

Title

Abstract

blahhh blahhh blah

Introduction

To survive in an ever-changing environment, animals must flexibly adapt their behavior based on previously encoded and novel information. This context-dependency must also be reflected in the information processing of neural networks underlying intelligent behavior. For instance, when you walk down some staircase in your fully lit basement, your brain might entirely rely on the feedforward (bottom-up) input your senses receive (Fig. 1axxx). In contrast, when you walk down the same stairs in complete darkness, your brain might rely entirely on feedback (top-down) signals generated from a staircase model it has formed over previous experiences (Fig. 1bxx).

The importance of these feedback inputs has been emphasized by observations showing that top-down projections outnumber feedforward connections (XXX) and that they modulate (XXX) or even entirely drive (XXX) neuron activity. But how do neural networks switch between a feedforward-dominated and a feedback-dominated processing mode? And how do neural networks combine both input streams wisely? For instance, if you hike down a mountain for the first time in very foggy conditions, your brain receives unreliable visual information. In addition, it can only draw on a shaky prediction about what to expect (Fig. 1cxxx).

A common hypothesis is that the brain weights different inputs according to their reliabilities which are a function of the variances of those inputs (XXX). A prominent example of this is Bayesian multisensory integration in which the information from multiple modalities is optimally combined (XXX). According to this theory, the representation of an object perceived through various senses is a linear combination of the single-modality estimates, weighted by their reliabilities. This theory has been backed up by a number of observations showing that xxx (XXX). The same concepts can be employed for the weighting of sensory feedforward inputs and predictions thereof (XXX). A central point in this weighting is the estimation of the sensory and prediction variances that underlie the reliability of each input. However, how this is implemented in the brain is still largely unknown.

We hypothesised that prediction error (PE) neurons can provide the basis for the neural computation of variances. PEs are an integral part of the theory of predictive processing that states that the brain constantly compares incoming sensory information with predictions thereof. When those predictions are wrong, the resulting PEs allow the network to update the model of the world, thereby ensuring that the predictions are more accurate (XXX). Experimental evidence suggests that these PEs may be represented in the activity of distinct groups of neurons, termed PE neurons (XXX). In brain areas with excitatory neurons that exhibit low spontaneous firing rates, these neurons may come in two types (XXX). Negative PE (nPE) neurons mainly increase their activity when the prediction is *stronger* than the sensory input, while positive PE (pPE) neurons mainly increase their activity when the prediction is *weaker* than the sensory input. Indeed, it has been shown that excitatory neurons in layer 2/3 of rodent primary sensory areas can encode negative or positive PEs (XXX).

Here, we show that xxx

Results

nPE and pPE neurons as the basis for estimating mean and variance of sensory stimuli

We hypothesise that the distinct response patterns of negative and positive prediction-error (nPE/pPE) neurons represent the backbone for estimating the mean and the variance of sensory stimuli. nPE neurons only increase their activity relative to a baseline when the sensory input is weaker than predicted, while pPE neurons only increase their activity relative to a baseline when the sensory input is stronger than predicted. Moreover, both nPE and pPE neurons remain at their baseline activity when the sensory input is fully predicted (XXX). Assuming that the prediction equals the mean of the sensory stimulus, the PE neurons, hence, encode the deviation from the mean. Thus, the squared sum of nPE and pPE activity represents the variance of the feedforward input.

To test our hypothesis, we studied a rate-based mean-field network the core of which is a prediction-error (PE) circuit with excitatory nPE and pPE neurons, as well as inhibitory parvalbumin-expressing (PV), somatostatin-expressing (SOM), and vasoactive intestinal peptide-expressing (VIP) interneurons (Fig. xxx). While the excitatory neurons are simulated as two coupled point compartments to emulate the soma and dendrites of elongated pyramidal cells, respectively, all inhibitory cell types were modeled as point neurons. The connectivity of and inputs to the network were chosen such that the excitatory (E) and inhibitory (I) pathways onto the pyramidal cells were balanced because it has been shown that this E/I balance is necessary for nPE and pPE neurons to emerge (XXX, see Methods).

