



Neural implementations of Bayesian inference

Hansem Sohn¹ and Devika Narain²

Abstract

Bayesian inference has emerged as a general framework that captures how organisms make decisions under uncertainty. Recent experimental findings reveal disparate mechanisms for how the brain generates behaviors predicted by normative Bayesian theories. Here, we identify two broad classes of neural implementations for Bayesian inference: a modular class, where each probabilistic component of Bayesian computation is independently encoded and a transform class, where uncertain measurements are converted to Bayesian estimates through latent processes. Many recent experimental neuroscience findings studying probabilistic inference broadly fall into these classes. We identify potential avenues for synthesis across these two classes and the disparities that, at present, cannot be reconciled. We conclude that to distinguish among implementation hypotheses for Bayesian inference, we require greater engagement among theoretical and experimental neuroscientists in an effort that spans different scales of analysis, circuits, tasks, and species.

Addresses

¹ McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA

² Dept. of Neuroscience, Erasmus University Medical Center, Rotterdam, 3015, CN, the Netherlands

Corresponding author: Narain, Devika (d.narain@erasmusmc.nl)

Current Opinion in Neurobiology 2021, **70**:121–129

This review comes from a themed issue on **Computational Neuroscience**

Edited by **Julijana Gjorgjieva** and **Ila Fiete**

For a complete overview see the [Issue](#) and the [Editorial](#)

Available online 19 October 2021

<https://doi.org/10.1016/j.conb.2021.09.008>

0959-4388/© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Introduction

Behavioral repertoires tend to be rife with feats of precision even though our observations of the environment can be ambiguous and subject to various sources of uncertainty. To mitigate such uncertainty, an organism may exploit several sources of information available to it. Normative theories enable us to better understand this process by quantifying the relative

influence of sources of information. For instance, we can formalize previous experience with relevant variables as a prior distribution and can model uncertainty in our measurements of these variables as a likelihood function (For an in-depth review, see ref [1]). Having defined these sources of information, Bayesian estimation theory [1–5] prescribes how these two sources of information, priors and likelihoods, could be combined to generate a posterior distribution¹ and thereafter enables the derivation of an estimate that minimizes some performance metric (e.g. expected error) over the posterior. Bayesian theoretical frameworks, therefore, provide a formal avenue for studying whether an organism's behavior adheres to rational inference principles as it makes choices under uncertainty [2,5–8].

The value of the Bayesian perspective lies in its ability to confer function on seemingly inaccurate or biased behaviors [9]. For instance, several studies have demonstrated the consistency of time perception with Bayesian models, especially in tasks where humans or monkeys must reproduce a previously observed time interval [10–12] or aim to continue a rhythmic beat [13,14]. When stimulus durations in these tasks are sampled from a prior distribution, responses become biased toward the mean of the distribution. Bayesian models provide us with the insight that such seemingly biased behavior could, in fact, represent an optimal strategy that mitigates uncertainty in measurements by leveraging prior knowledge to minimize overall error. Along these lines, Bayesian theory has helped in the characterization of numerous behaviors in different species in the sensory [4,15–21], motor [22–26], cognitive [9,27], and temporal domains [10–14,28–31]. Notwithstanding skepticism about the suitability of Bayesian models for certain behaviors [32–34], recent developments in modeling have expanded the considerations by which we deem behaviors optimal or suboptimal, for example, complex likelihoods, priors, and policies [35–37], bounded computational power to perform inference [38–41], and limits of learning strategies [42–45].

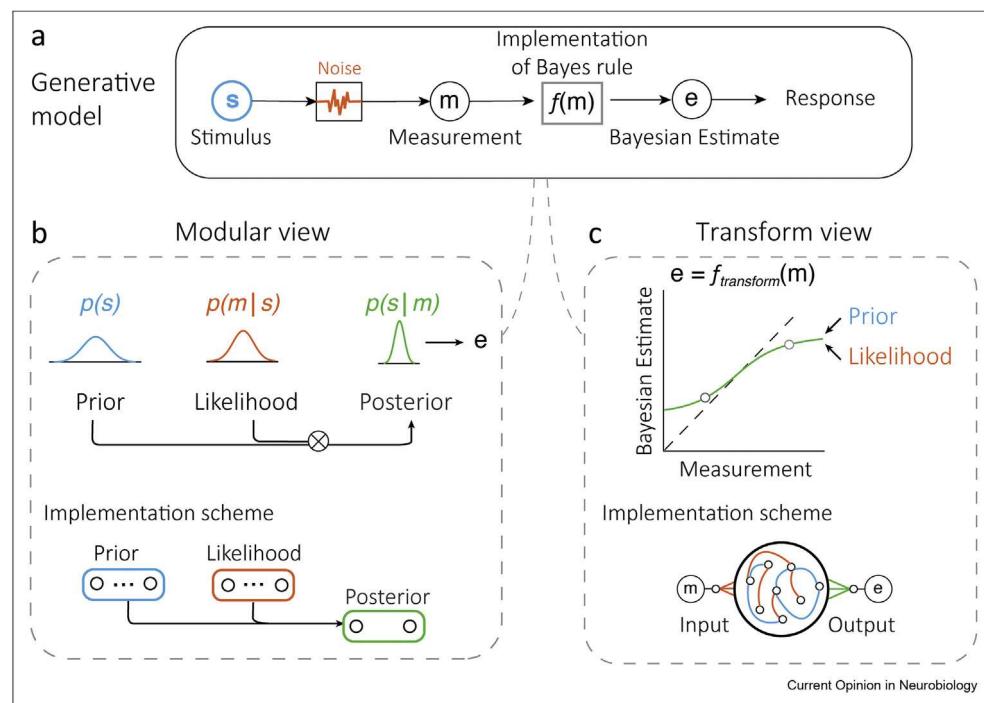
¹ Derived from the latin phrase *a posteriori*, which Immanuel Kant defined as knowledge that comes after empirical evidence has been considered. In the Bayesian case, *a posteriori* is subjective knowledge that comes after both the *a priori* knowledge (or prior) and observations, in the form of likelihoods, have been considered. We sometimes refer to these subjective probabilities as belief distributions.

