



A Biased Bayesian Inference for Decision-Making and Cognitive Control

Kaosu Matsumori^{1,2*}, Yasuharu Koike³ and Kenji Matsumoto^{1*}

¹ Tamagawa University Brain Science Institute, Machida, Tokyo, Japan, ² Department of Information Processing, Tokyo Institute of Technology, Yokohama, Kanagawa, Japan, ³ Institute of Innovative Research, Tokyo Institute of Technology, Yokohama, Kanagawa, Japan

OPEN ACCESS

Edited by:

Paul E. M. Phillips,
University of Washington,
United States

Reviewed by:

Vijay Mohan K. Nambodiri,
University of North Carolina at Chapel
Hill, United States
Hansem Sohn,
Massachusetts Institute
of Technology, United States

*Correspondence:

Kenji Matsumoto
matsumot@lab.tamagawa.ac.jp
Kaosu Matsumori
kaosu.matsumori@gmail.com

Specialty section:

This article was submitted to
Decision Neuroscience,
a section of the journal
Frontiers in Neuroscience

Received: 27 June 2018

Accepted: 24 September 2018

Published: 12 October 2018

Citation:

Matsumori K, Koike Y and
Matsumoto K (2018) A Biased
Bayesian Inference
for Decision-Making and Cognitive
Control. *Front. Neurosci.* 12:734.
doi: 10.3389/fnins.2018.00734

Although classical decision-making studies have assumed that subjects behave in a Bayes-optimal way, the sub-optimality that causes biases in decision-making is currently under debate. Here, we propose a synthesis based on exponentially-biased Bayesian inference, including various decision-making and probability judgments with different bias levels. We arrange three major parameter estimation methods in a two-dimensional bias parameter space (prior and likelihood), of the biased Bayesian inference. Then, we discuss a neural implementation of the biased Bayesian inference on the basis of changes in weights in neural connections, which we regarded as a combination of leaky/unstable neural integrator and probabilistic population coding. Finally, we discuss mechanisms of cognitive control which may regulate the bias levels.

Keywords: sub-optimality, parameter estimation, probability judgment, probabilistic population codes, gain modulation, two-alternative forced choice, cognitive control, computational psychiatry

INTRODUCTION

Decision-making and cognitive control, which are well developed in primates (especially humans), are important for adaptive behaviors in a changing environment with uncertainty (Fuster, 2015). Many studies in various research fields such as Psychology, Neuroscience, and Engineering have suggested that decision-making and cognitive control are implemented differently in brain regions which contribute to processing these functions (Stuss and Knight, 2013; Fuster, 2015). However, the computational principles underlying these processes are still under debate (Gazzaniga, 2009; Friston, 2010).

Recent theoretical studies suggest a unified view accounting for various cognitive functions including decision-making and cognitive control on the basis of Bayesian inference (which forces us to re-assess our prior subjective beliefs using currently observed data, while considering uncertainty as well), as means of obtaining the distributions (Friston, 2010; Beck et al., 2012; Pouget et al., 2013). This view has been supported by many empirical studies showing that information is integrated in a nearly Bayes-optimal way in motor control (Wolpert et al., 1995; Ernst and Banks, 2002; Körding and Wolpert, 2004), and also in perceptual decision-making (Yang and Shadlen, 2007; van Bergen et al., 2015).

Körding and Wolpert (2004) showed that previously-learned probability distributions were combined with newly-added probability distributions in a Bayesian fashion, in the performance of a reaching task (Körding and Wolpert, 2004). A cursor indicated the start and end points of the reaching act with a hidden hand, but an experimental shift was added at the end point, using Gaussian noise.

In this experiment, human subjects performed the reaching task by directing their index finger to the target. After learning the Gaussian distribution of the experimental shift, (first-probability distribution, i.e., the *prior*), additional feedback was given at the midpoint of the reaching with different Gaussian noises (second probability distribution, the *likelihood*). Subjects' behavior suggested that the two probability distributions were combined in a nearly Bayes-optimal way.

Brunton et al. (2013) showed that the variability of perceptual decision-making depends on sensory noise but not on the noise in the information-accumulation processes in an auditory pulse number-discrimination task as detailed in a mathematical model (Brunton et al., 2013). Because the information accumulation processes can be regarded as Bayesian inference (Bitzer et al., 2014), their findings suggest that the information accumulation for the perceptual decision-making is nearly Bayes-optimal.

On the other hand, models that assume bias in the information accumulation process, such as in Decision Field Theory and in the "leaky competing accumulator" model (Busemeyer and Townsend, 1993; Usher and McClelland, 2001), have also been proposed to explain the contextual effects (i.e., compromise effect, attractiveness effect, and similarity effect), as well as the primacy/recency effect. These models take into account deviations from Bayesian inference, as explained below. Interestingly, Powers et al. (2017) recently reported that hallucinations, abnormal perception in schizophrenia, can be explained by such deviations.

Classical economic decision-making studies have assumed that humans behave optimally (von Neumann and Morgenstern, 1947). However, the existence of cognitive biases in probability judgment has been amply shown (Kahneman et al., 1982). There is a well-known question that leads people to answer incorrectly: "It is known that the probability of contracting this disease is 1/10,000 and correct rates of positive and negative results obtained by a test for the disease are 99% for both ill and healthy subjects. Now, supposing the test results were positive for you. What would you estimate the probability is that you are actually ill?" People tend to answer greater than the correct rate such as "90%" despite the fact that, (mathematically), the correct answer is about 1% ($0.99 \cdot 0.0001 / (0.99 \cdot 0.0001 + 0.01 \cdot 0.9999) \approx 0.0098$). Here, they are considered to ignore the prior (or "base rate neglect") (Kahneman et al., 1982). Other biases such as "representativeness bias," "conservatism," and "anchoring and adjustment" can also be considered as deviations from optimal Bayesian inference (Kahneman et al., 1982).

Thus, various psychological phenomena have been described as optimal Bayesian inference or systematic deviations, pertaining perceptual decision-making and economic decision-making or probability judgment. However, there have been only limited attempts to explain perceptual decision making and economic decision-making within a unified framework (Summerfield and Tsetsos, 2012).

In this paper, we use the perspective of generalized Bayesian inference by considering the bias, and explain various decision-making, probability judgment, and cognitive control. We also

consider the possibility that the brain actually implements such biased Bayesian inference.

First, we describe a Bayesian inference model with two exponential biases. Next, we arrange major parameter estimation methods [maximum likelihood (ML) estimation, maximum posterior probability (MAP) estimation and usual Bayesian estimation] in a two-dimensional parameter space of biased Bayesian inference. This parameter space also accounts for biases in probability judgment which have served as a basis for the development of behavioral economics. Furthermore, we discuss the neural implementation of Bayesian inference with exponential biases on the basis of neural connections weight changes (Goldman et al., 2009), regarded as a combination of the leaky competing accumulator model (LCA) (Usher and McClelland, 2001), and probabilistic population codes (PPC) (Ma et al., 2006). Finally, we apply our framework to cognitive control.

BIASED BAYESIAN INFERENCE

The Bayesian method is a powerful tool that enables inference and decision-making even with a limited amount of data, succeeding where traditional statistical methods have failed to capture inherent dynamics. This method has been applied in many research fields, such as Genetics, Linguistics, Image processing, Cosmology, Ecology, Machine learning, Psychology, and Neuroscience (Stone, 2013).

In all these cases, the solutions were based on a very simple Bayes's theorem, which states that the probability of the hypothesis H after observation of data D is proportional to the product of the likelihood within the received data, and the prior probability of the hypothesis, or: ($P(H|D) \propto P(D|H)P(H)$). However, when it comes to human and animal behavior, deviations from Bayes-optimality have been observed in many cases (Kahneman et al., 1982). In some studies, such deviations have been explained by introducing exponential biases (i.e., inverse temperature parameters), on Bayesian inference (Nassar et al., 2010; Soltani and Wang, 2010; Payzan-LeNestour and Bossaerts, 2011; Payzan-LeNestour and Bossaerts, 2012; Payzan-LeNestour et al., 2013), mainly because these were found useful in expressing bias levels. Each exponential bias can be separately considered as a bias for a corresponding single distribution, and they are credited for introducing separate biases for prior probability and likelihood.

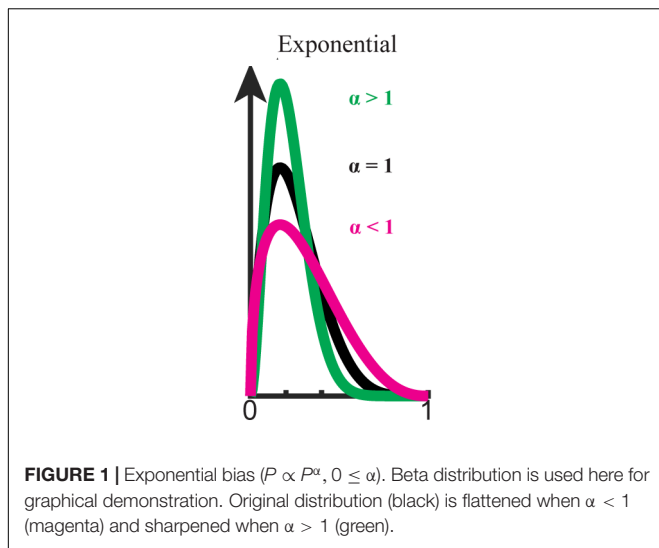
Here, We Propose the Importance of Introducing Both Biases on Prior Probability and Likelihood Simultaneously (see **Figure 1**).

$$P(H|D) \propto P(D|H)^\beta P(H)^\alpha \quad (1)$$

where \propto stands for proportionality, α is the exponential bias of the prior probability and β is the exponential bias of the likelihood. Equation (1) can be taken as its logarithm,

$$\log(P(H|D)) = \beta \log P(D|H) + \alpha \log P(H) + \text{const.} \quad (2)$$

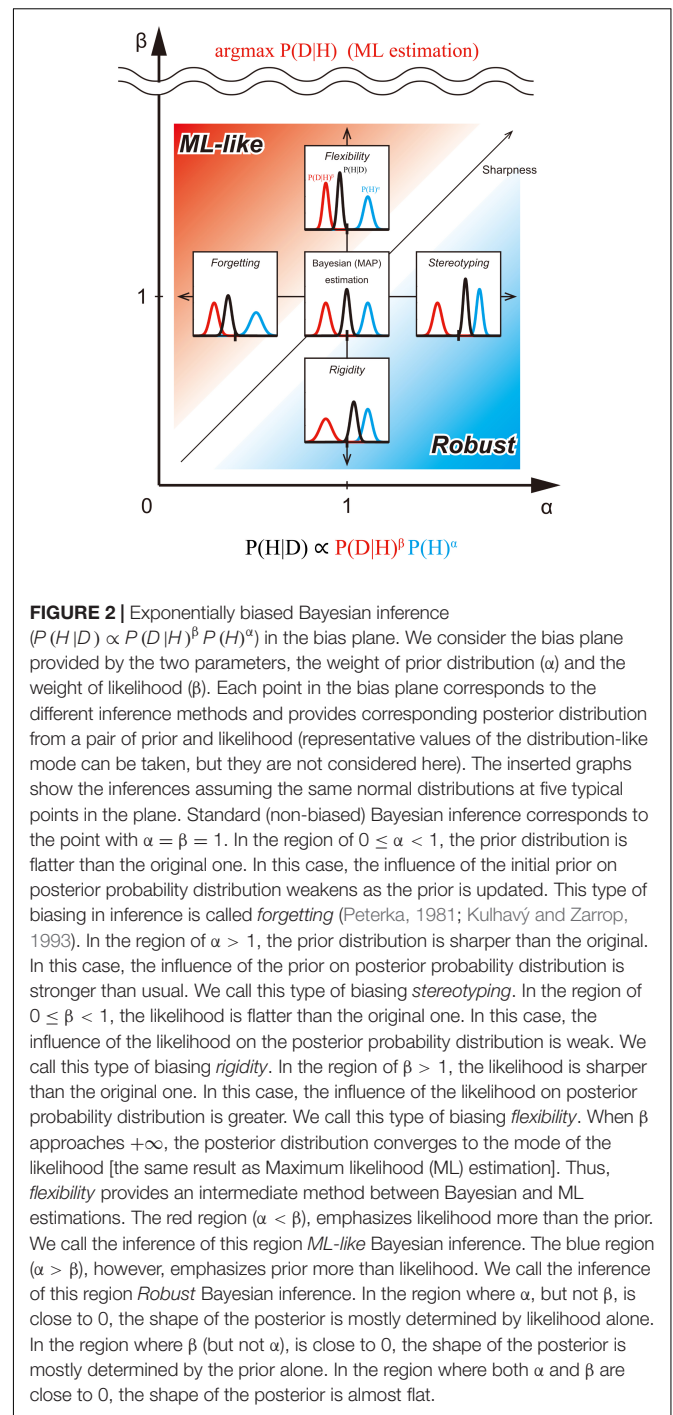
where α is the weight of the logarithm of prior probability, β is the weight of the logarithm of likelihood, and *const.* is the constant term for normalization (see **Supplementary Material**).