In addition to this core circuit, we model a memory (M) neuron that perfectly integrates the activity of the PE neurons. In accordance with XXX, we assume that the pPE neuron excites the memory neuron, while the nPE neuron inhibits this neuron (for instance, through lateral inhibition, here not modeled explicitly). The M neuron connects to the apical dendrites of the PE neurons and some of the interneurons (here, VIP and PV neurons, see Methods for more details). In this network, the M neuron serves as a prediction that is dynamically updated when new sensory information is available. We furthermore simulate a downstream neuron (termed V neuron), modeled as a leaky integrator with a squared activation function, that receives excitatory output synapses from the PE neurons. Hence, in this setting, the V neuron encodes the variance of the sensory stimuli.

To show that this network can indeed represent mean and variance in the respective neurons, we stimulate it with a sequence of step-wise constant inputs drawn from a uniform distribution (Fig. 2aXX), assuming that the sensory stimulus varies over time. In line with the distinct response patterns for nPE and pPE neurons, these neurons change only slightly with increasing stimulus mean but increase strongly with input variance (Fig. 2bXXX). This is in contrast to the three interneurons that strongly increase with stimulus mean while they only moderately increase with stimulus variance (Fig. 2cxxx). The activity of the memory neuron M gradually approaches the mean of the sensory inputs (Fig. 2d, middle), while the activity of the V neuron approaches the variance of the inputs (Fig. 2e, middle). This is true for a wide range of input statistics (Fig. XXX) and input distributions (Fig. SXXX). Small deviations from the true mean occur mainly for larger input variances, while the estimated variance is fairly independent of the input statistics tested.

XXX validated that also correct for population network (beyond mean-field network) XXX XXX assumptions (BL, gain-nPE = gain-pPE = 1) and how important really (see above)

XXX Summary paragraph XXX

Estimating variances of sensory inputs and predictions requires a hierarchy of PE circuits

Following the ideas of Bayesian multisensory integration (XXX), the weighting of sensory stimuli and predictions thereof would require knowledge of their variances. As we have shown in the previous section, the variance of the sensory stimulus can be estimated by means of PE neurons. We hypothesise that the same principles apply to estimating the variance of the prediction. Hence, we augment the network with a *higher* PE circuit that receives output synapses from the M neuron of the *lower* PE circuit (Fig. 1xxx). Both subnetworks are modeled the same, except that the M neuron in the higher PE circuit evolves more slowly than the M neuron in the lower PE circuit.

To test the network's ability to estimate the variances correctly, we stimulated the network with a sequence of inputs. In each trial one stimulus is shown to the network. To account for the stimulus variance, each stimulus is composed of n constant values drawn from a normal distribution with mean μ_{stimulus} and variance $\sigma_{\text{stimulus}}^2$, and presented one after the other. To account for potential changes in the environment, in each trial, we draw μ_{stimulus} from a uniform distribution (Fig. 2a). Hence, the inputs change on two different time scales, with stimulus variability (faster time scale) and trial variability (slower time scale).

As expected, the neurons' activity increase for both stimulus and trial variances (Fig. 2b). While the neurons in the lower PE circuit increase more strongly with stimulus variability, the neurons in the higher PE circuit increase more strongly with trial variability, indicating that the different subnetworks process different aspects of the inputs. We first consider two limit cases. In the first limit case, a different but low-variance stimulus is presented in each trial (Fig. 2c, left). In line with the ideas of multisensory integration (XXX), the network should therefore follow the sensory inputs closely and ignore the predictions. When we arithmetically calculate the weighted output (Fig. 2c, middle) based on the feedforward and feedback inputs, and the sensory weight (Fig. 2c, right), the network correctly represents mostly the sensory input (for more details, see Methods). In the second limit case, the same but high-

variance stimulus is presented in each trial (Fig. 2d, left). According to theory (XXX), the network should downscale the sensory feedforward inputs and weight the prediction more strongly. Indeed, the weighted output of the network shows a clear tendency to the mean of the stimuli (Fig. 2d, middle), also reflected in the low sensory weight (Fig. 2d, right).

In a next step, to validate the network responses more broadly, we systematically varied the trial and stimulus variability independently. If both variances are similar, the sensory weight approaches 0.5, reflecting equal contribution of sensory inputs and predictions to the weighted output. Only if both variances are zero, the network represents the sensory input perfectly. In line with the limit case examples above, if the stimulus variance is larger than the trial variance, the network weighs the prediction more strongly than the sensory input. This is reversed if the stimulus variance is smaller than the trial variance (Fig. 2e). Because the network dynamically estimates the mean and variances of the sensory inputs and the prediction, the weighted output and the sensory weight changes accordingly when the input statistics changes (Fig. SX).