Concomitant with the development of Bayesian models that describe behavior, several theories emerged to explain implementation mechanisms underlying probabilistic computations in neurobiology, which span synapses [46–48], circuits [13,49], and populations [50–54]. Here, we focus on aspects of these theories that pertain to similarly structured generative processes (Figure 1), which model relationships between the various elements used to carry out Bayesian inference. On the empirical front, recent years have witnessed a spate of findings [14,19,21,26,28,55,56] that investigated the neural basis of behaviors consistent with normative probabilistic theories. Moreover, other experimental studies have recently uncovered neural representations of individual components that could underlie Bayesian computations [56–59]. These findings relate to disparate behaviors in different species, which are likely to engage various brain regions. In other words, not only are we confronted with myriad empirical

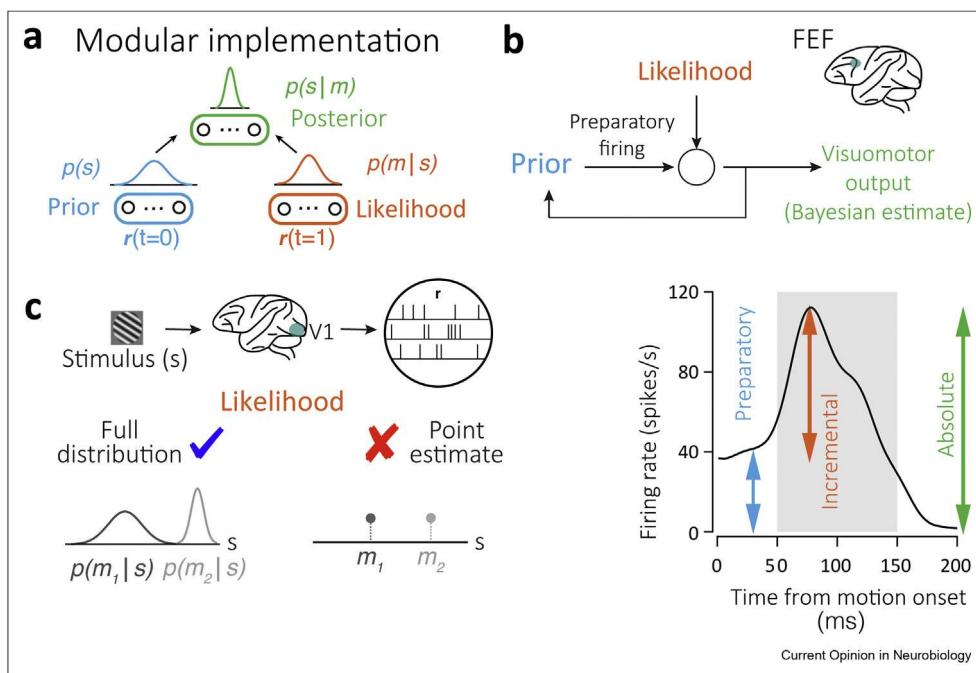
insights that require synthesis, we must also attempt to interpret these under various computational models. To address this many-to-many mapping conundrum, we recommend a conceptual categorization that seeks to simplify the interpretation of empirical findings under existing proposals of neural instantiations of Bayesian inference.

We propose two broad classes under which current implementation theories can be understood. Under the first class, which we refer to as the modular perspective, the likelihood and prior are encoded as independent entities and the mechanism entails a combination of these components to enable full Bayesian inference [1] (Figure 1b). The other case, which we call the transform perspective, specifies how uncertain sensory measurements can be directly mapped into Bayesian estimates via latent processes within which prior distributions are embedded. This process does not mandate encoding of

Figure 1



Two perspectives on neural implementation of Bayesian inference. (a) Generative model. Real-world stimuli, received by the sensory apparatus, are corrupted by various sources of internal and external uncertainty to yield a noisy measurement. In this scheme, the brain estimates the stimulus based on the measurement and prior knowledge of the real-world variable to elicit a response. Neural implementation of this inference process, $f(m)$, can be understood in terms of two possible mechanisms. (b) Modular view. Top. The likelihood (orange), $p(m|s)$, and prior distributions (blue), $p(s)$, are independently encoded and combined to generate a posterior distribution (green), $p(s|m)$. An estimate (e) is derived from the posterior using an appropriate cost function. Bottom: This modularity can be thought of in terms of a feedforward network structure where neural firing rates representing belief distributions are encoded independently and sequentially summed toward an output. For example, the rate at an earlier time, $r(t=0)$, would represent prior knowledge and be combined with the rate after a sensory measurement (likelihood) is made, $r(t=1)$, then subsequent activity could reflect the posterior ($t > 1$). (c) Transform view. Top. The Bayesian estimate (e) is obtained by transforming an uncertain measurement through a function $f_{\text{transform}}(\cdot)$. This function, which represents a continuous estimate of the posterior (green), carries within it a representation of the likelihood (orange) and prior (blue). Bottom: The fundamentals of this computation are captured well by a recurrent network/circuit model. Measurements arrive as inputs while the encoding of prior and likelihood is distributed in circuit connectivities and dynamics. An estimate of the posterior (e) can be obtained through an appropriate readout.

Figure 2

In the modular scheme for Bayesian inference, (a) the prior and likelihood are independently represented and then combined to generate the posterior distribution. We highlight two recent works that represent this view. (b) Recent work (Figure adapted from ref [26]) studying smooth pursuit behaviors in monkeys has shown that preparatory neural activity of frontal eye field neurons (example of firing rate in the panel below) before the presentation of the stimulus reflects the prior distribution and hypothesizes, using a variant of the PPC model, that higher visual cortices provide a signal representing sensory likelihood. Neural activity modulation during the behavior (orange arrow within the shaded region in the panel below) was consistent with the Bayesian estimate (not shown here. For details, consult the study reported by Darlington et al. [26]). This scheme reflects independent representations of various components of the Bayesian computation and appears to be consistent with the modular view. (c) Other recent work (Figure adapted from ref [56]) tests the assumptions underlying PPC, examining whether a monkey utilizes information of the full likelihood (bottom-left) decoded from the primary visual cortex in a trial-by-trial manner to make a decision. The full likelihood model better predicted trial-by-trial behaviors of animals than the alternative model with fixed uncertainty (bottom-right).

probabilistic distributions on each trial, which may not be necessary for certain tasks and can be resource-intensive for circuit implementation [60]. This implementation-based dichotomy might aid in interpreting existing insights while potentially refining empirical hypotheses of Bayesian theory for different circuits, species, and tasks.