α and β was regarded as *biases* in Equation (1) but as *weights* in Equation (2). We will take the former term, *biases*, in the present paper so as to keep consistency throughout this paper and emphasize the deviation from the optimal inference. Some authors have expressed Bayes' theorem as the sum of the log-likelihood ratio and odds ratio (Grether, 1980; Soltani and Wang, 2010). However, they only considered a 2-alternative choice task, which cannot address interesting phenomena such as preference reversals observed in an alternative choice task with more than two choices, while most decisions we make are among more than two choices (Churchland and Ditterich, 2012).

Here, we assume that the bias levels (α, β) are positive (but see **Supplementary Material**). When $\alpha = \beta = 1$, the inference is just as in usual Bayesian inference. When the weight of the prior distribution in the inference is smaller than usual ($0 < \alpha < 1$), the prior distribution is flatter than the original. In this case, the influence of the initial prior on the posterior probability distribution weakens as the prior is updated (see **Supplementary Material**). This type of biasing in inference is called *forgetting*, because the older the information becomes, the less influence it has (Peterka, 1981; Nassar et al., 2010; Payzan-LeNestour and Bossaerts, 2011, Payzan-LeNestour and Bossaerts, 2012). In a non-stationary environment fluctuating gradually, the older information is less important than the newer in predicting the next state. Thus, *forgetting* enables us to take into account the non-stationarity of the environment when we do not have a complete model of the environment (Peterka, 1981; Kulhavý and Zorrop, 1993).

When the weight of the prior distribution in the inference is larger than usual (i.e., $\alpha > 1$), the prior distribution is sharper than the original (see **Figure 2**). In this case, the influence of the prior on the posterior probability distribution is stronger than usual. We call this type of biasing *stereotyping*, because the influence of old information never decreases, but in fact increases.



When the weight of the likelihood in the inference is smaller than usual (i.e., $0 < \beta < 1$), the likelihood is flatter than the original. In this case, the influence of the likelihood on the posterior probability distribution is weak. We call this type of biasing *rigidity*, because observed data is taken less into account. When the observed data are outliers, the posterior distribution can be greatly influenced by the data. *Rigidity* reduces the influence of such outlier data (Agostinelli and Greco, 2012, 2013).

When the weight of the likelihood in the inference is larger than usual (i.e., $\beta > 1$), the likelihood is sharper than before. In this case, the influence of the likelihood on posterior probability distribution is greater. We call this type of biasing *flexibility*, because the observed data is strongly influential on the posterior probability. If $\beta \rightarrow +\infty$, the posterior distribution becomes the mode of the likelihood. Thus, *flexibility* allows for a compromise between Bayesian estimation and Maximum Likelihood (ML) estimations (see **Figure 2**).

Here, we would like to inform readers that more than two sources of data can be considered simultaneously. In such cases, the likelihood weights of the data sources (β_1, β_2, \dots) might be different; (see section *Cognitive control as gain modulation and biased Bayesian inference*).

Parameter Estimation Based on the Biased Bayesian Inference

Standard Bayesian inference has been used for Bayesian estimation and Maximum a-posteriori (MAP) estimation. In the previous section, we introduced Bayesian inference with exponential biases. Next, we consider a variety of parameter estimation methods in the bias plane ($\alpha - \beta$ plane) for biased Bayesian inference (**Figure 2**).

Each point in the bias plane provides a corresponding posterior distribution based on a pair of priors and likelihood. So, each point corresponds to the different inference methods. We can consider different estimation methods depending on what is returned, as well as representative values of the distribution such as mean, median and mode, or the distribution itself.

In the case of returning the representative value of distribution (here, in this paper, focusing on mode), the point with $(\alpha, \beta) = (1, 1)$ corresponds to the standard MAP estimation. Since the likelihood is weighted more than the prior in the region where $\alpha < \beta$ is satisfied, (as shown above the 45° line in **Figure 2**), we call the estimation of this region, *ML-like* estimation. In this region, we can consider intermediate estimation between MAP and ML. In particular, when $\alpha = k_1$ and $\beta = 1$, the inference corresponds to the ML estimation if $k_1 = 0$, MAP estimation if $k_1 = 1$, and the MAP-ML intermediate estimation (*forgetting*), if $0 < k_1 < 1$ (horizontal arrow in **Figure 2**). Furthermore, when $\alpha = 1$ and $k_2 = 1/\beta$, the inference corresponds to the ML method if $k_2 = 0$, MAP estimation if $k_2 = 1$, and another type of MAP-ML intermediate estimation (*flexibility*), if $0 < k_2 < 1$, (vertical arrow in **Figure 2**).

By contrast, since the prior is weighted more than the likelihood in the region where $\alpha > \beta$ is satisfied (below the 45° line in **Figure 2**), we call the estimation of this region the “*robust estimation*.”

In this region, we can consider intermediate estimation between MAP (and, so to speak), “*maximum prior (MP)*” estimation. In particular, when $\alpha = k_1$ and $\beta = 1$, the inference corresponds to MAP estimation if $k_1 = 1$, the *MP* estimation, if $k_1 \rightarrow +\infty$, and the MAP-*MP* intermediate estimation (*stereotyping*), if $k_1 > 1$ (horizontal arrow in **Figure 2**).

Furthermore, when $\alpha = 1$ and $k_2 = 1/\beta$, the inference corresponds to the MAP estimation if $k_2 = 1$, the *MP* method if $k_2 \rightarrow +\infty$ and another type of MAP-*MP* intermediate estimation (*rigidity*), if $k_2 > 1$; (vertical arrow in **Figure 2**).

In the case of returning distribution itself, the point with $(\alpha, \beta) = (1, 1)$ corresponds to the standard Bayesian estimation. In the *ML-like* region, when $\alpha = 1$ and $k_2 = 1/\beta$, the inference corresponds to the ML method if $k_2 = 0$ because it is considered a point estimation that returns $\arg \max P(D|H)$ (as the sharpness of likelihood should be the maximum above the scale break in **Figure 2**), thus, Bayesian estimation results if $k_2 = 1$, and the Bayesian-ML intermediate estimation (*flexibility*) when $0 < k_2 < 1$ (**Figure 2**). In the *Robust* region, when $\alpha = 1$ and $k_2 = 1/\beta$, the inference corresponds to the Bayesian estimation if $k_2 = 1$; the inference just returns the prior if $k_2 \rightarrow +\infty$, and the Bayesian-prior intermediate estimation (*rigidity*), if $k_2 > 1$ (vertical arrow in **Figure 2**).

Mapping the Psychological Biases of Probability Judgment on the Bias Plane

In this section, we map the psychological biases on the bias plane (**Figure 2**). In the context of human probability judgment, many deviations from Bayesian inference are known, such as base-rate fallacy, representativeness bias, conservation, anchoring and adjustment etc. (Tversky and Kahneman, 1974; Kahneman et al., 1982). In the bias plane, base-rate fallacy and representativeness correspond to the *ML-like* estimation when $\alpha < \beta$ (as in **Figure 2**), because the likelihood is more influential than the prior in these deviations. Here, we emphasize the distinction of the bias which comes from *forgetting* ($\alpha < 1$), and from *flexibility* ($\beta > 1$), which are similar in the aspect that the posterior distribution is drawn more to the likelihood rather than the prior, but arises from different mechanisms (Brock, 2012; Pellicano and Burr, 2012a; **Figure 2**).

Conservation and anchoring-and-adjusting correspond to *robustness* (where $\alpha > \beta$), because the prior is more influential than the likelihood in these deviations. Here we emphasize the distinction of the bias which comes from *stereotyping* (i.e., when $\alpha > 1$), and from *rigidity* (when $\beta < 1$), which are similar in the aspect that the posterior distribution is drawn more from the prior than the likelihood, but comes from different mechanisms (**Figure 2**). Note that our approach is applicable to variable problem situations and ones with more than two alternatives, while most previous attempts to quantitatively represent these biases have considered only two simple alternatives (Grether, 1980; Soltani and Wang, 2010; see also section 2AFC).

Biased Bayesian inference enables us to take into account the non-stationarity of the environment when we do not have a complete model of the environment. So, these psychological biases may have adaptive meanings and normative account as suggested by previous studies (Kulhavý and Zarrop, 1993; Gigerenzer and Goldstein, 1996). Their studies propose that the bias framework presented here provides insights into the sources of influences on these biases.

NEURAL IMPLEMENTATION OF BIASED BAYESIAN INFERENCE

Neural Representation of the Probability Distribution

Some authors have proposed computational models in which neural circuits perform Bayesian inference (Jazayeri and Movshon, 2006; Ma et al., 2006). Neural circuits need to represent probability distributions to perform Bayesian inference. Since synaptic inputs usually work additively (Kandel et al., 2000), Bayesian inference can be achieved by summation of neural activity in the brain. To make it possible to achieve Bayesian inference by summation, the neural activity should represent a logarithm of probabilities. In fact, several empirical studies have reported that neurons encode posterior probability in logarithmic form (Figure 3A; Yang and Shadlen, 2007; Kira et al., 2015).

Two theoretical methods have been proposed to encode the log probability distribution in a neural population (Pouget et al., 2013; Ma and Jazayeri, 2014). The first method is log probability coding in which the firing rate of each neuron is proportional to the log or logit of the probability of the event or belief coded by the neuron (Barlow, 1969; Jazayeri and Movshon, 2006; Pouget et al., 2013; Figure 3A). By simply pooling the activity of the neurons considered, the probability distribution can be represented by the neural population (Yang and Shadlen, 2007; Pouget et al., 2013). The second, more sophisticated method, is probabilistic population coding (PPC), in which a basis function is combined with the log probability coding. And $\log P(x)$ can be written as follows (Ma et al., 2006; Pouget et al., 2013):

$$\log P(x) = \sum_i r_i h_i(x) + \text{const.} = \mathbf{r} \cdot \mathbf{h} + \text{const.} \quad (3)$$

where r_i is the firing rate of i th neuron, $\mathbf{r} = (r_1, r_2, \dots, r_n)^T$, $h_{-i}(x)$ is basis function for an i th neuron, and $\mathbf{h} = (h_1(x), h_2(x), \dots, h_n(x))^T$.

In this paper, we propose a neural model of biased Bayesian inference on the basis of PPC.