The first limit case (Fig. 2 c) shows that even in a sensory-driven input regime, the prediction is weighted more at the beginning of a new trial than in the steady state. This is further confirmed in simulations in which the trial duration was shortened. For those simulations, the prediction even outweighs the sensory input, reflected in a very low sensory weight (Fig. 2e). This suggests that predictions influence neural activity more significantly in experiments that rely on very fast stimulus changes.

For some psychiatric disorders, it has been shown that the weighting of sensory inputs and predictions thereof is impaired (XXX), leading to an overweighting of one of these signals. We thus wondered which network properties might bias the estimation of the variances, and, consequently, the weighting of different input streams. In our network, the M neuron evolves faster in the lower PE circuit than in the higher PE circuit. We identified the speed at which the M neurons are updated in the different PE circuits as a decisive network factor. To show this, we varied the weights from the PE neurons onto the M neurons in either PE circuit. When xxxx, the network shows a bias for predictions, while when xxx, the network shows a bias for the sensory stimuli (Fig. SXX). While the speed at which the activity of the M neurons evolve in relation to each other may underlie pathological weighting, the precise activation function of the V neurons is of less importance. When we replaced the squared activation function with a purely linear function, the V neurons do not encode the variance but the averaged absolute deviation of the sensory stimuli. However, the sensory weight is only slightly shifted to larger values for low trial/high stimulus variability (Fig. SXXX).

XXX Summary XXX

Biasing the weighting of sensory inputs and predictions by neuromodulators acting on PE circuits

The brain's flexibility and adaptability are not least due to the fact that a plethora of neuromodulators influence the activity of neurons in a variety of ways (XXX). A prominent target of neuromodulatory inputs are inhibitory neurons (Cardin 2019, XXX). Moreover, distinct interneuron types are differently (in-)activated by those neuromodulators. For instance, it has been shown that xxx. We therefore wondered if and how the weighting of sensory inputs and predictions thereof may be biased when neuromodulators activate distinct interneuron types.

To this end, we modeled the presence of neuromodulators by injecting an additional, excitatory input into one or two interneuron types. We reasoned that the network effect of a neuromodulator not only depends on the interneuron type it targets but also on the inputs this neuron receives and the connections it makes with other neurons in the network. We, therefore, tested three different mean-field networks that differ with respect to the distribution of sensory inputs and predictions onto the interneurons and the underlying connectivity. The commonality across those networks is that all of them exhibit an E/I balance of excitatory and inhibitory pathways onto the PE neurons (XXX). Across the different mean-field networks tested, activating a SOM or VIP neuron individually forces the networks to weigh both inputs more equally. As a consequence, in a sensory-driven input regime, predictions are overrated. Similarly, in a prediction-driven regime, sensory inputs are overrated. Interestingly, when both interneuron types are activated to the same degree, this effect disappears (Fig. 5axx, left). In contrast, stimulating PV neurons largely biases the network output towards predictions. This is even more pronounced when PV and SOM, or PV and VIP neurons are activated simultaneously (Fig. 5a, middle and right).

In the previous simulations, we assumed that a neuromodulator acts globally, that is, on the interneurons ■

in both the lower and the higher PE circuit. While this is in line with experimental data showing that xxx (XXX), we note that neuromodulators may also act more locally. We hence tested how the weighting of sensory stimuli and predictions may change when a neuromodulator activates specific interneurons in the lower or higher PE circuit only. The effect of stimulating an interneuron type in the lower PE circuit on the sensory weight is mostly the opposite of activating the same interneuron in the higher PE circuit (Fig. SXXX). For instance, the sensory inputs are overrated when VIP neurons are activated in the higher PE circuit, while the prediction is overrated when VIP neurons in the lower PE circuit are stimulated. When VIP and SOM neurons are stimulated equally, the sensory weight remains unchanged, independently of which PE circuit is targeted by the neuromodulator.