Bayesian modular perspective

In this class of neural implementation, probabilistic computations are carried out using independent representations of likelihood, prior, and posterior distributions, followed by the generation of an estimate² (Figure 2a). Two prominent frameworks address pathways toward such an implementation. The first, probabilistic population coding (PPC) [61–63] formalizes how optimal inference ensues from population coding of belief distributions, which when linearly combined provide a close approximation to the

implementation of the Bayes rule.³ Important features of PPC that are consistent with the modular view are independent representations of likelihood functions [56] and prior knowledge⁴ in population codes, thus supporting flexible updates of trial-by-trial uncertainty.

Three recent studies in nonhuman primates reported findings consistent with the predictions of PPC in different circuits and behaviors, providing support for a modular implementation. The first of these studied smooth pursuit behaviors, where monkeys followed a moving target, whose speed was drawn from different prior distributions and whose reliability was modulated by changing contrasts [26]. The monkeys' pursuit behavior was consistent with the predictions of Bayesian inference, and the authors discovered that preparatory activity in the frontal eye fields encoded information about the prior distribution (Figure 2b). Moreover, the neural activity in the same area, during the pursuit behavior, encoded an estimate of target speed that was consistent with a Bayesian estimator. This echoes the

² Through the choice of an appropriate cost function. For more details see the study reported by Ma and Jazayeri, Jaynes, Kording and Wolpert, Knill and Richards, Ma, Berger, and Kersten Refs [1–7].

³ Under the assumption of Poisson-like neural variability. For more, see ref [38].

⁴ The logarithm of the prior, in many cases.

tenets of the modular view by demonstrating independent observability of prior knowledge before its combination with sensory measurement. Furthermore, authors showed how a linear PPC model supports their findings (**Figure 2b**). Similar findings pertaining to prior knowledge being reflected in preparatory activity have been reported in different species, areas, and tasks [59,64], which raises possibilities about the generality of such modular computations. The second study [55] consistent with predictions of PPC examined activity of neurons in the lateral intraparietal (LIP) cortex of monkeys performing a multisensory decision-making task.⁵ Here, monkeys were trained to judge perceived heading based on direction of vestibular (inertial cues) and visual (optic flow) stimuli. Lateral intraparietal neurons integrated visual stimulus velocity and vestibular acceleration over time in a manner consistent with linear PPC models, which as shown previously, implements a Bayes optimal solution [63,65]. These findings echo similar results obtained in earlier work [66] and argue for independent representations of likelihoods in different areas.

The third recent study [56] provides empirical evidence about the independent probabilistic encoding of likelihoods (**Figure 2c**). Monkeys were trained to perform a visual orientation discrimination task that required trial-by-trial use of stimulus uncertainty, thereby providing an incentive to retain a full representation of the likelihood function. Using a Bayesian decision model and machine learning techniques, the authors showed that the full likelihood function can be decoded from population activities in the primary visual cortex. Furthermore, the full likelihood model better predicted trial-by-trial behaviors of animals than the alternative model with fixed uncertainty. This work demonstrates how representations of entire likelihood functions can be encoded neural populations, thereby providing support for the PPC framework and the modular view.

The second neural implementation framework that has the potential to support modular probabilistic representations is sampling codes [51,67,68]. In this scheme, neural activity at a given point is assumed to be a sample of a probability distribution and neural variability across time reflects the uncertainty in the stimulus [69]. These predictions are consistent with the modular view in certain ways. For instance, neural responses in the ferret primary visual cortex confirmed model predictions of how spontaneous activity of neurons represented samples from a prior distribution, whereas responses that occurred after sensory stimuli (observation) reflected samples from a posterior [21]. Another important feature of the sampling-based model is that neural

variability is a direct consequence of probabilistic computation, not a nuisance feature in neural circuits, and therefore needs to be flexibly controlled depending on the stimulus uncertainty. Recent work [68] demonstrated that a computational model based on an excitatory-inhibitory motif could achieve such flexible control of neural variability, providing mechanistic accounts for several properties of cortical dynamics including stimulus-driven reduction of neural variability. This work provides a pathway to how biologically realistic circuits could perform Bayesian inference using sampling-based codes.

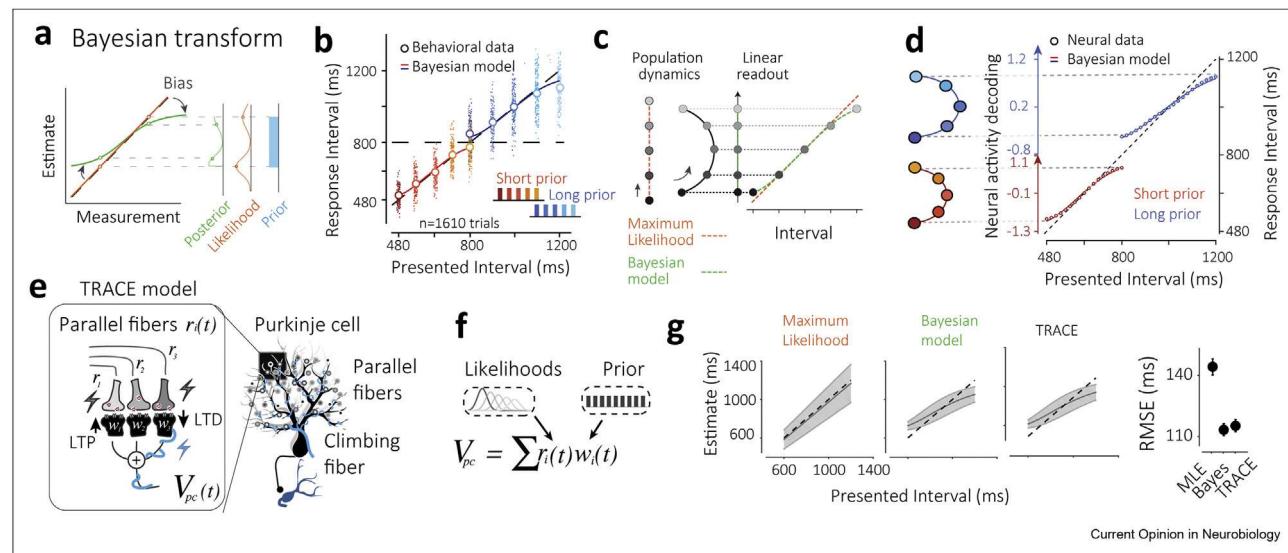
To summarize, modular implementations of Bayesian inference support independent representations of prior distributions and likelihood functions. Such implementations can sustain flexible computations with multiple representations and can support rapid sequential inference [70,71]. For several tasks, however, encoding full representations of priors and likelihoods can prove to be resource-intensive and inefficient [60,72]. Precluding the need for rapid trial-wise flexibility, one efficient solution would entail the implicit encoding of priors such that any observation may be nonlinearly transformed into a Bayesian estimate [10,60,73]. We explore this class of implementations in the next section.