Implementing Biased Bayesian Inference in a Neuronal Population

In considering the implementation of standard and biased Bayesian inference below, we assume two layers of neuronal populations with identical basis functions (see **Supplementary Material**). The first layer consists of two neuronal populations encoding the prior and the likelihood, and the second layer consists of a neuronal population encoding the posterior distribution (Figure 4A).

In PPC, standard Bayesian inference is achieved by summing the firing rates of corresponding neurons between the neuronal population encoding the prior, and that encoding the likelihood, at the next layer of a neuronal-population-encoded posterior distribution, because summation of their log probabilities is equivalent to the product of the prior and likelihood (Ma et al., 2006).

In this paper, we introduced biased Bayesian inference with exponential biases above ($P(H|D) \propto P(D|H)^\beta P(H)^\alpha$). Biased Bayesian inference can be achieved just by changing the gain (α, β) of the inputs for the next (posterior) layer considered in the implementation of the standard Bayesian inference (Equation 4; Figure 4A).

$$\begin{aligned} \log(P(D|H)^\beta P(H)^\alpha) &= \beta \log P(D|H) + \alpha \log P(H) \\ &= \beta \mathbf{r}_{\text{likelihood}} \cdot \mathbf{h} + \alpha \mathbf{r}_{\text{prior}} \cdot \mathbf{h} + \text{const.} \\ &= (\beta \mathbf{r}_{\text{likelihood}} + \alpha \mathbf{r}_{\text{prior}}) \cdot \mathbf{h} + \text{const.} \\ &= \mathbf{r}_{\text{posterior}} \cdot \mathbf{h} + \text{const.} \end{aligned} \quad (4)$$

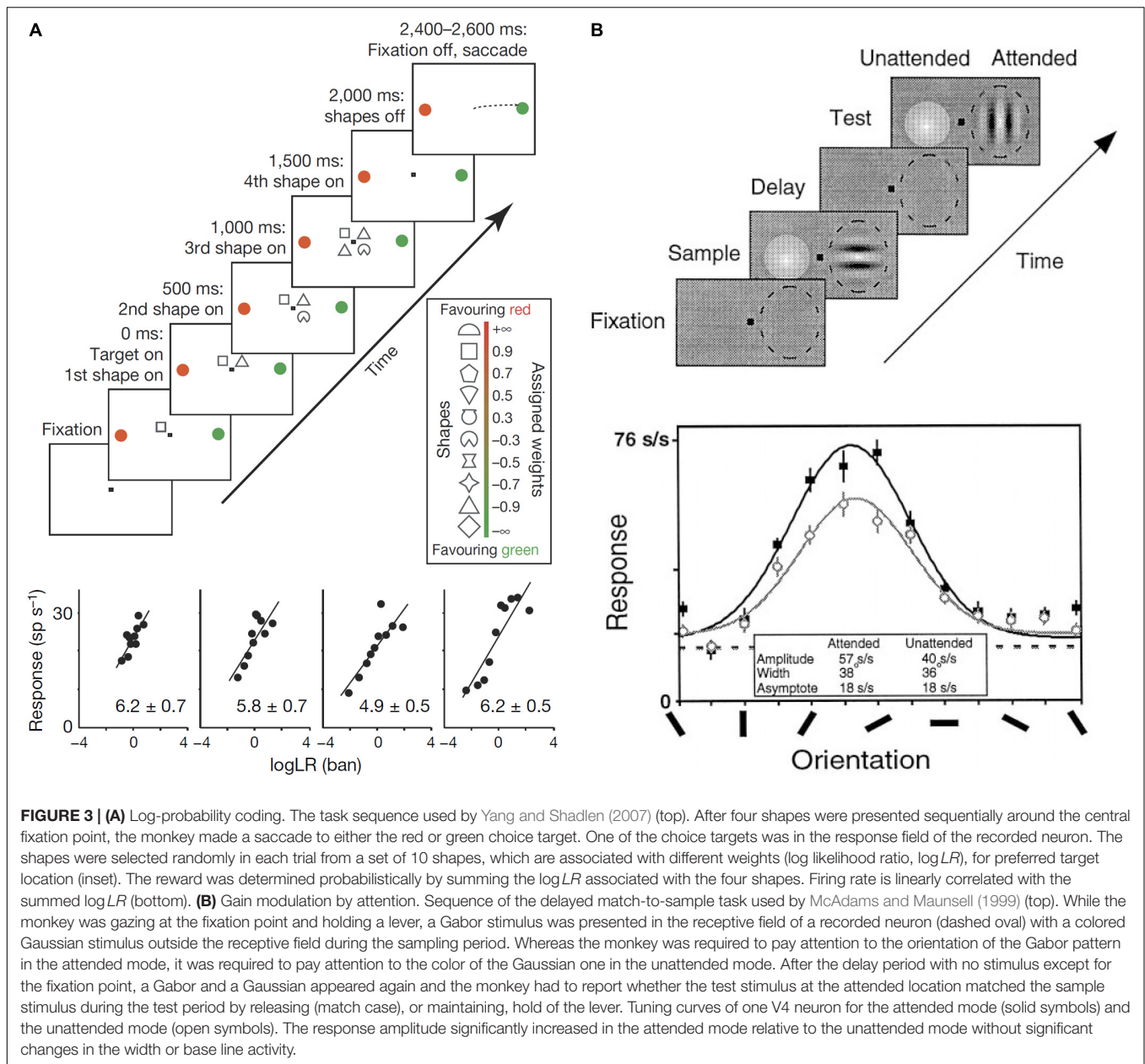
where $\mathbf{r}_{\text{prior}}$, $\mathbf{r}_{\text{likelihood}}$ and $\mathbf{r}_{\text{posterior}}$ are the firing rates of neural populations encoding the prior, likelihood, and posterior, respectively. α and β correspond to the bias for connection weights between prior-encoding neurons and posterior encoding-neurons and the bias for connection weight between likelihood-encoding neurons and posterior-encoding neurons, respectively. Thus, $\mathbf{r}_{\text{posterior}}$ is calculated as $\beta \mathbf{r}_{\text{likelihood}} + \alpha \mathbf{r}_{\text{prior}}$. Such gain modulation of neuronal firing rates is ordinarily used in the brain (Desimone and Duncan, 1995; Desimone, 1998) without clearly sophisticated, high cost mechanism, so biasing by gain modulation may be good a strategy of the brain to provide solutions that are approximately optimal.

Recurrent Connection and Neural Integrator

In Bayesian updating, the posterior distribution generally becomes the prior distribution for the following iterative time step. The resolution of the time steps for Bayesian updating depends on the task; it is trial-based when sensory evidence for updating the prior is given only at the end of a trial (Glimcher, 2003), whereas it is based on finer temporal steps in a single trial when sensory evidence is sequentially (or continuously) given throughout the trial (Bogacz et al., 2006; Kira et al., 2015). Such repetitive updating can be achieved regardless of the temporal resolution by a recurrent neural circuit, in which the firing rates of neurons last across the time steps (Goldman et al., 2009; Bitzer et al., 2014; Kira et al., 2015; Chandrasekaran, 2017). Neural integrator models have been considered as a mechanism for maintain the firing rate with external and recurrent inputs (Goldman et al., 2009). Since the dynamics of neural integrators can be described using a firing rate equation (Dayan and Abbott, 2001; Goldman et al., 2009), biased Bayesian inference can therefore be expressed by Equation (5) (Figure 4B).

$$\tau_{\text{neuron}} \frac{d\mathbf{r}}{dt} = -\mathbf{r} + \text{Input} = -\mathbf{r} + \alpha \mathbf{r} + \beta \mathbf{r}_{\text{likelihood}}(t) \quad (5)$$

where τ_{neuron} is the intrinsic decay time constant of neurons, \mathbf{r} represents the firing rate of the neurons that encode prior/posterior distribution, and Input is input to the neurons that encode prior/posterior distribution ($\alpha \mathbf{r} + \beta \mathbf{r}_{\text{likelihood}}(t)$), α is strength of recurrent connection, β is strength of feedforward connection and $\mathbf{r}_{\text{likelihood}}$ is the firing rate of the neuronal



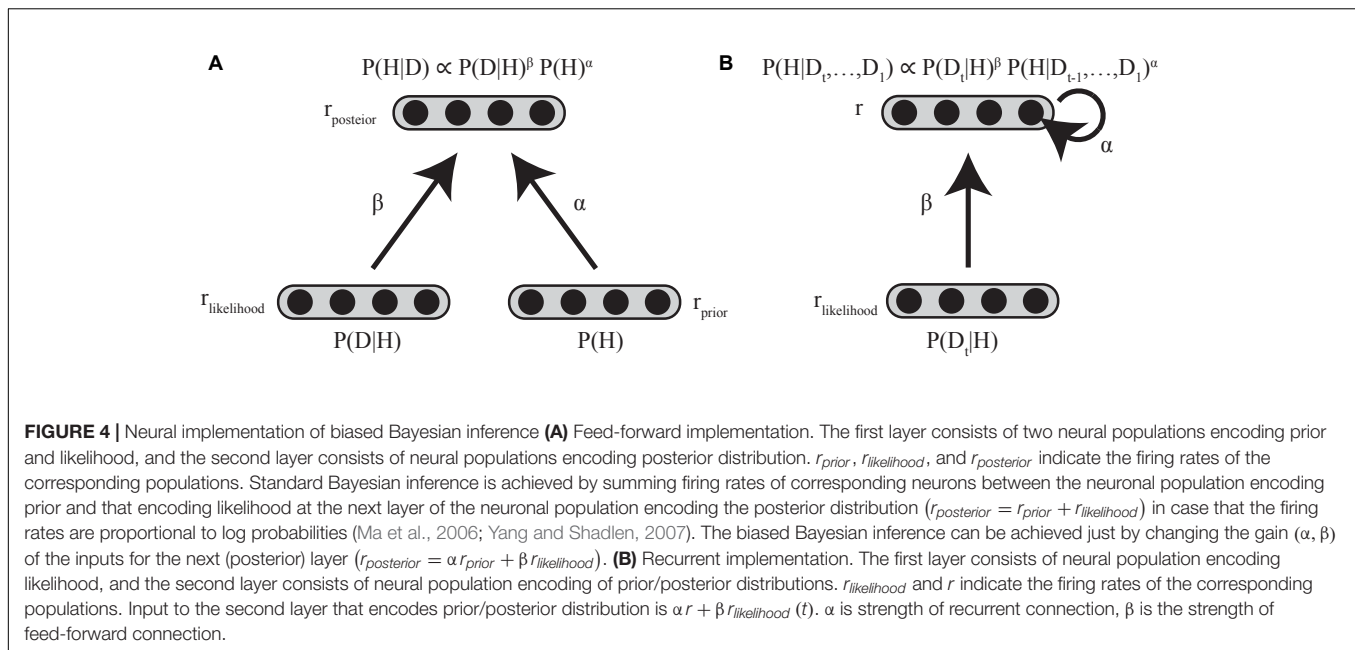
population that encodes the likelihood (with external input to be integrated).

In the neural integrator model, the strength of recurrent input required for neurons to independently maintain their firing rates is α (Goldman et al., 2009). When a neuron maintains its firing rate in the absence of external input, the neural integrator is called *balanced* ($\alpha = 1$), because the firing rate changes due to recurrent input and intrinsic decay are regarded as balanced. If the firing rate decreases due to weak recurrent input, it does not perfectly compensate for the intrinsic decay of the firing rate, and therefore, the neural integrator is called *leaky* ($\alpha < 1$), and in this case, encoded distribution decays over time (*forgetting*). If the firing rate increases due to recurrent input and overcompensates for the intrinsic decay of the firing

rate, the neural integrator is called *unstable* ($\alpha > 1$). In this case, encoded distribution is sharpened over time (what we call *stereotyping*).