What are the mechanisms that give rise to these effects? And how do the combined local changes give rise to the global one observed in our network simulations? The sensory weight is chosen to be a function of the V neurons of the lower and higher PE circuit. Hence, any changes to the sensory weight are a consequence of changes to the neurons encoding the variance (Fig. 5xx). In our network, the V neurons only receive excitatory output synapses from both nPE and pPE neurons. Hence, any changes in the sensory weights upon activation of interneurons must be due to changes in the PE neurons. To disentangle the effect of nPE and pPE neurons, we perturbed those neurons individually in both the lower or higher subnetwork by injecting either an inhibitory or excitatory additional input (Fig. 5cxxx). Stimulating either PE neuron in the lower subnetwork increases the activity of the lower V neuron strongly. Moreover, the higher V neuron is also slightly affected. At first, this is counterintuitive because the V neuron in the higher subnetwork does not receive direct synapses from the PE neurons in the lower subnetwork. However, the activity of the M neuron encoding the prediction increases with an excitatory input onto the pPE neuron and decreases when the same input targets the nPE neuron (the opposite is true for an inhibitory input). Because neurons in the higher PE circuit receive synapses from the M neuron of the lower subnetwork, the estimation of the variance in the higher subnetwork is also affected. In contrast, stimulating either PE neuron in the higher subnetwork increases the activity of the higher V neuron but leaves the M and V neurons in the lower subnetwork unaffected (Fig. 5cxxx).

The stimulation of the PE neurons may cause both an increase in the baseline activity and a change in the neuron's gain. To disentangle the effect of elevated baseline activity and gain change, we illustrate each contribution separately by means of a mathematically tractable toy model (see Methods for more details). The variance estimated in the lower subnetwork increases for both an increase in the baseline and the gain of the PE neurons in the lower subnetwork. In contrast, when the gain of those PE neurons decreases, so does the variance. As we have seen before, changes in the baseline and the gain of PE neurons in the higher PE circuit do not affect the lower-level variance estimation. The variance estimated in the higher subnetwork changes equally for baseline increase or gain change of the PE neurons in the same subnetwork. However, in line with the perturbation experiments above, the higher-level variance estimation is affected by changes to the baseline or gain of the lower-level PE neurons as a result of the biased prediction. The variance increases slightly with increasing baseline or increasing deviation from a gain equal to 1 (Fig 5dxxx). Moreover, the mean of the sensory stimuli is overpredicted when the baseline or the gain for the pPE neuron increases, or when the gain of the nPE neuron decreases. In contrast, the mean of the sensory stimuli is underpredicted when the baseline or the gain of the nPE neuron increases, or when the gain of the pPE neuron decreases. In summary, any change in interneurons that cause an additional inhibitory input to the PE neurons reduces the gain of those neurons. Changes in the inhibitory neurons that cause an additional excitatory input to the PE neurons increases the baseline and potentially the gain of those neurons.

This suggests that to understand the effect of neuromodulators on the sensory weight, we need to investigate the effect of interneuron activation on baseline and gain of PE neurons. To this end, we stimulated either PV, SOM, or VIP neurons independently for all three mean-field networks and measured the changes in baseline and gain of both PE neuron types (Fig. 5exx). In all three networks tested, activating PV neurons decreases both quantities, leading to a decrease in the variance estimated (Fig. 5exx & Fig. 5Sxxx). Activating SOM or VIP neurons decreases the overall gain but increases the baseline activity of the PE neurons. Whether and how much the gain of the nPE or pPE neuron is reduced depends on the inputs onto SOM/VIP neuron and the connectivity to the other neurons in the network. Similarly, how much the baseline is elevated depends on the specifics of the mean-field network (Fig. 5exx & Fig. 5Sxxx). Hence, whether stimulating SOM or VIP neurons decreases or increases the activity of the V neurons depend on the input statistics. For low-mean stimuli, the effect of the elevated baseline dominates, while for high-mean stimuli the effect of the gain changes dominates the combined effect.

XXX Summary ... effect of neuromodulators acting on INs on sensory weight best understood as

combined effect on the baseline and gain of nPE and pPE neurons that in turn determine the activity of the V neuron. Those effects not only depend on the target IN but also the network in which it is embedded, that is the connectivity and the inputs onto INs. Moreover, it depends on whether local or global. The effects of activation on lower is mainly the opposite of same in higher. However, the effects don't cancel each other in global modulation. Mostly because 1) V higher additionally changes due to changes in prediction but also 2) effects in higher are stronger !? and 3) $1/(1+(V_s + D)/(V_p + D))$... that is it cannot cancel per se. Even though the underlying reasons for a change in the V neuron depends on the the network characteristics, there are some consistent changes in sensory weight across MFN ... name

The contraction bias as a result of the weighting of sensory inputs and predictions thereof

The weighting of sensory inputs and predictions thereof manifests in all-day behavior, in the form of a phenomenon called *contraction bias*. The contraction bias describes the tendency to overestimate sensory stimuli that are at the lower end of a stimulus distribution and to underestimate stimuli that are at the upper end of the same distribution (Fig. 5aXX). This *bias towards the mean* has been reported in different species and modalities (XXX).