Bayesian transform perspective

The transform view makes the assumption that neural mechanisms underlying Bayesian computations carry out an input–output transformation that encodes relevant belief distributions from which a Bayesian estimate can be derived [10,60]. Particularly compelling theories that support the Bayesian transform view are recent instantiations of efficient coding theory with heterogeneous neural populations that embed representations of prior knowledge within them [53]. This theory has found support in empirical evidence from human visual behavior [17], auditory behaviors in owls [20], primate temporal perception [28], and associative learning [13]. At a conceptual level, this theory aligns with the transform view because it predicts that latent encoding of priors in neural substrates facilitates a transformation of incoming observations into Bayesian estimates, which distinguishes itself from the independent encoding of full belief distributions and computations carried out under the modular view.

The transform view (**Figure 3a**) can be best understood from two vantage points, the level of collective dynamics of neural populations [28] or at the level of prior knowledge embedded in local circuits [13,49,53,74]. Beginning from the population viewpoint, recent advances in neuroscience have enabled the recording of hundreds of neurons as animals perform behavioral tasks [75]. Recent studies have provided a glimpse into how neural

⁵ This study does not directly test Bayesian aspects of learning and is an example of multisensory integration. The framework used in this study (PPC) proposes how different sources of information can be combined in accordance with their reliability and although not highlighted in this specific study possesses the capacity to explain implementations of Bayesian integration [63].

Figure 3

The Bayesian transform scheme. (a) An uncertain sensory measurement can be transformed into a Bayesian estimate by means of a nonlinear deterministic mapping $f(\cdot)$. (b) In a recent study [28], monkeys were trained to reproduce observed time intervals, which were sampled from two prior distributions, short (warm colors) and long (cool colors). Because response variability increases with duration, Bayesian models predict larger biases for longer priors, which were observed in behavioral responses. (c) One hypothesized mechanism by which such biased responses could be generated is the warping of neural population dynamics during interval measurements into semicircular geometries. Linear projections of these geometries onto their own diameters yield states that become compressed toward the mean of the prior. If these readouts are used to generate behavioral responses, these responses will also be biased toward the mean of the prior. (d) Neural population dynamics of the dorsomedial frontal cortex (DMFC) exhibited different curved geometries for each prior, which when decoded (circles) matched the biases predicted by Bayesian models (lines) [28]. Because representations of likelihoods and priors are entangled, this form of implementation is consistent with the transform view. (e) A biophysical model for the Bayesian transform in cerebellar circuits [13]. A biophysical model (TRACE) that provides a basis for how synaptic plasticity dynamics in the form of long term depression (LTD) and long term potentiation (LTP) occurring at the confluence of parallel/climbing fiber inputs at the Purkinje cells of the cerebellar cortex can encode likelihood functions $r_i(t)$ and priors $p(s(t))$. (f) This implementation strategy is consistent with the transform view because measurements arriving through the parallel fibers are transformed by the prior knowledge embedded in synaptic weights $w_i(t)$, giving rise to the Bayesian estimate. (g) The output of the TRACE model is consistent with Bayesian models and deviates significantly from non-Bayesian models such as maximum likelihood estimation (MLE) as measured by the overall error of timing responses (root-mean-squared-error; RMSE). Error bars represent standard deviation.

populations perform task-relevant computations within the activity state space spanned by a large set of neurons [76–80]. An important insight gained was that although the full neural state space occupies a high-dimensional space, projections of neural activity onto a low-dimensional, behaviorally relevant subspace can provide substrates for input–output transformations [81].

To illustrate this in light of the Bayesian transform view (Figure 3a), we examine a recent study where monkeys observed time intervals and then waited for a matching duration before responding [28]. When stimulus intervals across trials were sampled from different prior distributions, monkeys' waiting times were found to be biased toward the mean of the prior distribution (Figure 3b). Because timing responses become more variable with longer durations [82], Bayesian models predict greater uncertainty in likelihoods associated with long durations, and consequently, greater bias toward the mean [10]. This prediction was tested by using two prior distributions with short and long durations. During the task, neurons in the dorsomedial

frontal cortex modulated differently during the range of each prior, which when viewed collectively in neural state space, manifested as semicircular population trajectories with different curvatures for each prior (Figure 3c and d). The study reports that when neural states were projected linearly onto the diameters of these semicircular trajectories, the resulting states became compressed toward the mean of the prior (Figure 3c). They predicted, the greater the compression in projected neural states, the larger the expected bias toward the mean in the behavior. These predictions were borne out in comparisons with Bayesian models and the monkeys' behavior (Figure 3d). Recurrent neural network models, which recapitulated the animals' behavior and dorsomedial frontal cortex population dynamics, helped further test this hypothesis by enabling direct perturbation of the subspace occupied by the semicircular population trajectories, which when compressed elicited larger behavioral bias [28].

The modus operandi uncovered in these neural dynamics represents a deterministic transformation of an

uncertain time measurement into a Bayesian estimate, which represents the mean of the posterior. It, therefore, supports a mechanism that best reflects the transform class. Although this study did not directly test the assumption, the mechanism elucidated in this work shares many similarities with the theoretical framework called distributed distributional code (DDC), which does not require full probabilistic representations but uses statistical moments of the prior distribution [52,83]. Furthermore, DDC proposes a linear readout to estimate the mean of the posterior, which is consistent both with the population mechanism and type of Bayesian estimator described in the aforementioned study [28]. For these reasons, we propose that DDC is well-aligned with the transform perspective. Further work is needed, however, to develop a quantifiable approach to interpret different model frameworks and experimental findings under the modular or transform classes.

On the issue of flexibility underlying Bayesian transforms, recent experimental evidence elucidates how transform class mappings can be sensitive to changes in the relative reliability of likelihoods and priors [14]. In this study, monkeys were trained to make two time measurements before reproducing an interval, whereas time intervals across trials were sampled from a prior distribution. The task employed, therefore, jointly tests within-trial decrease of measurement uncertainty and across-trial influence of prior knowledge. In accordance with predictions of Bayesian estimation theory, neural states decoded from responses in the DMFC were biased during the first time measurement and decreased their bias during the second measurement, consistent with normative expectations of a corresponding decrease in uncertainty. This work demonstrates that parameters underlying such Bayesian transforms can flexibly adapt to changes in uncertainty. Other recent work has also explored neural correlates of flexible sequential Bayesian computations in rodents performing a spatial navigation task under sensory uncertainty [19].

