Title

Abstract

blahhh blahhh blah

Introduction

XXX Behavior of intelligent life is context-dependent and requires a constant adaptation to the ever-changing life. This must also be reflected in the underlying neural networks and the inputs that drive those. For instance, when you walk down your your basement stairs and the lighting is well, your brain might rely entirely on the feedforward sensory input your senses receive to take decisions and adapt your body in the environment/world (Fig 1a). However, when you walk down the same basement stairs you walked million of times before but suddenly the light bulb is broken and the visual input is absent, you don't freeze forever but walk the stairs almost as confidently as before. Your brain now relies on the prediction it had formed over the last years. (Fig. 1b). The former example resonates with the feedforward-centric view that has dominated the neuroscience literature for decades. However, in recent years, the importance of internally generated, feedback inputs has increased significantly. Not only because it was shown that top-down projects xxx outsize feedforward connections (ref) but also because these connection modulate (ref) or even drive (ref) neuron dynamics. These feedback projects have been hypothesized to carry expectations, predictions ... and may help us to navigate (in the physical and metaphysical sense) the world even in the absence of sensory input. But how do our brains decide whether it follows more the sensory feedforward input or the feedback predictions?

For instance, if you hike down a mountain for the first time under very foggy conditions, you brain may receive unreliable sensory input but at the same time can only have a very rough prediction about what to expect. In this case, your brain might be better off combining both the sensory input and the prediction thereof (Fig 1c). A common hypothesis is that the brain weights those signals according to their reliabilties, that is, a function of the variances of those signals (ref). These ideas come from Bayesian integration in which it is believed that xxx (ref). A prominent example of this is multisensory integration in which xxx (ref). Similarly, the brain must have mechanisms to estimate the variances of both the sensory input and the predictions thereof that allow it to weight those signals accordingly and context-dependent. But how this is implemented in the brain is still largely unknown.

Here, we hypothesis that xxx. We show that xxx

Results

nPE and pPE neurons as the basis for computing 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

Weighting external and internal signals requires two sets of PE neurons

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 $\sigma^2_{\text{stimulus}}$, 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 trail 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. Moreover, it has been hypothesised that factors like stress or cognitive load may also influence the processing of feedforward and feedback inputs (XXX). Naturally, we were wondering which factors might influence 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. When we xxxx, the network shows a bias for predictions, while when we xxx, the network shows a bias for the sensory stimuli (Fig. SXX). This indicates that a non-pathological weighting requires xxx. Another factor that may contribute to a distorted weighting is the baseline activity of PE neurons that was set to zero in our model, in line with very low spontaneous firing rates of excitatory neurons in rodent primary sensory areas (XXX). Increasing this baseline activity for the nPE neuron (Fig. SXXX), pPE neuron (Fig. SXXX) or both pushes the network to weight sensory stimuli and predictions more equally. We speculate that an increase of baseline activity may be a natural result of an increased cognitive load or stress.

XXX Summary XXX

XXX IN influence, neuromodulators ... mechanisms XXX

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 were therefore wondering 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, activating 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 activating 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 activated. 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?

XXX contraction bias

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 feefdowradly 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 of obeys (**?**)

$$\tau_E \, \frac{dr_{\mathrm{E}}}{dt} = -r_{\mathrm{E}} + w_{\mathrm{ED}} \cdot r_{\mathrm{D}} - w_{\mathrm{EP}} \cdot r_{\mathrm{P}} + I_{\mathrm{E}}, \tag{1}$$

where τ_{E} denotes the excitatory rate time constant (τ_{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 (**?**)

$$\tau_E \frac{dr_{\mathrm{D}}}{dt} = -r_{\mathrm{D}} + w_{\mathrm{DE}} \cdot r_{\mathrm{E}} - w_{\mathrm{DS}} \cdot r_{\mathrm{S}} + I_{\mathrm{D}}, \tag{2}$$

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 (**?**),

$$\tau_I \frac{dr_{\mathrm{X}}}{dt} = - r_{\mathrm{X}} + I_{\mathrm{X}} + w_{\mathrm{XE}} \cdot r_{\mathrm{E}} - w_{\mathrm{XP}} \cdot r_{\mathrm{P}} - w_{\mathrm{XS}} \cdot r_{\mathrm{S}} - w_{\mathrm{XV}} \cdot r_{\mathrm{V}}, \tag{3}$$

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 $\mathrm{GABA_A}$ 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

$$\tau_m \cdot \frac{dr_{\mathrm{M}}}{dt} = w_{\mathrm{M}\leftarrow\mathrm{pPE}} \cdot r_{\mathrm{pPE}} - w_{\mathrm{M}\leftarrow\mathrm{nPE}} \cdot r_{\mathrm{nPE}} \tag{4}$$

$$\tau_v \cdot \frac{dr_{\mathrm{V}}}{dt} = -r_{\mathrm{V}} + (w_{\mathrm{V}\leftarrow\mathrm{pPE}} \cdot r_{\mathrm{pPE}} - w_{\mathrm{V}\leftarrow\mathrm{nPE}} \cdot r_{\mathrm{nPE}})^2 \tag{5}$$

Weighted output

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

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

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

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

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

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

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 $U\left(s - \frac{\sigma_{\text{S}}}{12}, s + \frac{\sigma_{\text{S}}}{12}\right)$. P_0 is drawn from $U\left(\mu - \frac{\sigma_{\text{P}}}{12}, \mu + \frac{\sigma_{\text{P}}}{12}\right)$. P_{n} is then given by

$$P_{\text{n}} = P_0 \cdot \mathrm{e}^{-t/\tau} + \left(1 - \mathrm{e}^{-t/\tau}\right) \sum_{i=1}^{n} s_i \cdot \mathrm{e}^{-(n-1)\cdot t/\tau}$$

This needs to be averaged over all possible states

$$P_{\text{n}} = \mathrm{e}^{-t/\tau} \int_{\mu - \frac{\sigma_{\text{P}}}{12}}^{\mu + \frac{\sigma_{\text{P}}}{12}} P_0 \; f(P_0) \; dP_0 + \left(1 - \mathrm{e}^{-t/\tau}\right) \sum_{i=1}^{n} \cdot \mathrm{e}^{-(n-1)\cdot t/\tau} \int_{s - \frac{\sigma_{\text{S}}}{12}}^{s + \frac{\sigma_{\text{S}}}{12}} s \; f(s) \; ds$$

This gives

$$P_{\text{n}} = \mu \cdot \mathrm{e}^{-T/\tau} + \left(1 - \mathrm{e}^{-t/\tau}\right) \sum_{i=1}^{n} \mathrm{e}^{-(n-i)\cdot t/\tau} \cdot S$$

By making use of the geometric series, this simplifies to

$$P_{\text{n}} = \mu \cdot \mathrm{e}^{-T/\tau} + \left(1 - \mathrm{e}^{-T/\tau}\right) \cdot S$$

Together, this yields for the weighted output

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

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

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

If the trail duration is short ($T \to 0$), 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

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

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

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

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

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

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

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

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

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

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

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

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,

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

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

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

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

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

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

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

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

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

Simplifying the expression, leads to

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

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:

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

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

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

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