The weighted output of our network can be interpreted as a neural manifestation of the contraction bias (see Methods for a thorough analysis). The bias increases with stimulus variance (Fig. 5bXX), decreasing the slope of the linear fit modeling the relationship between the true and estimated stimuli (Fig. 5 b, right XXX, compare with Fig. 5a). In contrast, the bias decreases with trial variance, so that the slope of the linear fit approaches 1 (Fig. 5b rightXXX).

What are the underlying network factors that contribute to this phenomenon? To disentangle the potentially different sources of the bias, we first simulated a network without stimulus variability (variance set to zero) for two different trial variabilities. In this case, a contraction bias emerges but is independent of the volatility of the environment (Fig. 5c, leftXXX). We show mathematically that the bias is a result of the recent stimulus history and vanishes if the trial duration approaches infinity (see Methods for more details, and Fig. 5d). We next resume the limit case in which the same but high-variance stimulus is shown in every trial. In this case, the weighted output exhibits a contraction bias that is largely independent of the exact stimulus variances tested. As shown mathematically (see Methods), the bias is a result of xxx and approaches xxx (Fig. 5dxxx).

So far, we assumed that the stimulus variance is independent of the trial mean. A consequence of this choice is that the bias on either end of the stimulus distribution is the same (but with reversed signs). However, behavioral (neural?) data (XXX) shows that the bias increases for stimuli drawn from the upper end of the distribution, a phenomenon usually attributed to *scalar variability*. To capture this in the model, we assume that the stimulus standard deviation linearly increases with the trial mean. In these simulations, as expected, the bias increases for a stimulus distribution shifted to higher trial means (Fig. 5eXXX).

XXX Summary XXX

Discussion

We solved the brain.

Models and methods

Network model

Network consists of two subnetworks. Each subnetwork consists of a PE circuit, a memory neuron and a neuron representing the variance. XXX The memory neuron of subnetwork feedforwardly connects to the PE circuit of the second subnetwork.

Prediction-error network model

Consider a mean field network in which each population is represented by one representative neuron. The mean-field PE network consists of an excitatory nPE and pPE neuron, as well as two inhibitory PV neurons (one receiving S, the other P), as well as inhibitory SOM and VIP neurons.

Each excitatory pyramidal cell (that is, nPE or pPE neuron) is divided into two coupled compartments, representing the soma and the dendrites, respectively. The dynamics of the firing rate r_E of the somatic compartment obeys (?)

$$\tau_E \frac{dr_E}{dt} = -r_E + w_{ED} \cdot r_D - w_{EP} \cdot r_P + I_E, \quad (1)$$

where τ_E denotes the excitatory rate time constant ($\tau_E=60$ ms), the weight w_{ED} describes the connection strength between the dendritic compartment and the soma of the same neuron, and w_{EP} denotes the strength of somatic inhibition from PV neurons. The overall input I_E comprises external background and feedforward sensory inputs (see “Inputs” below). Firing rates are rectified to ensure positivity.

The dynamics of the activity r_D of the dendritic compartment obeys (?)

$$\tau_E \frac{dr_D}{dt} = -r_D + w_{DE} \cdot r_E - w_{DS} \cdot r_S + I_D, \quad (2)$$

where the weight w_{DE} denotes the recurrent excitatory connections between PCs, including backpropagating activity from the soma to the dendrites. w_{DS} represents the strength of dendritic inhibition from SOM neurons. The overall input I_D comprises fixed, external background inputs and feedback predictions (see “Inputs” below). We assume that any excess of inhibition in a dendrite does not affect the soma, that is, the dendritic compartment is rectified at zero.

Just as for the excitatory neurons, the firing rate dynamics of each interneuron is modeled by a rectified, linear differential equation (?),

$$\tau_I \frac{dr_X}{dt} = -r_X + I_X + w_{XE} \cdot r_E - w_{XP} \cdot r_P - w_{XS} \cdot r_S - w_{XV} \cdot r_V, \quad (3)$$

where r_X denotes the firing rate of neuron type X , and the weight matrices w_{XY} denote the strength of connection between the presynaptic neuron population Y and the postsynaptic neuron population X ($X, Y \in \{P, S, V\}$). The rate time constant τ_I was chosen to resemble a fast GABA_A time constant, and set to 2 ms for all interneuron types included. The overall input I_X comprises fixed, external background inputs, as well as feedforward sensory inputs and feedback predictions (see “Inputs” below).