Linking Biased Bayesian Inference to the Two-Alternative Forced Choice (2AFC)

2AFC decision-making has been well-studied (Stone, 1960; Ratcliff, 1978; Busemeyer and Townsend, 1993; Gold and Shadlen, 2001; Roe et al., 2001; Shadlen and Newsome, 2001; Usher and McClelland, 2001, 2004; Wang, 2002; Mazurek et al., 2003; Bogacz et al., 2006; Wong and Wang, 2006; Kiani et al., 2008; Tsetsos et al., 2012), because this is one of the simplest cases of decision-making.



The accumulator models of 2AFC such as the Drift-diffusion model (DDM) (Ratcliff, 1978), decision field theory (DFT) (Busemeyer and Townsend, 1993; Roe et al., 2001), and the leaky competing accumulator model (LCA) (Usher and McClelland, 2001, 2004), typically make three assumptions:

- (i) Evidence favoring each alternative is integrated over time in a single trial;
- (ii) The process is subject to random fluctuations; and,
- (iii) A decision is made when sufficient evidence has accumulated favoring one alternative over another (Bogacz et al., 2006).

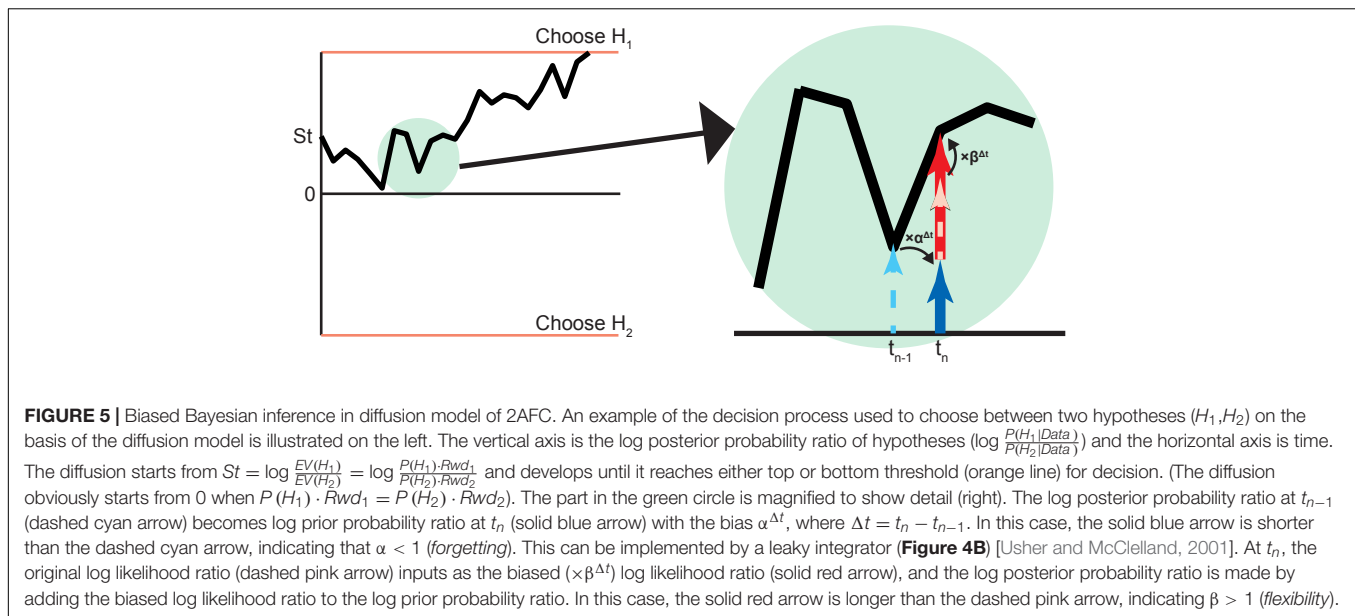
These models are drawn from neuronal data particularly in monkey lateral intraparietal cortex (LIP) (Kiani et al., 2008; Brunton et al., 2013).

DDM uses an optimal integrator, and this model can be seen as the continuum limit of the Sequential Probability Ratio test (SPRT) – the optimal sequential hypothesis-testing method (Neyman and Pearson, 1933; Wald and Wolfowitz, 1948; Wald, 1973; Bogacz et al., 2006; Kira et al., 2015). Because the instantaneous drift is considered the likelihood, the instantaneous drift of DDM can be considered as a form of Bayesian inference (Bogacz et al., 2006; Bitzer et al., 2014) by taking the starting point of accumulation as the initial prior.

In contrast to DDM, DFT, and LCA assume leaky neural integrators instead of the optimal integrator, thereby explaining many behavioral biases such as the primacy/recency effect and three preference reversal effects (compromise effect, attractiveness effect, and similarity effect) (Busemeyer and Townsend, 1993; Roe et al., 2001; Usher and McClelland, 2001, 2004; Bogacz et al., 2006) in two- or multi-alternative forced choice tasks. However, the normative Bayesian perspective has never been applied to DFT or LCA. A simplified form of the DFT or LCA, the Ornstein–Uhlenbeck process (Bogacz et al., 2006),

is consistent with the recurrent network model (Figure 4B), whose inputs to the second layer have Gaussian noise. Thus, DFT and LCA can be considered forms of biased Bayesian inference (Figure 5). Here, we would like to emphasize that biased Bayesian inference is a key tool in combining the descriptive richness of models of 2AFC and the normative Bayesian perspective which are implemented in PPC (Beck et al., 2012) (Supplementary Material).

In addition to Bayesian inference, an intrinsic gain/loss function should also be considered in the selection of “an action in the environment” (Ernst and Bulthoff, 2004). When different reward values are contingent on corresponding action selections, the reward function $Rwd(x)$ dependent on an executed action x must also be considered (Matsumoto et al., 2003, 2006; Matsumoto and Tanaka, 2004). The $Rwd(x)$ can be encoded by population firing rates r_{reward} , as posterior probability $P(x)$ denoted here as $r_{\text{prior/posterior}}$. Since the expected value of reward $EV(x)$ is calculated by multiplying the probability of obtaining the reward with the reward value on the basis of action x , the combination between the Bayesian inference and intrinsic gain/loss function may be obtained by simple multiplication of $Rwd(x)$ and $P(x)$. In other words, $EV(x)$ can be denoted as $r_{\text{prior/posterior}} + r_{\text{reward}}$, where, ($\log EV(x) = \log P(x) + \log Rwd(x)$), (in considering that the firing rates are proportional to log probabilities). Actually, recent studies have shown that the difference in reward between directions changes the starting point of integration, which corresponds to the prior knowledge, irrespective of mean drift rate in the diffusion model on the basis of fronto-parietal activity (Rorie et al., 2010; Summerfield and Koechlin, 2010; Mulder et al., 2012). Thus, Bayesian inference and the calculation of the expected value of reward can be achieved by the same mechanism if one takes the reward value as a prior (Friston et al., 2013).



Importantly, the mean drift rate in value integration is also modulated by spatial visual attention (Krajbich et al., 2010). This finding suggests that the level of exponential bias for $Rwd(x)$ changes depending on attention, which could be shifted more by the higher value of reward associated with the target processed in the fronto-striatal network (Schultz, 2006; Hikosaka, 2007).

Empirically Testable Predictions in 2AFC Tasks

There is some concern that Bayesian approaches in psychology and neuroscience are often “just-so stories” because it can often provide many degrees of freedom for explanations and they are not falsifiable (Bowers and Davis, 2012). In the comment paper for Bowers and Davis (2012) and Griffiths et al. (2012) wrote: “In evaluating claims about falsifiability, it is useful to distinguish between a model and a theoretical framework. A model is proposed to account for a specific phenomenon and makes specific assumptions in order to do so. A theoretical framework provides a general perspective and a set of tools for making models. [...] Models are falsifiable, but frameworks are typically not” (Griffiths et al., 2012).

Here, we provide some falsifiable predictions about the behavioral and neural models of biased Bayesian inference for 2AFC tasks, which is a typical model on the basis of our biased Bayesian inference, the theoretical framework in the present paper.

As for the biased Bayesian inference behavioral models for 2AFC task, it would be easily falsified merely if the decision processes were not based on the prior, likelihood, and gain/loss function, because our model for 2AFC assumes the subjects’ decision processes are based on them as typical Bayesian inference model is. “Bayesian transfer” is a useful experimental procedure to determine whether a decision process is based on the prior, likelihood, and gain/loss function (Maloney and Mamassian, 2009). In this procedure, subjects are overtrained for

two decision tasks which contain different priors, likelihoods, and gain/loss functions (Task 1, Task 2) until subjects’ performances come close to maximizing expected gain in both tasks. Then, it is tested whether the subjects can transfer the knowledge about priors, likelihoods, and gain/loss functions acquired in Task 1 and Task 2 into a new task (transfer task). For example, the transfer task contains the same likelihood and gain/loss function as those in Task 1 and the same prior as that in Task 2. The subjects are familiar with the prior, likelihood, and gain/loss function in the resulting transfer task, but the combination of prior, likelihood, and gain function is novel. One can state that subjects’ decision processes are based on the prior, likelihood, and gain/loss function when the subjects’ performances in the transfer task immediately close to ideal without further practice or learning (Maloney and Mamassian, 2009). Actually, Bayesian transfer should be efficiently applied to 2AFC tasks, because the prior and gain/loss function can be easily changed by instruction and the likelihood functions is easily switched with another in the tasks.

Even if the biased Bayesian inference behavioral models for 2AFC passed the Bayesian transfer test, the models would be further falsified if they were non-biased. This falsification check is important because it has been considered that the information accumulation for the perceptual decision-making is nearly Bayes-optimal rather than biased or leaky/unstable, at least in short time integration (Gold and Shadlen, 2007; Brunton et al., 2013). In order to check whether the integration is leaky/unstable ($\alpha \neq 1$) by behavioral data with the assumption of constant α and β , one can use psychophysical reverse correlation (Ahumada, 1996; Neri et al., 1999; Okazawa et al., 2018). Psychophysical reverse correlation is a technique that estimates how sensory information is weighted to guide decisions by quantifying the spatiotemporal stimulus fluctuations that precede each choice (Okazawa et al., 2018). If sensory weights are constant throughout the integration process, the integration can be seen as

perfect integration. Otherwise, one can state that the integration is leaky/unstable.

As for the biased Bayesian inference neural models, we proposed it on the basis of the empirical findings that LIP neurons encode the accumulated information (prior/posterior distribution) and middle temporal cortex (MT) neurons encode instant sensory information (likelihood function) (Mazurek et al., 2003; Gold and Shadlen, 2007). Therefore, the biased Bayesian inference neural models would be falsified, simply if firing rates of LIP neurons were explained by perfect integration ($\alpha = 1$) because it would just deny our assumption of biased, leaky/unstable integration ($\alpha \neq 1$).

The other bias for likelihood (β) can also be considered. One may be interested in the biases originated from inference process itself ($\beta_{connection}$), which is different from the biases originated from representation of probability distributions (β_{rate}) (Supplementary Material). One can state that the origin of bias is inference process itself ($\beta_{connection} \neq 1$) if the choice behaviors change associated with the corresponding changes in firing rates of LIP neurons without changes in firing rate of MT neurons. Actually, such a situation can be available in switching between two tasks, one to discriminate motion direction and the other to discriminate stereoscopic depth in the same moving random dot stereogram stimuli (Sasaki and Uka, 2009; Kumano et al., 2016). Therefore, by using such a context-dependent task switching, the biased Bayesian inference neural models whose biases originated from inference process itself would be falsified, if firing rates of LIP and MT neurons were rather consistent to the neural models with the biases originated from representation of probability distributions.

COGNITIVE CONTROL AS GAIN MODULATION AND BIASED BAYESIAN INFERENCE

We have already seen that the biased Bayesian inference can be explained by gain modulation in neuronal populations. Here, we explain how the bias levels may be regulated by cognitive control.