The second vantage point through which the transform view can be considered is that of synaptic connectivity. Recent attempts have aimed to uncover mechanistic origins of Bayesian inference by studying how prior knowledge can be encoded in synaptic connectivity [13,49]. One of these studies provides the basis for learning and encoding of prior distributions over time intervals across the synaptic weights of the principal cerebellar neurons, Purkinje cells. The authors propose a computational mechanism by which cerebellar microcircuits transform incoming temporal measurements into states consistent with Bayesian estimates (Figure 3c). Conceptually, this work is consistent with a recent efficient coding hypothesis [53] that describes how neurons tune their activity to reflect prior

knowledge in order to maximize information content. It is worth noting that this circuit model [13] and theoretical framework [53] were developed to explain Bayesian inference for entirely different systems (cerebellum vs visual cortex) and tasks (timing vs visual behaviors) and yet share many common mechanistic features. Within the transform perspective, these unexpected similarities highlight the potential for generalizability of mechanisms for Bayesian inference across different tasks and regions.

In summary, the transform view emphasizes that belief distributions need not be independently encoded in neural circuits to give rise to Bayesian estimates. When trial-wise flexibility is not a prerequisite, neural circuits can efficiently render measurements into Bayesian estimates through latent processes that encode prior knowledge.

Synthesis and divergence

Having examined various recent works on neural implementations of behaviors consistent with Bayesian inference, we ask whether the modular or transform views are entirely distinct or whether they can be synthesized under a broader overarching idea. A point of divergence arrives when we turn our attention to the site of encoding of prior knowledge in the neural circuit. For instance, regarding the study by Darlington et al. [26], we hope future work can elucidate whether prior knowledge is embedded in the synaptic weights of the frontal cortex (transform view) or encoded in a different area from where the preparatory signal is inherited (modular view). In other words, is the preparatory activity reflecting prior knowledge also the key driver behind Bayesian computations [21,59,64], or is it an epiphenomenon while the latent transform within the synapses provides the key mechanism [13,49,74,84]? Resolving this distinction [85,86] may require technological advances in our ability to directly observe the influence of synaptic weights on a large scale and measure interareal inputs. At the population level, we may also hope for the means to dissociate latent population dynamics into their individual biophysical and cell-type interactions to bridge the two vantage points within the Bayesian transform perspective by leveraging recent advances in theory [87,88]. On the empirical side, it is interesting to note that while the toolkit for circuit interrogation to test these distinctions is currently available in rodent animal models, almost all the studies mentioned here were performed in nonhuman primates, where our ability to study circuits, cell types, and anatomy is significantly curtailed. Achieving these kinds of reductionist unifications will not only require us to look across species, it will also require theory and experiments to bridge insights across cellular, circuit, and population scales.

One point of synthesis arises naturally when considering the possibility that the brain may adopt both modular and transform strategies depending on task demand and computational power of the local task-relevant network. If optimal behavior in a task requires constant monitoring of uncertainty in variables [70,71], it makes sense to independently encode such variables in circuits [56], as the modular view purports. On the other hand, if stable learning of environmental statistics is sufficient to optimally perform the task, latent encoding of environmental statistics into the circuit may provide the most efficient solution, as captured by the transform view [53]. An important metric that allows us to gauge task demands and computational power is the expected dimensionality incurred by a task of given complexity in the neural state space [76]. For instance, a neural network that performs a Bayesian transformation for one task with low demands for flexibility will potentially occupy a lower-dimensional space than a network that performs modular Bayesian computations with independent representations of likelihood functions, prior, and posterior distributions. *In silico* investigations in recurrent neural networks may allow us to determine whether the same network can be trained to implement higher or lower dimensionality solutions that enable the switch between modular and transform instantiations along a spectrum. One important question that arises is how the brain might arbitrate between these two approaches across different brain regions given the discovery of high dimensional encoding of stimuli in sensory cortices [89,90] versus low dimensional encoding of tasks associated with frontal and motor cortices [77,78,91–99].

In conclusion, neuroscientists are increasingly embracing the need to rigorously quantify behaviors in pursuit of their underlying neural mechanisms [100]. Modeling behaviors with normative theories can further aid this effort by providing simultaneous constraints over both behavior and neural mechanisms, potentially increasing our ability to discover general principles of neural function. Future work in this domain will benefit from continuation of the recent tradition of interaction among theorists and experimentalists working together at the population [26,28,55,56], circuit, and cellular levels [13,48,49] across different animal models [20,21].

Author contributions

Hanse Sohn and Devika Narain: Conceptualization, Writing, Editing and Review.

Hanse Sohn and Devika Narain: Visualization.

Conflict of interest statement

Nothing declared.

Acknowledgements

The authors thank Joshua Dudman and Matthew Botvinick for inspiring discussions. The authors are grateful to Mehrdad Jazayeri and Seth Egger for their comments on the article. H.S. is supported by the Center for

Sensorimotor Neural Engineering and the NARSAD young investigator grant (2020) by the Brain and Behavior Research Foundation. D.N. is supported by the Vidi grant (21-15016) from the Netherlands Organization for Scientific Research and by the EU Marie Skłodowska Curie reintegration grant (PredOpt 796577).

References

Papers of particular interest, published within the period of review, have been highlighted as:

- * of special interest
- ** of outstanding interest

1. Ma WJ, Jazayeri M: **Neural coding of uncertainty and probability.** *Annu Rev Neurosci* 2014, **37**:205–220.
2. Jaynes ET: **Probability theory: the logic of science.** Cambridge University Press; 2003.
3. Kording KP, Wolpert DM: **Bayesian decision theory in sensorimotor control.** *Trends Cognit Sci* 2006, **10**:319–326.
4. Knill DC, Richards W: **Perception as Bayesian inference.** Cambridge University Press; 1996.
5. Ma WJ: **Bayesian decision models: a primer.** *Neuron* 2019, **104**:164–175.
6. Berger JO: **Statistical decision theory and Bayesian analysis.** Springer Science & Business Media; 2013.
7. Kersten D, Mamassian P, Yuille A: **Object perception as Bayesian inference.** *Annu Rev Psychol* 2004, **55**:271–304.
8. McNamee D, Wolpert DM: **Internal models in biological control.** *Annu Rev Control Robot Auton Syst* 2019, **2**:339–364.
9. Griffiths TL, Kemp C, Tenenbaum JB: **Bayesian models of cognition.** *Camb Handb Comput Psychol* 2001, <https://doi.org/10.1017/cbo9780511816772.006>.
10. Jazayeri M, Shadlen MN: **Temporal context calibrates interval timing.** *Nat Neurosci* 2010, **13**:1020–1026.
11. Miyazaki M, Nozaki D, Nakajima Y: **Testing Bayesian models of human coincidence timing.** *J Neurophysiol* 2005, **94**:395–399.
12. Acerbi L, Wolpert DM, Vijayakumar S: **Internal representations of temporal statistics and feedback calibrate motor-sensory interval timing.** *PLoS Comput Biol* 2012, **8**, e1002771.
13. Narain D, Remington ED, Zeeuw CID, Jazayeri M: **A cerebellar mechanism for learning prior distributions of time intervals.** *Nat Commun* 2018, **9**:469.
- This work studied how the cerebellar circuitry encodes likelihoods for time intervals, which are transformed through a network of synaptic weights that encode prior knowledge. This model provides a mechanism for learning of prior knowledge in realistic circuits and of the generation of trial-by-trial Bayesian estimates for time intervals. Authors also showed that timing behaviors known to be encoded in the modeled cerebellar circuitry are consistent with Bayesian inference.
14. Egger SW, Remington ED, Chang C-J, Jazayeri M: **Internal models of sensorimotor integration regulate cortical dynamics.** *Nat Neurosci* 2019, **22**:1871–1882.
- This work studied a task that simultaneously gauged effects of within-trial decrease of sensory uncertainty and across-trial influence of prior knowledge, in a novel timing task in monkeys that exhibited Bayes-consistent behavior. Authors provide an account for flexible updating of prediction errors through nonlinear transformations.
15. Mamassian P, Landy MS: **Interaction of visual prior constraints.** *Vis Res* 2001, **41**:2653–2668.
16. Stocker AA, Simoncelli EP: **Noise characteristics and prior expectations in human visual speed perception.** *Nat Neurosci* 2006, **9**:578–585.
17. Girshick AR, Landy MS, Simoncelli EP: **Cardinal rules: visual orientation perception reflects knowledge of environmental statistics.** *Nat Neurosci* 2011, **14**:926–932.
18. Knill DC: **Learning Bayesian priors for depth perception.** *J Vis* 2007, **7**:13.