Memory and variance neuron

$$\tau_m \cdot \frac{dr_M}{dt} = w_{M \leftarrow pPE} \cdot r_{pPE} - w_{M \leftarrow nPE} \cdot r_{nPE} \quad (4)$$

$$\tau_v \cdot \frac{dr_V}{dt} = -r_V + (w_{V \leftarrow pPE} \cdot r_{pPE} - w_{V \leftarrow nPE} \cdot r_{nPE})^2 \quad (5)$$

Weighted output

$$r_{out} = \alpha \cdot S + (1 - \alpha) \cdot P \quad (6)$$

$$\begin{aligned} \alpha &= \frac{1/r_{V1}}{1/r_{V1} + 1/r_{V2}} \\ &= \left(1 + \frac{r_{V1}}{r_{V2}} \right)^{-1} \end{aligned} \quad (7)$$

Connectivity

Inputs

Simulations

Acknowledgments

Supplementary Information

Sensory weight and contraction bias

If P is rather constant, the slope in the contraction bias is exactly the sensory weight

$$r_{\text{out}} = \alpha_S \cdot S + (1 - \alpha_S) \cdot P \equiv m \cdot S + n$$

However, P is usually/normally a function of S . For simplicity, let's assume that P decays exponentially to a new value of S :

$$P = P_0 \cdot e^{-t/\tau} + f(S) \cdot (1 - e^{-t/\tau})$$

Within each trial with trial duration T , P can be expressed by n sections of length t in which the stimulus is constant and, for the sake of simplicity, drawn from a uniform distribution $U(s - \frac{\sigma_S}{12}, s + \frac{\sigma_S}{12})$. P_0 is drawn from $U(\mu - \frac{\sigma_P}{12}, \mu + \frac{\sigma_P}{12})$. P_n is then given by

$$P_n = P_0 \cdot e^{-t/\tau} + (1 - e^{-t/\tau}) \sum_{i=1}^n s_i \cdot e^{-(n-i) \cdot t/\tau}$$

This needs to be averaged over all possible states

$$P_n = e^{-t/\tau} \int_{\mu - \frac{\sigma_P}{12}}^{\mu + \frac{\sigma_P}{12}} P_0 f(P_0) dP_0 + (1 - e^{-t/\tau}) \sum_{i=1}^n \cdot e^{-(n-i) \cdot t/\tau} \int_{s - \frac{\sigma_S}{12}}^{s + \frac{\sigma_S}{12}} s f(s) ds$$

This gives

$$P_n = \mu \cdot e^{-T/\tau} + (1 - e^{-T/\tau}) \sum_{i=1}^n e^{-(n-i) \cdot t/\tau} \cdot S$$

By making use of the geometric series, this simplifies to

$$P_n = \mu \cdot e^{-T/\tau} + (1 - e^{-T/\tau}) \cdot S$$

Together, this yields for the weighted output

$$r_{\text{out}} = \left[\alpha_S e^{-T/\tau} + (1 - e^{-T/\tau}) \right] \cdot S + (1 - \alpha_S) e^{-T/\tau} \mu$$

Hence, the slope is a function of both the sensory weight and the trial duration.

In a prediction-driven input regime ($\alpha_S \sim 0$), the slope is independent of the sensory weight and only determined by the trial duration, $m \sim (1 - e^{-T/\tau})$. In a sensory-driven input regime ($\alpha_S \sim 1$), the contraction bias vanishes ($m \sim 1$).

If the trail duration is short ($T \rightarrow 0$), the slope is given by the sensory weight. If the trail duration approaches infinity, the slope would be 1 again (however, this seems rather unrealistic, this would only be true in an ideal system without memory decay or reproduction and accumulation noise ...).

