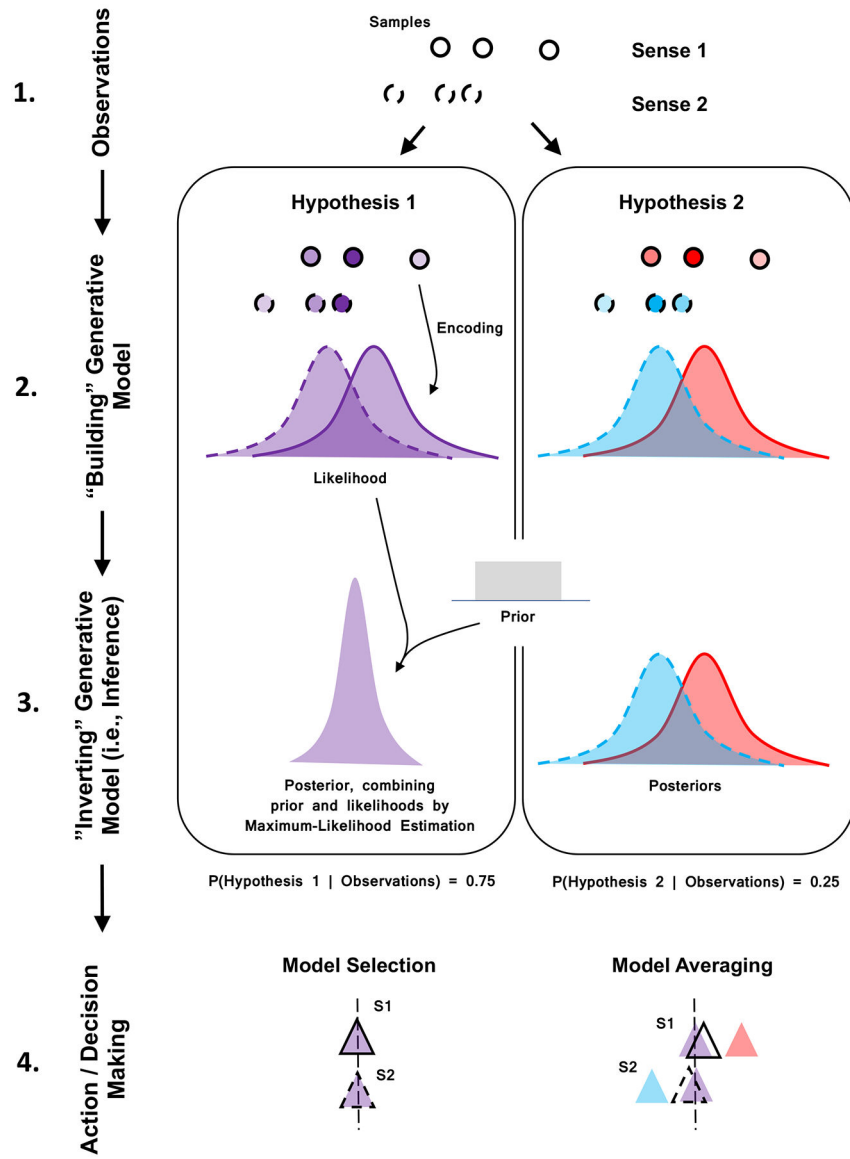


Figure 2. Causal Inference.

Our sensory periphery redundantly samples from the environment (empty circles, step 1). Based on these samples, we build an internal model of the potential causal structure of the world that may have given rise to the observed sensory data (Eq. 5, step 2). In the first hypothesis illustrated here (Hypothesis 1), the two senses index a common object in the environment (purple). As such, the samples that best reflect the state of affairs is the middle sample for sense 1, and the right-most for sense 2 (color-coded, darker = sample falling closer to the mean of the inferred distribution). Since signals from both sense 1 and 2 are taken to come from the same source, we may integrate this information, together with a prior, according to maximum-likelihood estimation, Eq. 2, step 3). Conversely, we may hypothesize that the two senses reflect different objects in the external environment (a red one and a blue one). If this were the case, the central sample, both for sense 1 and 2 (darkest red and blue respectively), is best aligned with the mean of the inferred

distribution (again, stronger hue indicating the sample closest to the mean of its distribution). Under this hypothesis, we would not integrate the different signals (step 3). Lastly, we may combine (or not) world views (i.e., hypotheses) in acting on the external world (step 4). Two potential solutions are illustrated here. In a model selection strategy (left), we would commit to the most likely hypothesis. In this example, we assume hypothesis 1 is most likely, and thus the final estimates correspond to the estimates from this model. In a model averaging strategy (right), observers may weigh estimates according to the relative certainty of the hypothesis. Again, hypothesis 1 is most likely in this example. Thus, the final estimates (empty triangles) will fall somewhere in between the estimates derived from hypothesis 1 (purple) and hypothesis 2 (blue and red), but closer to the former.