Cognitive control is thought to occur in two steps: first, the anterior cingulate cortex (ACC) monitors cognitive control demands on the basis of uncertainty in response to outcomes and sends signals to the prefrontal cortex (PFC), and then, the PFC recruits this cognitive control by sending top-down signals to various cortical areas (Botvinick et al., 2001, 2004; Matsumoto, 2004; Shenhav et al., 2013).

Application to Stroop task

The Stroop task is a well-known task that has been used to examine the neural mechanisms of cognitive control. This task consists of word reading and color naming. In the word reading task, the subject is asked to report the written color name of a color, printed in the same color (*congruent*); or another (*incongruent*) (wherein, “green” is the answer for GREEN printed in red). Since human subjects in modern countries are trained more to read words as opposed to reporting colors, relatively low cognitive control is recruited in this task. In the color

naming task, a subject is asked to name the ink color of a word denoting the same color (*congruent*) or a different color (*incongruent*) (e.g., “red” being the correct answer for the word “GREEN” printed in red), and relatively high cognitive control must be recruited to overcome the cognitive interference. Particularly, the reaction times are longer and the error rates are higher in the *incongruent* condition than in the *congruent* one – what is now known as the “Stroop effect” (Stroop, 1935).

Botvinick and colleagues proposed a neural network model of cognitive control focusing on the conflict monitoring function of the ACC (Figure 6; Botvinick et al., 2001, 2004; Shenhav et al., 2013). Here, we show that this model can be considered a neural model of a Bayesian network with biases (Rao, 2005). When the stimulus is presented, word-and color-encoding neurons in the perception layer are activated. We conjecture that these neurons encode log-likelihoods of word and color.

The succeeding response layer receives input from the perception layer. We conjecture that the neurons in the response layer encode prior/posterior distributions for responses such as “green,” “red” and so on; (here, a flat prior being assumed). According to Botvinick’s model, because we are more adept at reading words rather than naming colors due to the more prevalent experience of enforcement on reading, the connection between the word-encoding neurons and the response layer is stronger than that between the color-encoding neurons and the response layer.

An *incongruent* stimulus, however (e.g., “GREEN” printed in red), activates not only the neurons encoding “green” but also the neurons encoding “red” in the response layer. And, accordingly, the responses of the two types of neurons should be closer to each other than the responses to the congruent stimulus, i.e., “GREEN” printed in green. The comparable firing rates of the neurons encoding different responses come into conflict, thus activating the ACC. The ACC is then believed to send a cognitive control signal to PFC. In turn, the PFC regulates the gains of word and color encoding neurons through a top-down signal, which amounts to regulation of bias levels ($\beta_{word}, \beta_{color}$) in our biased Bayesian inference (Figure 6). The top-down signal augments the gains of neurons encoding task-relevant information (i.e., word-encoding neurons in the word reading task, and color-encoding neurons in the color naming task), and then the firing rates of the two types of neurons in response layers for “green” and “red” become differentiated in accordance with task requirements. The result being, that the appropriate response is achieved.

Application to Working Memory

Cognitive control works in a variety of situations requiring flexible behaviors that use goal-directed, top-down selection of relevant information, as in, “maintain it tentatively despite irrelevant distractors, and utilize it to solve a problem” (a process akin to working memory) (Baddeley, 1986).

Now, we consider working memory within the framework of biased Bayesian inference. Top-down attentional regulation of working memory can be regarded as being governed by neural

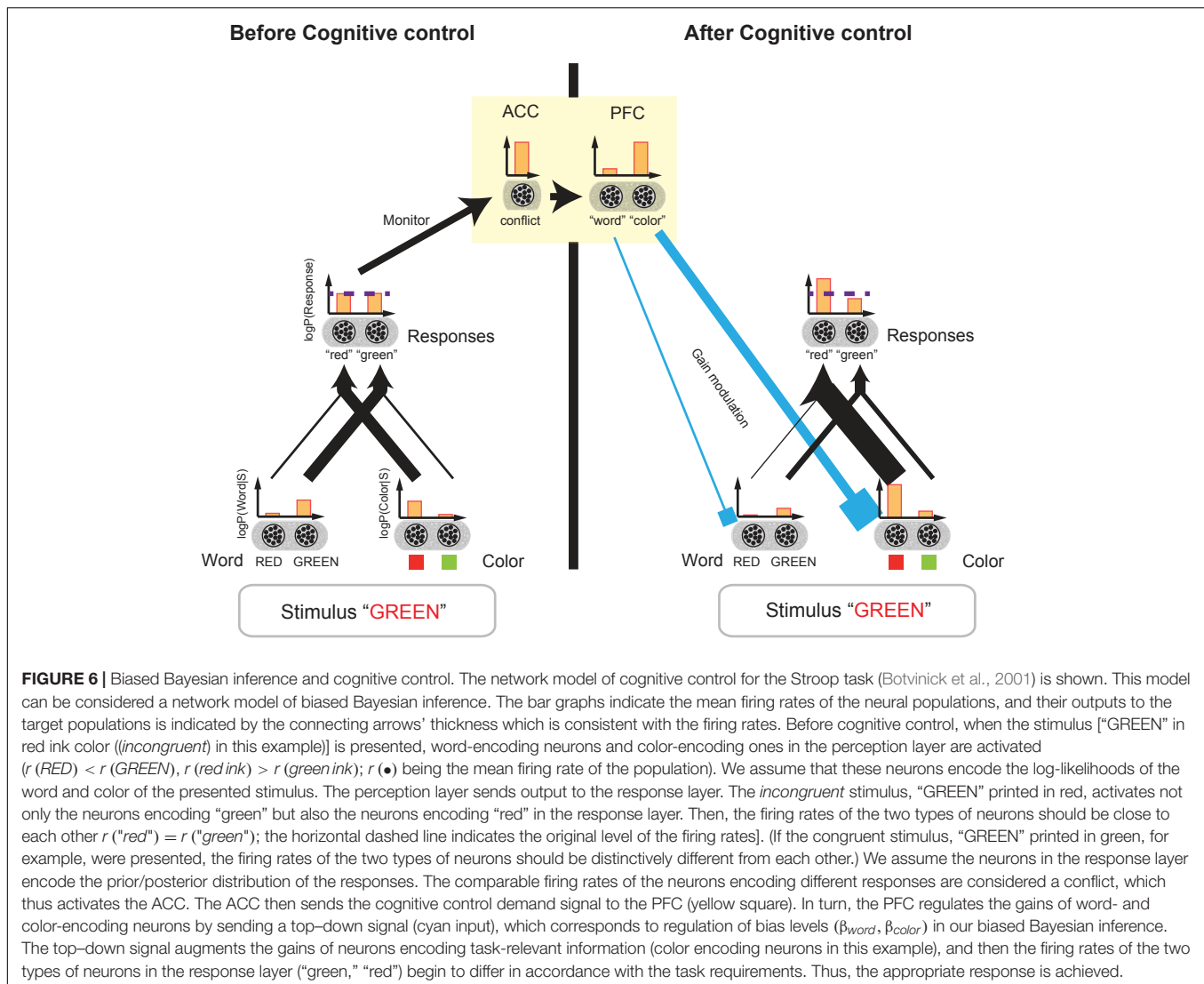


FIGURE 6 | Biased Bayesian inference and cognitive control. The network model of cognitive control for the Stroop task (Botvinick et al., 2001) is shown. This model can be considered a network model of biased Bayesian inference. The bar graphs indicate the mean firing rates of the neural populations, and their outputs to the target populations is indicated by the connecting arrows' thickness which is consistent with the firing rates. Before cognitive control, when the stimulus ["GREEN" in red ink color (*incongruent* in this example)] is presented, word-encoding neurons and color-encoding ones in the perception layer are activated ($r(\text{RED}) < r(\text{GREEN})$, $r(\text{red ink}) > r(\text{green ink})$; $r(\bullet)$ being the mean firing rate of the population). We assume that these neurons encode the log-likelihoods of the word and color of the presented stimulus. The perception layer sends output to the response layer. The *incongruent* stimulus, "GREEN" printed in red, activates not only the neurons encoding "green" but also the neurons encoding "red" in the response layer. Then, the firing rates of the two types of neurons should be close to each other $r(\text{red}) = r(\text{green})$; the horizontal dashed line indicates the original level of the firing rates]. (If the congruent stimulus, "GREEN" printed in green, for example, were presented, the firing rates of the two types of neurons should be distinctively different from each other.) We assume the neurons in the response layer encode the prior/posterior distribution of the responses. The comparable firing rates of the neurons encoding different responses are considered a conflict, which thus activates the ACC. The ACC then sends the cognitive control demand signal to the PFC (yellow square). In turn, the PFC regulates the gains of word- and color-encoding neurons by sending a top-down signal (cyan input), which corresponds to regulation of bias levels (β_{word} , β_{color}) in our biased Bayesian inference. The top-down signal augments the gains of neurons encoding task-relevant information (color encoding neurons in this example), and then the firing rates of the two types of neurons in the response layer ("green," "red") begin to differ in accordance with the task requirements. Thus, the appropriate response is achieved.

integrators whose gains are considered the biases in a biased Bayesian inference.

Working memory is actively maintained longer than several seconds even across distracting stimuli (Miller et al., 1996; Miyake and Shah, 1999). Persistent activity based on top-down attention might be sustained by a reciprocal positive feedback loop within a population of neurons in certain cortical regions including the PFC (Curtis and D'Esposito, 2003; Curtis and Lee, 2010).

The attentional top-down signals from the PFC improve working memory (Desimone and Duncan, 1995; Miller and Cohen, 2001; Noudoost et al., 2010; Gazzaley and Nobre, 2012), suggesting that the top-down signals work to change the gain of neural connections implemented in working memory circuits. Here, the attentional top-down signals should contribute to adaptive modulation of either α or β (Norman and Shallice, 1986; Desimone and Duncan, 1995; McAdams and Maunsell, 1999; Botvinick et al., 2001; Miller and Cohen, 2001; Egner and Hirsch, 2005; Maunsell and Treue, 2006; Reynolds and Heeger, 2009;

Cole et al., 2013, 2014), as the gain of the neural connections corresponds to α in the biased Bayesian inference. Therefore, this can be regarded as a change in the balance of neural integrators (i.e., $\alpha = 1$ corresponds to the balanced integrator, $\alpha < 1$ to the leaky integrator, and $\alpha > 1$, to the unstable integrator; Roe et al., 2001; Gazzaley and Nobre, 2012).

Application to Top-Down Attention

Top-down attention amplifies the tuning curves of activity in sensory cortical regions (McAdams and Maunsell, 1999; Treue and Trujillo, 1999; Maunsell and Treue, 2006; **Figure 3B**). This is because the tuning curve amplification itself can be considered a form of noise reduction (Dayan and Zemel, 1999), and the effect of the bias (β) on the target of attention should be based on the top-down attentional signal that reduces the noise of the target.

Tuning curves are not completely flattened even for task-irrelevant features such as distractors. For example, it is well known that subjects' behaviors are affected by distractors in a variety of tasks that require top-down attention (Stroop, 1935;

Eriksen and Eriksen, 1974). This cannot be explained by standard, non-biased Bayesian inference assuming binary β s for task-relevant and task-irrelevant features, but by biased Bayesian inference that can take arbitrary values as for the β s. This can be considered a Bayesian version of the biased competition model (Desimone and Duncan, 1995; Desimone, 1998).

Here, we do not make strong assumptions of normality for represented probability distributions as Friston et al. (2006) and Friston (2009), and Friston (2010) did in their considerations of gain modulation in the context of predictive coding based on a free energy principle. Our approach sacrifices the mathematical tractability of theirs, and instead makes the models applicable to recently developed decision-making and cognitive control theories by allowing the representation of arbitrary probability distributions.