Impact of PE neurons' gain on estimating mean and variance

Only if the the gain of the nPE neuron (g_{nPE}) equals the gain of the pPE neuron (g_{pPE}) in the mean-field network, the activity of the M neuron represents the mean of the inputs,

$$\begin{aligned} g_{pPE} \langle \text{nPE} \rangle &= g_{nPE} \langle \text{pPE} \rangle \\ g_{pPE} \langle [S - P]_+ \rangle &= g_{nPE} \langle [P - S]_+ \rangle \\ g_{pPE} \int_P^b (x - P) f(x) dx &= g_{nPE} \int_a^P (P - x) f(x) dx. \end{aligned} \tag{8}$$

In case of a uniform distribution ($f(x) = 1/(b-a)$ when $x \in [a, b]$ and 0 otherwise) from which the input values are drawn, this condition yields

$$P = \int_a^b x f(x) dx = \frac{a+b}{2} \quad (9)$$

for $g_{nPE} = g_{pPE} = g$, and

$$g_{pPE} \left[\frac{1}{2} (b^2 - P^2) - P(b-P) \right] = g_{nPE} \left[P(P-a) - \frac{1}{2} (P^2 - a^2) \right] \quad (10)$$

for $g_{nPE} \neq g_{pPE}$, which can be further summaries by

$$P = \frac{g_{pPE} b - g_{nPE} a \pm \sqrt{g_{nPE} g_{pPE}}(a-b)}{g_{pPE} - g_{nPE}}. \quad (11)$$

Hence, estimating the mean correctly requires $g_{nPE} = g_{pPE} = g$. For the V neuron to represent the variance of the inputs, this condition must be tightened to $g_{nPE} = g_{pPE} = 1$. The variance is given by

$$\begin{aligned} V &= \langle (S-P)^2 \rangle \\ &= g_{pPE} \langle [S-P]_+^2 \rangle + g_{nPE} \langle [P-S]_+^2 \rangle. \end{aligned} \quad (12)$$

In case of a uniform distribution from which the input values are drawn, this condition yields

$$\begin{aligned} V &= \frac{g_{pPE}}{b-a} \int_P^b (u-P)^2 du + \frac{g_{nPE}}{b-a} \int_a^P (P-u)^2 du \\ &= \frac{g_{pPE}}{3} \cdot \frac{(b-P)^3}{b-a} + \frac{g_{nPE}}{3} \cdot \frac{(P-a)^3}{b-a}. \end{aligned} \quad (13)$$

Only if $g_{nPE} = g_{pPE} = 1$ and $P = \frac{a+b}{2}$, the variance is given by $\frac{(b-a)^2}{12}$, otherwise the V neuron's activity is given by

$$V = \frac{(b-a)^2}{3 (g_{pPE} - g_{nPE})^3} \cdot [g_{nPE} \cdot (g_{pPE} \mp \sqrt{g_{nPE} g_{pPE}})^3 - g_{pPE} \cdot (g_{nPE} \mp \sqrt{g_{nPE} g_{pPE}})^3]. \quad (14)$$

Impact of PE neurons' baseline on estimating mean and variance

Only if the the baseline of the nPE neuron (n_0) equals the baseline of the pPE neuron (p_0) in the mean-field network, the activity of the M neuron represents the mean of the inputs,

$$\begin{aligned} \langle pPE \rangle &= \langle nPE \rangle \\ \langle [S-P]_+ + p_0 \rangle &= \langle [P-S]_+ + n_0 \rangle \\ \int_P^b (x-P) f(x) dx + p_0 \underbrace{\int_a^b f(x) dx}_{=1} &= \int_a^P (P-x) f(x) dx + n_0 \underbrace{\int_a^b f(x) dx}_{=1}. \end{aligned} \quad (15)$$

In case of a uniform distribution ($f(x) = 1/(b-a)$ when $x \in [a, b]$ and 0 otherwise) from which the input values are drawn, this condition yields

$$P = \frac{b+a}{2} + \frac{p_0 - n_0}{b-a}. \quad (16)$$

Hence, if $p_0 = n_0$, the mean can be estimated correctly. For the V neuron to represent the variance of the inputs, this condition must be tightened to $p_0 = n_0 = 0$. The variance is given by

$$\begin{aligned} V &= \langle (pPE + nPE)^2 \rangle \\ &= \langle [S-P]_+^2 \rangle + \langle [P-S]_+^2 \rangle + (p_0 + n_0)^2 + 2 (p_0 + n_0) (\langle [S-P]_+ \rangle + \langle [P-S]_+ \rangle) \end{aligned} \quad (17)$$

In case of a uniform distribution from which the inputs to the mean-field network are drawn, this expression simplifies to

$$V = \frac{1}{3(b-a)} [(b-P)^3 + (P-a)^3] + (p_0 + n_0)^2 + \frac{(p_0 + n_0)}{b-a} [(b-P)^2 + (a-P)^2]. \quad (18)$$