19. Funamizu A, Kuhn B, Doya K: **Neural substrate of dynamic Bayesian inference in the cerebral cortex.** *Nat Neurosci* 2016, **19**:1682–1689.
20. Fischer BJ, Peña JL: **Owl's behavior and neural representation predicted by Bayesian inference.** *Nat Neurosci* 2011, **14**: 1061–1066.
21. Berkes P, Orbán G, Lengyel M, Fiser J: **Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment.** *Science* 2011, **331**:83–87.
22. Kording KP, Ku S-P, Wolpert DM: **Bayesian integration in force estimation.** *J Neurophysiol* 2004, **92**:3161–3165.
23. Kording KP, Wolpert DM: **Bayesian integration in sensorimotor learning.** *Nature* 2004, **427**:244–247.
24. Berniker M, Voss M, Kording K: **Learning priors for Bayesian computations in the nervous system.** *PLoS One* 2010, **5**.
25. Darlington TR, Tokiyama S, Lisberger SG: **Control of the strength of visual-motor transmission as the mechanism of rapid adaptation of priors for Bayesian inference in smooth pursuit eye movements.** *J Neurophysiol* 2017, **118**:1173–1189.
26. Darlington TR, Beck JM, Lisberger SG: **Neural implementation of Bayesian inference in a sensorimotor behavior.** *Nat Neurosci* 2018, **21**:1442–1451.
- This work studied smooth pursuit behavior that was consistent with Bayesian inference in monkeys. Authors showed how preparatory activity in the frontal cortex reflects the prior distribution whereas activity during the pursuit behavior reflected a Bayesian estimate. Authors proposed a model to explain these findings that is a variation of Probabilistic Population Coding frameworks.
27. Oaksford M, Chater N: **Bayesian rationality: the probabilistic approach to human reasoning.** Oxford University Press; 2007.
28. Sohn H, Narain D, Meirhaeghe N, Jazayeri M: **Bayesian computation through cortical latent dynamics.** *Neuron* 2019, **103**:934–947.e5.
- This work studied the neural population mechanisms underlying Bayes-consistent timing behavior in monkeys and used recurrent neural networks to corroborate the population dynamics hypothesis. Authors found that prior knowledge manifested as a geometrical property of the neural population trajectory, which when linearly readout exhibited states consistent with Bayesian estimates.
29. Egger SW, Le NM, Jazayeri M: **A neural circuit model for human sensorimotor timing.** *Nat Commun* 2020, **11**:3933.
30. Li Y, Dudman JT: **Mice infer probabilistic models for timing.** *Proc Natl Acad Sci U S A* 2013, **110**:17154–17159.
31. Shi Z, Church RM, Meck WH: **Bayesian optimization of time perception.** *Trends Cognit Sci* 2013, **17**:556–564.
32. Rahnev D, Denison RN: **Suboptimality in perceptual decision making.** *Behav Brain Sci* 2018, **41**:1–66.
33. Chambers C, Kording KP: **Although optimal models are useful, optimality claims are not that common.** *Behav Brain Sci* 2018, **41**.
34. Jones M, Love BC: **Bayesian Fundamentalism or Enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition.** *Behav Brain Sci* 2011, **34**:169–188.
35. Acerbi L, Vijayakumar S, Wolpert DM: **On the origins of suboptimality in human probabilistic inference.** *PLoS Comput Biol* 2014, **10**, e1003661.
36. Chambers C, Fernandes H, Kording KP: **Policies or knowledge: priors differ between a perceptual and sensorimotor task.** *J Neurophysiol* 2019, **121**:2267–2275.
37. Sohn H, Jazayeri M: **Validating model-based Bayesian integration using prior-cost metamers.** *Proc Natl Acad Sci Unit States Am* 2021, **118**, e2021531118.
38. Genewein T, Leibfried F, Grau-Moya J, Braun DA: **Bounded rationality, abstraction, and hierarchical decision-making: an information-theoretic optimality principle.** *Front Robot AI* 2015, **2**.
39. Ortega PA, Braun DA: **Thermodynamics as a theory of decision-making with information-processing costs.** *Proc Math Phys Eng Sci* 2013, **469**:20120683.
40. Bhui R, Lai L, Gershman SJ: **Resource-rational decision making.** *Curr Opin Behav Sci* 2021, **41**:15–21.
41. Lieder F, Griffiths TL: **Resource-rational analysis: understanding human cognition as the optimal use of limited computational resources.** *Behav Brain Sci* 2019, **43**:e1.
42. Laquaitaine S, Gardner JL: **A switching observer for human perceptual estimation.** *Neuron* 2018, **97**:462–474. e6.
43. Narain D, Mamassian P, van Beers RJ, Smeets JBJ, Brenner E: **How the statistics of sequential presentation influence the learning of structure.** *PLoS One* 2013, **8**, e62276.
44. Genewein T, Braun DA: **Occam's Razor in sensorimotor learning.** *Proc Biol Sci* 2014, **281**:20132952.
45. Narain D, Smeets JBJ, Mamassian P, Brenner E, van Beers RJ: **Structure learning and the Occam's razor principle: a new view of human function acquisition.** *Front Comput Neurosci* 2014, **8**:121.
46. Pfister J-P, Dayan P, Lengyel M: **Synapses with short-term plasticity are optimal estimators of presynaptic membrane potentials.** *Nat Neurosci* 2010, **13**:1271–1275.
47. Soltani A, Wang X-J: **Synaptic computation underlying probabilistic inference.** *Nat Neurosci* 2010, **13**:112–119.
48. Aitchison L, Jegminat J, Menendez JA, Pfister J-P, Pouget A, Latham PE: **Synaptic plasticity as Bayesian inference.** *Nat Neurosci* 2021, <https://doi.org/10.1038/s41593-021-00809-5>.
49. Hiratani N, Latham PE: **Rapid Bayesian learning in the mammalian olfactory system.** *Nat Commun* 2020, **11**:3845.
- This work formulates local synaptic plasticity in mammalian olfactory circuits as a Bayesian optimization process to account for rapid unsupervised odor-reward associated learning.
50. Pouget A, Beck JM, Ma WJ, Latham PE: **Probabilistic brains: knowns and unknowns.** *Nat Neurosci* 2013, **16**:1170–1178.
51. Fiser J, Berkes P, Orbán G, Lengyel M: **Statistically optimal perception and learning: from behavior to neural representations.** *Trends Cognit Sci* 2010, **14**:119–130.
52. Sahani M, Dayan P: **Doubly distributional population codes: simultaneous representation of uncertainty and multiplicity.** *Neural Comput* 2003, **15**:2255–2279.
53. Ganguli D, Simoncelli EP: **Efficient sensory encoding and Bayesian inference with heterogeneous neural populations.** *Neural Comput* 2014, **26**:2103–2134.
- This work provided a theoretical account for how prior knowledge can warp the tuning curves of a population of neurons consistent with efficient coding theory. While this work was proposed for the visual cortex, it has provided the basis for recent circuit models in other brain regions underlying Bayesian inference.
54. Doya K: **Bayesian brain: probabilistic approaches to neural coding.** MIT Press; 2007.
55. Hou H, Zheng Q, Zhao Y, Pouget A, Gu Y: **Neural correlates of optimal multisensory decision making under time-varying reliabilities with an invariant linear probabilistic population code.** *Neuron* 2019, **104**:1010–1021. e10.
- This work studied cue combination of vestibular and visual cues in the lateral intraparietal cortex of monkeys. The authors showed that the features of the uncovered population are consistent with predictions of the probabilistic population coding framework.
56. Walker EY, Cotton RJ, Ma WJ, Tolias AS: **A neural basis of probabilistic computation in visual cortex.** *Nat Neurosci* 2020, **23**:122–129.
- Authors tested a prominent assumption of probabilistic population coding, whether full likelihood functions are represented in the brain. The work studied the primary visual cortex of monkeys as they performed a visual orientation discrimination task that required trial-by-trial use of stimulus uncertainty. They showed that the visual cortex data was most consistent with a full representation of the likelihood function.
57. Basso MA, Wurtz RH: **Modulation of neuronal activity by target uncertainty.** *Nature* 1997, **389**:66–69.
58. Glaser JI, Perich MG, Ramkumar P, Miller LE, Kording KP: **Population coding of conditional probability distributions in dorsal premotor cortex.** *Nat Commun* 2018, **9**.