One important issue is how to determine the bias levels as suggested by Shenhav et al. (2013). Also, we would like to note to readers that some studies have treated the prior as attention (Angela and Dayan, 2004; Chikkerur et al., 2010). However, these issues are beyond the scope of this paper.

POSSIBLE RELATIONSHIP AMONG NEUROMODULATORS/NEUROTRANSMITTERS, GAIN MODULATION, AND PSYCHIATRY

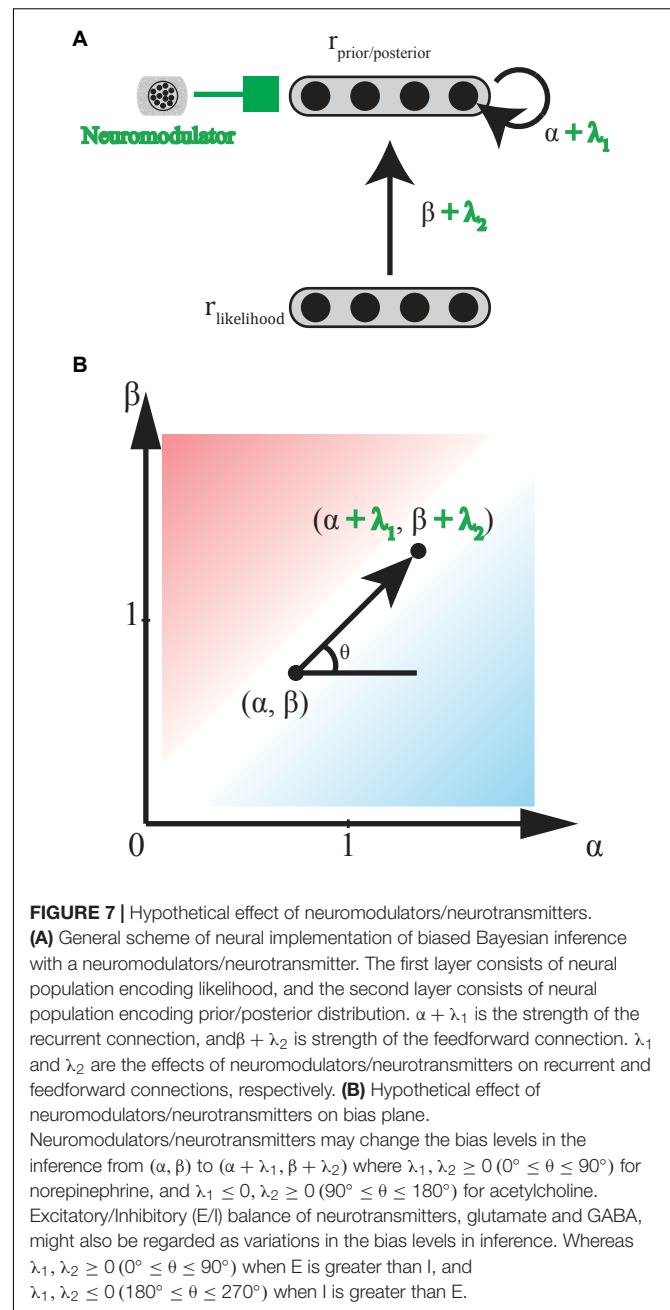
Detailed mechanisms of gain modulation and application to psychiatry are also important (**Supplementary Material**) (Friston et al., 2006; Friston, 2009; Friston, 2010; Mante et al., 2013; Miller and Buschman, 2013), since psychiatric diseases could be caused by changes in connectivity between the PFC and other regions which lead to defects in cognitive control (Cole et al., 2014; Stephan et al., 2015).

Potent candidates that modulate gain for cognitive or top-down attention control are classical neuromodulators/neurotransmitters such as norepinephrine, acetylcholine, glutamate, λ -aminobutyric acid (Friston et al., 2006; Friston, 2009; Friston, 2010).

Norepinephrine

Norepinephrine (NE)-containing cells in the locus coeruleus (LC) receive ongoing task utility values [as evaluated by ACC and orbitofrontal cortex (OFC)], release NE at cortical areas involved in stimulus judgment and behavioral decision making to modulate the gain of current task-relevant information specifically (exploitation) or possible alternatives (exploration) (Servan-Schreiber et al., 1990; Aston-Jones and Cohen, 2005; Yu and Dayan, 2005). This gain modulation corresponds to bias levels (α and β) in the biased Bayesian inference, where the sharper distribution is led by higher levels of NE (**Figure 7**).

Attention deficit/hyperactivity disorder (ADHD) patients show NE dysfunction. They not only show erratic trial-to-trial exploratory behavior but also high trial-to-trial variability in their reaction times in 2AFC tasks with probabilistic reinforcement (Frank et al., 2007; Hauser et al., 2016). Consistent with



these behavioral findings, a computational simulation suggests that appropriate NE release leads to sharper motor cortical representations and sharper distribution of reaction times (Frank et al., 2007).

Acetylcholine

Acetylcholine (ACh) release from basal forebrain cells modulates the gain of processing in cortical circuits involved in memory in two ways (Hasselmo and McGaughy, 2004; Hasselmo, 2006). ACh enhances the cortical responses to afferents inputting new inputs for encoding on one hand, but suppresses the cortical responses to recurrent feedback for memory maintenance on the other

hand. Thus, whereas high ACh levels enhance new encoding, low ACh levels enhance maintenance and following consolidation. Yu and Dayan (2005) proposed a consistent model suggesting that ACh levels are high when a presented stimulus does not indicate an appropriate response with certainty, (here, called the expected uncertainty), while other inputs may still be informative. This gain modulation corresponds to a function with an ML-like-robust axis ($\beta - \alpha$) in the biased Bayesian inference (Figure 7): where high and low ACh levels correspond to ML-like and robust inferences, respectively (but see Hasselmo, 2006; Hasselmo and Sarter, 2011).

Patients with Alzheimer's disease (associated with ACh hypofunction), show working memory impairment (Baddeley et al., 1991). This may be caused by robust inference on the basis of a least input (reduced β), and/or *unstable* neural integrator (enlarged α).

Excitatory/Inhibitory (E/I) Balance

Glutamate (Glu) and gamma-aminobutyric acid (GABA), respectively, act as important excitatory and inhibitory neurotransmitters in the brain (Kandel et al., 2000). The ratio of these two neurotransmitters (i.e., "E/I balance" in a functional unit of neural circuitry), should determine the gain of intrinsic activity in the circuit. This gain corresponds to α in the biased Bayesian inference. The functional unit of neural circuitry receives plenty of glutamatergic and GABAergic inputs externally. Thus, the E/I balance of these external inputs should determine the gain of the activity which they evoke in the circuit. This gain corresponds to α and β in the biased Bayesian inference (Figure 7).

The E/I balance in cerebral cortex is altered in schizophrenic patients (Wang, 2001; Kehrer et al., 2008; Murray et al., 2014). There are two major positive symptoms of schizophrenia: hallucinations (i.e., perceiving something that is not actually there), and delusions (erroneous beliefs that usually involve a misinterpretation of perceptions or experiences). It has been suggested that neither hallucinations nor delusions occur when gain levels based on both prior knowledge and external input from the environment are correct. On the basis of the predictive coding hypothesis, Corlett et al. (2009) have suggested that hallucinations occur when the gain level of prior knowledge exceeds that of external input and delusions occur when the gain level of external input exceeds that of prior knowledge (Corlett et al., 2009; Fletcher and Frith, 2009; Teufel et al., 2015; Powers et al., 2017). Because the gain levels of prior knowledge and external input are expected to depend on the E/I balance of their corresponding connections, hallucinations and delusions can be explained by an abnormal E/I balance.

Autism is a heritable, lifelong neurodevelopmental condition characterized by difficulties in social communication and social interaction (called social symptoms), and a range of restricted activities and sensory abnormalities (called non-social symptoms) (Pellicano and Burr, 2012b). Some forms of autism are also thought to be caused by an altered E/I balance in sensory, mnemonic, social and emotional systems (Rubenstein and Merzenich, 2003; Nelson and Valakh, 2015). Since E/I balance determines bias levels in the biased Bayesian inference,

the altered E/I balance in autism implies inappropriately-biased neural processing (Figure 4). Here, we focus on the process causing a sensory abnormality in autism, since the construction of a Bayesian model for social communication and social interaction has not yet reached consensus.

As we normally experience strong visual illusions such as in the Kanizsa triangle (Kanizsa, 1955) and Shepard's table (Shepard, 1990), the brain postulates the most likely interpretation for the noisy, ambiguous sensory signals corresponding to strong priors (Pellicano and Burr, 2012b). By contrast, individuals with autism are less susceptible to such illusions and show more accurate or overly realistic perceptions, suggesting that their priors are flattened (Pellicano and Burr, 2012b) (i.e., $\alpha < 1$, or *forgetting* in the bias plane). However, as Brock (2012) has noted, the same results can be obtained with a sharpened likelihood, when $\beta > 1$ (with *flexibility* in the bias plane) as well, because these two biases make similar posterior distributions (ML-like inferences), as we have explained in Figure 2.

LIMITATIONS

The biased Bayesian models especially in the context of 2AFC tasks would be empirically identifiable to some extent as described in section Empirically Testable Predictions in 2AFC Tasks. However, details of neural implementation of biased Bayesian inference such as discrimination between the biases originated from inference process and those originated from representation of probability distributions and importance of gain/loss function in the model identifiability issue are not fully addressed. Also, the proposed models for cognitive control and those based on assumed neuromodulators/neurotransmitters functions for understanding mental disorders are insufficient to test their identifiability. Because the identifiability issue should be severer due to more parameters than that for usual Bayesian models, rigorous model comparisons must be done in future studies that apply the biased Bayesian models.

CONCLUSION

In the present paper, we have discussed the importance of introducing exponentially-biased Bayesian inference into a cognitive framework with testable neural correlates. The bias plane for the biased Bayesian inference is useful in providing a view-integrating ML estimation, Bayesian estimation, and MAP estimation, thus enabling us to make intermediate parameter estimations between ML estimation and Bayesian (or MAP) estimation. It also explains many psychological decision biases, such as base-rate fallacy, representativeness, conservatism, anchoring-and-adjusting, primacy/recency effect, three-preference reversal effects (compromise effect, attractiveness effect, and similarity effect), as well as cognitive control. Furthermore, it may provide further insight into the mechanisms of ADHD, Alzheimer's disease, and schizophrenia (including hallucinations and delusions), and, potentially, autism as well.

AUTHOR CONTRIBUTIONS

KaM conceptualized the data. KaM and KeM wrote the paper. KaM, KeM, and YK revised the paper.

FUNDING

This research was partially supported by JSPS KAKENHI Grant No. JP15H03124, and by the program for Brain Mapping by Integrated Neurotechnologies for Disease Studies (Brain/MINDS) from the Japan Agency for Medical Research and development (AMED).