Inserting the expression for P (by itself modulated by the baseline activities of the PE neurons) yields

$$V = \frac{1}{3(b-a)} \left[\left(\frac{b-a}{2} - \frac{p_0 - n_0}{b-a} \right)^3 + \left(\frac{b-a}{2} + \frac{p_0 - n_0}{b-a} \right)^3 \right] + (p_0 + n_0)^2 + \frac{(p_0 + n_0)}{b-a} \left[\left(\frac{b-a}{2} + \frac{p_0 - n_0}{b-a} \right)^2 + \left(\frac{b-a}{2} - \frac{p_0 - n_0}{b-a} \right)^2 \right] \quad (19)$$

Simplifying the expression, leads to

$$V = \frac{(b-a)^2}{12} + \frac{(p_0 - n_0)^2}{(b-a)^2} \left(1 + 2 \frac{p_0 + n_0}{b-a} \right) + (p_0 + n_0) \left(p_0 + n_0 + \frac{b-a}{2} \right) \quad (20)$$

Modelling the impact of neuromodulators on the weighting of sensory inputs and predictions thereof

Assume that neuromodulator act through activating INs. INs on the other hand modulate both gain and BL of nPE and pPE neurons Changes in gain and BL will affect the overall activity and the balance between nPE and pPE neurons =; this will affect both the prediction and the variance.

Modulate PE in lower: variance in lower is combination of direct changes in PE and indirect through prediction (that is changed through PE mod) variance in higher is just because of differences in prediction

Modulation in higher: variance in lower should be unaffected variance in higher is combination of direct changes in PE and indirect through prediction (that is changed through PE mod)

Variance as function of bias in mean:

$$\begin{aligned} V &= \frac{1}{n} \sum_i (x_i - (\mu \pm \delta\mu))^2 \\ &= \frac{1}{n} \sum_i \{(x_i - \mu)^2 + \delta\mu^2 \mp 2\delta\mu(x_i - \mu)\} \\ &= V_{\text{unmod}} + \delta\mu^2 \mp 2\delta\mu \left(\frac{1}{n} \sum_i x_i - \mu \right) \\ &= V_{\text{unmod}} + \delta\mu^2 \end{aligned}$$

Influence of a population of nPE and pPE neurons

XXX

Analysis of simplified network model, effect of time constants

simplified model: dynamics and steady state of rM and rV, rM and rV as a function of time constants and trial duration etc., weighting, then use those expressions to discuss when weighting goes awry and how long transitions take from one state to another ...

Comparison to Kalman filter and Bayes Factor surprise

Kalman filter. Initialisation

$$\begin{aligned} x_{0|init} &= 0 \\ P_{0|init} &= \sigma^2 I \end{aligned}$$

with x being the system state (in my terms the prediction), P is the covariance matrix of the errors of x (in my terms the var of the predictions) and I is the identity matrix. Then the "correction" is given by

$$\begin{aligned} K_k &= P_{k|k-1} H_k^T (H_k P_{k|k-1} H_k^T + R_k)^{-1} \\ x_k &= x_{k|k-1} + K_k (z_k - H_k x_{k|k-1}) \\ P_k &= (I - K_k H_k) P_{k|k-1} \end{aligned}$$

with K the kalman gain matrix, H the observation matrix ($z_k = H_k x_k + noise$), R the covaraince of the measurement noise and z a new observation. The last part of the Kalman filter is the "prediction":

$$\begin{aligned} x_{k|k-1} &= F_{k-1} x_{k-1} + B_{k-1} u_{k-1} \\ P_{k|k-1} &= F_{k-1} P_{k-1} F_{k-1}^T + Q_{k-1} \end{aligned}$$

with F the transition matrix ($x_{k|k-1} = F_{k-1} x_{k-1}$), u a deterministic perturbation, B the dynamics of the deterministic perturbation. In our terms

$$\alpha = K_k = \frac{P_{k|k-1}}{R_k + P_{k|k-1}}$$

$P_{k|k-1}$, is however σ_P^2 in my implementation and R_k is fixed variance of inputs σ_S^2 . Hence, my implementation represents (?) the Kalman filter. Important to note is, that in my implementation we estimate the variance of inputs dynamically, so it is not set! Another nice advantage here is that I don't need a good estimate for P . I can basically initiate it as I want. Another difference is that I consider the optimal weighting in my "output neuron" and not the prediction itself

XXX Comparison to Bayes Factor surprise