59. Akrami A, Kopec CD, Diamond ME, Brody CD: **Posterior parietal cortex represents sensory history and mediates its effects on behaviour.** *Nature* 2018, **554**:368–372.
60. Simoncelli EP: **Optimal estimation in sensory systems.** *Cognit Neurosci* 2009, **IV**.
61. Sanger TD: **Probability density estimation for the interpretation of neural population codes.** *J Neurophysiol* 1996, **76**: 2790–2793.
62. Pouget A, Dayan P, Zemel RS: **Inference and computation with population codes.** *Annu Rev Neurosci* 2003, **26**:381–410.
63. Ma WJ, Beck JM, Latham PE, Pouget A: **Bayesian inference with probabilistic population codes.** *Nat Neurosci* 2006, **9**: 1432–1438.
64. Rao V, DeAngelis GC, Snyder LH: **Neural correlates of prior expectations of motion in the lateral intraparietal and middle temporal areas.** *J Neurosci* 2012, **32**:10063–10074.
65. Beck JM, Ma WJ, Kiani R, Hanks T, Churchland AK, Roitman J, Shadlen MN, Latham PE, Pouget A: **Probabilistic population codes for Bayesian decision making.** *Neuron* 2008, **60**:1142–1152.
66. Fetsch CR, Pouget A, DeAngelis GC, Angelaki DE: **Neural correlates of reliability-based cue weighting during multisensory integration.** *Nat Neurosci* 2011, **15**:146–154.
67. Hoyer PO, Hyvärinen A: **Interpreting neural response variability as Monte Carlo sampling of the posterior.** In *Advances in neural information processing systems 15*, Edited by Becker S, Thrun S, Obermayer K, MIT Press; 2003:293–300.
68. Echeveste R, Aitchison L, Hennequin G, Lengyel M: **Cortical-like dynamics in recurrent circuits optimized for sampling-based probabilistic inference.** *Nat Neurosci* 2020, **23**:1138–1149.
- Recent work that proposed a multi-scale model for how Bayesian inference behaviors can be implemented using sampling codes in sensory cortices. This work provides a pathway towards a mechanistic account that seeks to bridge implementation hypotheses between the theoretical sampling-based frameworks and realistic neural circuits.
69. Orbán G, Berkes P, Fiser J, Lengyel M: **Neural variability and sampling-based probabilistic representations in the visual cortex.** *Neuron* 2016, **92**:530–543.
70. Kanitscheider I, Fiete I: **Training recurrent networks to generate hypotheses about how the brain solves hard navigation problems.** *arXiv:1609.09059 [q-bio.NC]*; 2016.
71. Cheung A, Ball D, Milford M, Wyeth G, Wiles J: **Maintaining a cognitive map in darkness: the need to fuse boundary knowledge with path integration.** *PLoS Comput Biol* 2012, **8**, e1002651.
72. Lennie P: **The cost of cortical computation.** *Curr Biol* 2003, **13**: 493–497.
73. Raphan M, Simoncelli EP: **Least squares estimation without priors or supervision.** *Neural Comput* 2011, **23**:374–420.
74. Nessler B, Pfeiffer M, Buesing L, Maass W: **Bayesian computation emerges in generic cortical microcircuits through spike-timing-dependent plasticity.** *PLoS Comput Biol* 2013, **9**, e1003037.
75. Jun JJ, Steinmetz NA, Siegle JH, Denman DJ, Bauza M, Barbarits B, Lee AK, Anastassiou CA, Andrei A, Aydön C, et al.: **Fully integrated silicon probes for high-density recording of neural activity.** *Nature* 2017, **551**:232–236.
76. Gao P, Ganguli S: **On simplicity and complexity in the brave new world of large-scale neuroscience.** *Curr Opin Neurobiol* 2015, **32**:148–155.
77. Churchland MM, Cunningham JP, Kaufman MT, Foster JD, Nuyujukian P, Ryu SI, Shenoy KV: **Neural population dynamics during reaching.** *Nature* 2012, **487**:51–56.
78. Mante V, Sussillo D, Shenoy KV, Newsome WT: **Context-dependent computation by recurrent dynamics in prefrontal cortex.** *Nature* 2013, **503**:78–84.
79. Kobak D, Brendel W, Constantinidis C, Feierstein CE, Kepcs A, Mainen ZF, Qi X-L, Romo R, Uchida N, Machens CK: **Demixed principal component analysis of neural population data.** *Elife* 2016, **5**.
80. Cunningham JP, Yu BM: **Dimensionality reduction for large-scale neural recordings.** *Nat Neurosci* 2014, **17**:1500–1509.
81. Ganguli S, Sompolinsky H: **Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis.** *Annu Rev Neurosci* 2012, **35**:485–508.
82. Gibbon J: **Scalar expectancy theory and Weber's law in animal timing.** *Psychol Rev* 1977, **84**:279–325.
83. Vertes E, Sahani M: **Flexible and accurate inference and learning for deep generative models.** *arXiv [statML]*; 2018. This work is a recent instantiation of the framework Distributed Distributional Codes (DDC), which enable flexible inference of complex probabilistic generative models. This formulation enables the expectation of the posterior, to be readout with linear weights from an appropriately designed basis set.
84. Wu S, Amari S-I: **Neural implementation of Bayesian inference in population codes.** *Adv Neural Inf Process Syst* 2002, **14**: 317–323.
85. Churchland PS, Sejnowski TJ: *The computational brain*. MIT Press; 2016.
86. Sohn H, Meirhaeghe N, Rajalingham R, Jazayeri M: **A network perspective on sensorimotor learning.** *Trends Neurosci* 2020, <https://doi.org/10.1016/j.tins.2020.11.007>.
87. Mastrogiovanni F, Ostoicic S: **Linking connectivity, dynamics, and computations in low-rank recurrent neural networks.** *Neuron* 2018, **99**:609–623. e29.
88. Dubreuil A, Valente A, Beiran M, Mastrogiovanni F, Ostoicic S: **The role of population structure in computations through neural dynamics.** 2021, <https://doi.org/10.1101/2020.07.03.185942> [date unknown].
89. Stringer C, Pachitariu M, Steinmetz N, Carandini M, Harris KD: **High-dimensional geometry of population responses in visual cortex.** *Nature* 2019, **571**:361–365.
90. Seely JS, Kaufman MT, Ryu SI, Shenoy KV, Cunningham JP, Churchland MM: **Tensor analysis reveals distinct population structure that parallels the different computational roles of areas M1 and V1.** *PLoS Comput Biol* 2016, **12**, e1005164.
91. Kaufman MT, Churchland MM, Ryu SI, Shenoy KV: **Cortical activity in the null space: permitting preparation without movement.** *Nat Neurosci* 2014, **17**:440–448.
92. Mazor O, Laurent G: **Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons.** *Neuron* 2005, **48**:661–673.
93. Yu BM, Cunningham JP, Santhanam G, Ryu SI, Shenoy KV, Sahani M: **Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity.** *J Neurophysiol* 2009, **102**:614–635.
94. Wang J, Narain D, Hosseini EA, Jazayeri M: **Flexible timing by temporal scaling of cortical responses.** *Nat Neurosci* 2018, **21**: 102–110.
95. Remington ED, Narain D, Hosseini EA, Jazayeri M: **Flexible sensorimotor computations through rapid reconfiguration of cortical dynamics.** *Neuron* 2018, **98**:1005–1019.e5.
96. Gallego JA, Perich MG, Miller LE, Solla SA: **Neural manifolds for the control of movement.** *Neuron* 2017, **94**:978–984.
97. Gallego JA, Perich MG, Chowdhury RH, Solla SA, Miller LE: **Long-term stability of cortical population dynamics underlying consistent behavior.** *Nat Neurosci* 2020, **23**:260–270.
98. Gallego JA, Perich MG, Naufel SN, Ethier C, Solla SA, Miller LE: **Cortical population activity within a preserved neural manifold underlies multiple motor behaviors.** *Nat Commun* 2018, **9**:4233.
99. Sadler PT, Quick KM, Golub MD, Chase SM, Ryu SI, Tyler-Kabara EC, Yu BM, Batista AP: **Neural constraints on learning.** *Nature* 2014, **512**:423–426.
100. Krakauer JW, Ghazanfar AA, Gomez-Marin A, MacLver MA, Poeppl D: **Neuroscience needs behavior: correcting a reductionist bias.** *Neuron* 2017, **93**:480–490.