REFERENCES

- Agostinelli, C., and Greco, L. (2012). "Weighted likelihood in Bayesian inference," in *Proceedings of the 46th Scientific Meeting of the Italian Statistical Society*, New York, NY.
- Agostinelli, C., and Greco, L. (2013). A weighted strategy to handle likelihood uncertainty in Bayesian inference. *Comput. Stat.* 28, 319–339. doi: 10.1007/s00180-011-0301-1
- Ahumada, A. J. (1996). Perceptual classification images from vernier acuity masked by noise. *Perception* 25, 1831–1840. doi: 10.1068/v9610501
- Angela, J. Y., and Dayan, P. (2004). "Inference, attention, and decision in a Bayesian neural architecture," in *Proceedings of the Advances in Neural Information Processing Systems*, Vancouver, 1577–1584.
- Aston-Jones, G., and Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.* 28, 403–450. doi: 10.1146/annurev.neuro.28.061604.135709
- Baddeley, A. (1986). *Working Memory*. Oxford: Clarendon Press.
- Baddeley, A. D., Bressi, S., Della Sala, S., Logie, R., and Spinnler, H. (1991). The decline of working memory in Alzheimer's disease. A longitudinal study. *Brain* 114(Pt 6), 2521–2542. doi: 10.1093/brain/114.6.2521
- Barlow, H. B. (1969). Pattern recognition and the responses of sensory neurons. *Ann. N. Y. Acad. Sci.* 156, 872–881. doi: 10.1111/j.1749-6632.1969.tb14019.x
- Beck, J. M., Ma, W. J., Pitkow, X., Latham, P. E., and Pouget, A. (2012). Not noisy, just wrong: the role of suboptimal inference in behavioral variability. *Neuron* 74, 30–39. doi: 10.1016/j.neuron.2012.03.016
- Bitzer, S., Park, H., Blankenburg, F., and Kiebel, S. J. (2014). Perceptual decision making: drift-diffusion model is equivalent to a Bayesian model. *Front. Hum. Neurosci.* 8:102. doi: 10.3389/fnhum.2014.00102
- Bogacz, R., Brown, E., Moehlis, J., Holmes, P., and Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol. Rev.* 113, 700–765. doi: 10.1037/0033-295x.113.4.700
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., and Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychol. Rev.* 108, 624–652. doi: 10.1037/0033-295X.108.3.624
- Botvinick, M. M., Cohen, J. D., and Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends Cogn. Sci.* 8, 539–546. doi: 10.1016/j.tics.2004.10.003
- Bowers, J. S., and Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychol. Bull.* 138, 389–414. doi: 10.1037/a0026450
- Brock, J. (2012). Alternative Bayesian accounts of autistic perception: comment on Pellicano and Burr. *Trends Cogn. Sci.* 16, 573–574. doi: 10.1016/j.tics.2012.10.005
- Brunton, B. W., Botvinick, M. M., and Brody, C. D. (2013). Rats and humans can optimally accumulate evidence for decision-making. *Science* 340, 95–98. doi: 10.1126/science.1233912
- Busmeyer, J. R., and Townsend, J. T. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychol. Rev.* 100, 432–459. doi: 10.1037/0033-295X.100.3.432

ACKNOWLEDGMENTS

We thank William T. Newsome, Yukihito Yomogida, Kazuki Iijima, Ayaka Sugiura, Ryuta Aoki, Kou Murayama, and the reviewers for helpful discussions and comments on previous versions of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2018.00734/full#supplementary-material>

- Chandrasekaran, C. (2017). Computational principles and models of multisensory integration. *Curr. Opin. Neurobiol.* 43, 25–34. doi: 10.1016/j.conb.2016.11.002
- Chikkerur, S., Serre, T., Tan, C., and Poggio, T. (2010). What and where: a Bayesian inference theory of attention. *Vis. Res.* 50, 2233–2247. doi: 10.1016/j.visres.2010.05.013
- Churchland, A. K., and Ditterich, J. (2012). New advances in understanding decisions among multiple alternatives. *Curr. Opin. Neurobiol.* 22, 920–926. doi: 10.1016/j.conb.2012.04.009
- Cole, M. W., Repovs, G., and Anticevic, A. (2014). The frontoparietal control system: a central role in mental health. *Neuroscientist* 20, 652–664. doi: 10.1177/1073858414525995
- Cole, M. W., Reynolds, J. R., Power, J. D., Repovs, G., Anticevic, A., and Braver, T. S. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nat. Neurosci.* 16, 1348–1355. doi: 10.1038/nn.3470
- Corlett, P. R., Frith, C. D., and Fletcher, P. C. (2009). From drugs to deprivation: a Bayesian framework for understanding models of psychosis. *Psychopharmacology* 206, 515–530. doi: 10.1007/s00213-009-1561-0
- Curtis, C. E., and D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends Cogn. Sci.* 7, 415–423. doi: 10.1016/S1364-6613(03)00197-9
- Curtis, C. E., and Lee, D. (2010). Beyond working memory: the role of persistent activity in decision making. *Trends Cogn. Sci.* 14, 216–222. doi: 10.1016/j.tics.2010.03.006
- Dayan, P., and Abbott, L. F. (2001). *Theoretical Neuroscience*. Cambridge, MA: MIT Press.
- Dayan, P., and Zemel, R. S. (1999). *Statistical Models and Sensory Attention. IET Conference Proceedings [Online]*. Available at: http://digital-library.theiet.org/content/conferences/10.1049/cp_19991246
- Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philos. Trans. R. Soc. Lon. B Biol. Sci.* 353, 1245–1255. doi: 10.1098/rstb.1998.0280
- Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222. doi: 10.1146/annurev.ne.18.030195.001205
- Egner, T., and Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nat. Neurosci.* 8, 1784–1790. doi: 10.1038/nn1594
- Eriksen, B. A., and Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Percept. Psychophys.* 16, 143–149. doi: 10.3758/BF03203267
- Ernst, M. O., and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429–433. doi: 10.1038/415429a
- Ernst, M. O., and Bulthoff, H. H. (2004). Merging the senses into a robust percept. *Trends Cogn. Sci.* 8, 162–169. doi: 10.1016/j.tics.2004.02.002
- Fletcher, P. C., and Frith, C. D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat. Rev. Neurosci.* 10, 48–58. doi: 10.1038/nrn2536

- Frank, M. J., Santamaria, A., O'Reilly, R. C., and Willcutt, E. (2007). Testing computational models of dopamine and noradrenaline dysfunction in attention deficit/hyperactivity disorder. *Neuropsychopharmacology* 32, 1583–1599. doi: 10.1038/sj.npp.1301278
- Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends Cogn. Sci.* 13, 293–301. doi: 10.1016/j.tics.2009.04.005
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787
- Friston, K., Kilner, J., and Harrison, L. (2006). A free energy principle for the brain. *J. Physiol. Paris* 100, 70–87. doi: 10.1016/j.jphysparis.2006.10.001
- Friston, K., Schwartenbeck, P., Fitzgerald, T., Moutoussis, M., Behrens, T., and Dolan, R. J. (2013). The anatomy of choice: active inference and agency. *Front. Hum. Neurosci.* 7:598. doi: 10.3389/fnhum.2013.00598
- Fuster, J. (2015). *The Prefrontal Cortex*, 5th Edn. New York, NY: Elsevier. doi: 10.1016/B978-0-12-407815-4.00002-7
- Gazzaley, A., and Nobre, A. C. (2012). Top-down modulation: bridging selective attention and working memory. *Trends Cogn. Sci.* 16, 129–135. doi: 10.1016/j.tics.2011.11.014
- Gazzaniga, M. S. (2009). *The Cognitive Neurosciences*. Cambridge, MA: MIT press.
- Gigerenzer, G., and Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychol. Rev.* 103, 650–669. doi: 10.1037/0033-295X.103.4.650
- Glimcher, P. W. (2003). The neurobiology of visual-saccadic decision making. *Annu. Rev. Neurosci.* 26, 133–179. doi: 10.1146/annurev.neuro.26.010302.081134
- Gold, J. I., and Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends Cogn. Sci.* 5, 10–16. doi: 10.1016/S1364-6613(00)01567-9
- Gold, J. I., and Shadlen, M. N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.* 30, 535–574. doi: 10.1146/annurev.neuro.29.051605.113038
- Goldman, M. S., Compte, A., and Wang, X. J. (2009). “Neural integrator models,” in *Encyclopedia of Neuroscience*, ed. L. R. Squire (Oxford: Academic Press), 165–178. doi: 10.1016/B978-008045046-9.01434-0
- Grether, D. M. (1980). Bayes rule as a descriptive model: the representativeness heuristic. *Q. J. Econ.* 95, 537–557. doi: 10.2307/1885092
- Griffiths, T. L., Chater, N., Norris, D., and Pouget, A. (2012). How the Bayesians got their beliefs (and what those beliefs actually are): comment on Bowers and Davis (2012). *Psychol. Bull.* 138, 415–422. doi: 10.1037/a0026884
- Hasselmo, M. E. (2006). The role of acetylcholine in learning and memory. *Curr. Opin. Neurobiol.* 16, 710–715. doi: 10.1016/j.conb.2006.09.002
- Hasselmo, M. E., and McGaughy, J. (2004). High acetylcholine levels set circuit dynamics for attention and encoding and low acetylcholine levels set dynamics for consolidation. *Prog. Brain Res.* 145, 207–231. doi: 10.1016/s0079-6123(03)45015-2
- Hasselmo, M. E., and Sarter, M. (2011). Modes and models of forebrain cholinergic neuromodulation of cognition. *Neuropsychopharmacology* 36, 52–73. doi: 10.1038/npp.2010.104
- Hauser, T. U., Fiore, V. G., Moutoussis, M., and Dolan, R. J. (2016). Computational psychiatry of ADHD: neural gain impairments across marrian levels of analysis. *Trends Neurosci.* 39, 63–73. doi: 10.1016/j.tins.2015.12.009
- Hikosaka, O. (2007). Basal ganglia mechanisms of reward-oriented eye movement. *Ann. N. Y. Acad. Sci.* 1104, 229–249. doi: 10.1196/annals.1390.012
- Jazayeri, M., and Movshon, J. A. (2006). Optimal representation of sensory information by neural populations. *Nat. Neurosci.* 9, 690–696. doi: 10.1038/nn1691
- Kahneman, D., Slovic, P., and Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511809477
- Kandel, E. R., Schwartz, J. H., and Jessell, T. M. (2000). *Principles of Neural Science*. New York, NY: McGraw-Hill.
- Kanizsa, G. (1955). Margini quasi-percettivi in campi con stimolazione omogenea. *Riv. Psicol.* 49, 7–30.
- Kehrer, C., Maziashvili, N., Dugladze, T., and Gloveli, T. (2008). Altered excitatory-inhibitory balance in the NMDA-hypofunction model of schizophrenia. *Front. Mol. Neurosci.* 1:6. doi: 10.3389/neuro.02.006.2008
- Kiani, R., Hanks, T. D., and Shadlen, M. N. (2008). Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *J. Neurosci.* 28, 3017–3029. doi: 10.1523/JNEUROSCI.4761-07.2008
- Kira, S., Yang, T., and Shadlen, M. N. (2015). A neural implementation of Wald's sequential probability ratio test. *Neuron* 85, 861–873. doi: 10.1016/j.neuron.2015.01.007
- Körding, K. P., and Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature* 427, 244–247. doi: 10.1038/nature02169
- Krajbich, I., Armel, C., and Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nat. Neurosci.* 13, 1292–1298. doi: 10.1038/nn.2635
- Kulhavy, R., and Zarp, M. B. (1993). On a general concept of forgetting. *Int. J. Control* 58, 905–924. doi: 10.1080/00207179308923034
- Kumano, H., Suda, Y., and Uka, T. (2016). Context-dependent accumulation of sensory evidence in the parietal cortex underlies flexible task switching. *J. Neurosci.* 36, 12192–12202. doi: 10.1523/jneurosci.1693-16.2016
- Ma, W. J., Beck, J. M., Latham, P. E., and Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nat. Neurosci.* 9, 1432–1438. doi: 10.1038/nn1790
- Ma, W. J., and Jazayeri, M. (2014). Neural coding of uncertainty and probability. *Annu. Rev. Neurosci.* 37, 205–220. doi: 10.1146/annurev-neuro-071013-014017
- Maloney, L. T., and Mamassian, P. (2009). Bayesian decision theory as a model of human visual perception: testing Bayesian transfer. *Vis. Neurosci.* 26, 147–155. doi: 10.1017/s0952523808080905
- Mante, V., Sussillo, D., Shenoy, K. V., and Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 503, 78–84. doi: 10.1038/nature12742
- Matsumoto, K. (2004). Conflict and cognitive control. *Science* 303, 969–970. doi: 10.1126/science.1094733
- Matsumoto, K., Matsumoto, M., and Abe, H. (2006). Goal-based action selection and utility-based action bias. *Neural Netw.* 19, 1315–1320. doi: 10.1016/j.neunet.2006.05.036
- Matsumoto, K., Suzuki, W., and Tanaka, K. (2003). Neuronal correlates of goal-based motor selection in the prefrontal cortex. *Science* 301, 229–232. doi: 10.1126/science.1084204
- Matsumoto, K., and Tanaka, K. (2004). The role of the medial prefrontal cortex in achieving goals. *Curr. Opin. Neurobiol.* 14, 178–185. doi: 10.1016/j.conb.2004.03.005
- Maunsell, J. H., and Treue, S. (2006). Feature-based attention in visual cortex. *Trends Neurosci.* 29, 317–322. doi: 10.1016/j.tins.2006.04.001
- Mazurek, M. E., Roitman, J. D., Ditterich, J., and Shadlen, M. N. (2003). A role for neural integrators in perceptual decision making. *Cereb. Cortex* 13, 1257–1269. doi: 10.1093/cercor/bhg097
- McAdams, C. J., and Maunsell, J. H. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J. Neurosci.* 19, 431–441. doi: 10.1523/JNEUROSCI.19-01-00431.1999
- Miller, E. K., and Buschman, T. J. (2013). Cortical circuits for the control of attention. *Curr. Opin. Neurobiol.* 23, 216–222. doi: 10.1016/j.conb.2012.11.011
- Miller, E. K., and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202. doi: 10.1146/annurev.neuro.24.1.167
- Miller, E. K., Erickson, C. A., and Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J. Neurosci.* 16, 5154–5167. doi: 10.1523/JNEUROSCI.16-16-05154.1996
- Miyake, A., and Shah, P. (1999). *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139174909
- Mulder, M. J., Wagenmakers, E.-J., Ratcliff, R., Boekel, W., and Forstmann, B. U. (2012). Bias in the brain: a diffusion model analysis of prior probability and potential payoff. *J. Neurosci.* 32, 2335–2343. doi: 10.1523/JNEUROSCI.4156-11.2012
- Murray, J. D., Anticevic, A., Gancsos, M., Ichinose, M., Corlett, P. R., Krystal, J. H., et al. (2014). Linking microcircuit dysfunction to cognitive impairment: effects of disinhibition associated with schizophrenia in a cortical working memory model. *Cereb. Cortex* 24, 859–872. doi: 10.1093/cercor/bhs370

- Nassar, M. R., Wilson, R. C., Heasley, B., and Gold, J. I. (2010). An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *J. Neurosci.* 30, 12366–12378. doi: 10.1523/JNEUROSCI.0822-10.2010
- Nelson, S. B., and Valakh, V. (2015). Excitatory/inhibitory balance and circuit homeostasis in autism spectrum disorders. *Neuron* 87, 684–698. doi: 10.1016/j.neuron.2015.07.033
- Neri, P., Parker, A. J., and Blakemore, C. (1999). Probing the human stereoscopic system with reverse correlation. *Nature* 401, 695–698. doi: 10.1038/44409
- Neyman, J., and Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philos. Trans. R. Soc. Lon.* 231, 289–337. doi: 10.1098/rsta.1933.0009
- Norman, D. A., and Shallice, T. (1986). *Attention to Action*. Berlin: Springer.
- Noudoust, B., Chang, M. H., Steinmetz, N. A., and Moore, T. (2010). Top-down control of visual attention. *Curr. Opin. Neurobiol.* 20, 183–190. doi: 10.1016/j.conb.2010.02.003
- Okazawa, G., Sha, L., Purcell, B. A., and Kiani, R. (2018). Psychophysical reverse correlation reflects both sensory and decision-making processes. *Nat. Commun.* 9:3479. doi: 10.1038/s41467-018-05797-y
- Payzan-LeNestour, E., and Bossaerts, P. (2011). Risk, unexpected uncertainty, and estimation uncertainty: bayesian learning in unstable settings. *PLoS Comput. Biol.* 7:e1001048. doi: 10.1371/journal.pcbi.1001048
- Payzan-LeNestour, E., Dunne, S., Bossaerts, P., and O'Doherty, J. P. (2013). The neural representation of unexpected uncertainty during value-based decision making. *Neuron* 79, 191–201. doi: 10.1016/j.neuron.2013.04.037
- Payzan-LeNestour, É., and Bossaerts, P. (2012). Do not bet on the unknown versus try to find out more: estimation uncertainty and “unexpected uncertainty” both modulate exploration. *Front. Neurosci.* 6:150. doi: 10.3389/fnins.2012.00150
- Pellicano, E., and Burr, D. (2012a). Response to brock: noise and autism. *Trends Cogn. Sci.* 16, 574–575. doi: 10.1016/j.tics.2012.10.004
- Pellicano, E., and Burr, D. (2012b). When the world becomes “too real”: a Bayesian explanation of autistic perception. *Trends Cogn. Sci.* 16, 504–510. doi: 10.1016/j.tics.2012.08.009
- Peterka, V. (1981). Bayesian approach to system identification. *Trends Progr. Syst. Identif.* 1, 239–304. doi: 10.1016/B978-0-08-025683-2.50013-2
- Pouget, A., Beck, J. M., Ma, W. J., and Latham, P. E. (2013). Probabilistic brains: knowns and unknowns. *Nat. Neurosci.* 16, 1170–1178. doi: 10.1038/nn.3495
- Powers, A. R., Mathys, C., and Corlett, P. R. (2017). Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors. *Science* 357, 596–600. doi: 10.1126/science.aan3458
- Rao, R. P. N. (2005). Bayesian inference and attentional modulation in the visual cortex. *NeuroReport* 16, 1843–1848. doi: 10.1097/01.wnr.0000183900.92901.fc
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychol. Rev.* 85, 59–108. doi: 10.1037/0033-295X.85.2.59
- Reynolds, J. H., and Heeger, D. J. (2009). The normalization model of attention. *Neuron* 61, 168–185. doi: 10.1016/j.neuron.01.002
- Roe, R. M., Busemeyer, J. R., and Townsend, J. T. (2001). Multialternative decision field theory: a dynamic connectionist model of decision making. *Psychol. Rev.* 108, 370–392. doi: 10.1037/0033-295X.108.2.370
- Rorie, A. E., Gao, J., McClelland, J. L., and Newsome, W. T. (2010). Integration of sensory and reward information during perceptual decision-making in lateral intraparietal cortex (lip) of the macaque monkey. *PLoS One* 5:e9308. doi: 10.1371/journal.pone.0009308
- Rubenstein, J. L., and Merzenich, M. M. (2003). Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes Brain Behav.* 2, 255–267. doi: 10.1034/j.1601-183X.2003.00037.x
- Sasaki, R., and Uka, T. (2009). Dynamic readout of behaviorally relevant signals from area mt during task switching. *Neuron* 62, 147–157. doi: 10.1016/j.neuron.2009.02.019
- Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annu. Rev. Psychol.* 57, 87–115. doi: 10.1146/annurev.psych.56.091103.070229
- Servan-Schreiber, D., Printz, H., and Cohen, J. D. (1990). A network model of catecholamine effects: gain, signal-to-noise ratio, and behavior. *Science* 249, 892–895. doi: 10.1126/science.2392679
- Shadlen, M. N., and Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J. Neurophysiol.* 86, 1916–1936. doi: 10.1152/jn.2001.86.4.1916
- Shenhav, A., Botvinick, M. M., and Cohen, J. D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron* 79, 217–240. doi: 10.1016/j.neuron.2013.07.007
- Shepard, R. N. (1990). *Mind Sights: Original Visual Illusions, Ambiguities, and Other Anomalies, With a Commentary on the Play of Mind in Perception and Art*. New York, NY: W.H. Freeman.
- Soltani, A., and Wang, X.-J. (2010). Synaptic computation underlying probabilistic inference. *Nat. Neurosci.* 13, 112–119. doi: 10.1038/nn.2450
- Stephan, K. E., Iglesias, S., Heinze, J., and Diaconescu, A. O. (2015). Translational Perspectives for Computational Neuroimaging. *Neuron* 87, 716–732. doi: 10.1016/j.neuron.2015.07.008
- Stone, J. V. (2013). *Bayes' Rule: A Tutorial Introduction to Bayesian Analysis*. Sheffield: Sebtel Press.
- Stone, M. (1960). Models for choice-reaction time. *Psychometrika* 25, 251–260. doi: 10.1007/bf02289729
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18, 643–662. doi: 10.1037/h0054651
- Stuss, D. T., and Knight, R. T. (2013). *Principles of Frontal Lobe Function*. Oxford: Oxford University Press.
- Summerfield, C., and Koehlin, E. (2010). Economic value biases uncertain perceptual choices in the parietal and prefrontal cortices. *Front. Hum. Neurosci.* 4:208. doi: 10.3389/fnhum.2010.00208
- Summerfield, C., and Tsetsos, K. (2012). Building bridges between perceptual and economic decision-making: neural and computational mechanisms. *Front. Neurosci.* 6:70. doi: 10.3389/fnins.2012.00070
- Teufel, C., Subramaniam, N., Dobler, V., Perez, J., Finnemann, J., Mehta, P. R., et al. (2015). Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. *Proc. Natl. Acad. Sci. U.S.A.* 112, 13401–13406. doi: 10.1073/pnas.1503916112
- Treue, S., and Trujillo, J. C. M. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399, 575–579. doi: 10.1038/21176
- Tsetsos, K., Gao, J., McClelland, J. L., and Usher, M. (2012). Using time-varying evidence to test models of decision dynamics: bounded diffusion vs. the leaky competing accumulator model. *Front. Neurosci.* 6:79. doi: 10.3389/fnins.2012.00079
- Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science* 185, 1124–1131. doi: 10.1126/science.185.4157.1124
- Usher, M., and McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychol. Rev.* 108, 550–592. doi: 10.1037/0033-295X.108.3.550
- Usher, M., and McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychol. Rev.* 111, 757–769. doi: 10.1037/0033-295X.111.3.757
- van Bergen, R. S., Ji Ma, W., Pratte, M. S., and Jehee, J. F. (2015). Sensory uncertainty decoded from visual cortex predicts behavior. *Nat. Neurosci.* 18, 1728–1730. doi: 10.1038/nn.4150
- von Neumann, J., and Morgenstern, O. (1947). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton university press.
- Wald, A. (1973). *Sequential Analysis*. New York, NY: Courier Corporation.
- Wald, A., and Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *Ann. Math. Stat.* 37, 326–339. doi: 10.1214/aoms/1177730197
- Wang, X.-J. (2001). Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci.* 24, 455–463. doi: 10.1016/S0166-2236(00)01868-3
- Wang, X.-J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. *Neuron* 36, 955–968. doi: 10.1016/S0896-6273(02)01092-9

- Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science* 269, 1880–1882. doi: 10.1126/science.7569931
- Wong, K. F., and Wang, X. J. (2006). A recurrent network mechanism of time integration in perceptual decisions. *J. Neurosci.* 26, 1314–1328. doi: 10.1523/jneurosci.3733-05.2006
- Yang, T., and Shadlen, M. N. (2007). Probabilistic reasoning by neurons. *Nature* 447, 1075–1080. doi: 10.1038/nature05852
- Yu, A. J., and Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron* 46, 681–692. doi: 10.1016/j.neuron.2005.04.026

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Matsumori, Koike and Matsumoto. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.