

# Message Passing in the Hierarchical Gaussian Filter

October 23, 2018

## 1 Computations of Nodes in the HGF

In the HGF, each node has to perform several computations within an experimental trial, which can be ordered in time as shown in the box:

NODE  $i$  AT TRIAL  $k$

(compute prediction $_i^{(k)}$ )  
 $\leftarrow$  receive PE $_{i-1}^{(k)}$  from node $_{i-1}$

UPDATE step  
compute posterior $_i^{(k)}$   
*given:* PE $_{i-1}^{(k)}$  and prediction $_i^{(k)}$   
 $\rightarrow$  send posterior $_i^{(k)}$  to node $_{i-1}$

PE step  
compute PE $_i^{(k)}$   
*given:* prediction $_i^{(k)}$  and posterior $_i^{(k)}$   
 $\rightarrow$  send PE $_i^{(k)}$  to node $_{i+1}$   
 $\leftarrow$  receive posterior $_{i+1}^{(k)}$  from node $_{i+1}$

PREDICTION step  
compute prediction $_i^{(k+1)}$   
*given:* posterior $_i^{(k)}$  and posterior $_{i+1}^{(k)}$

The exact computations in each step depend on the nature of the coupling (via VAPes vs. VOPEs) with the parent and children nodes and will be

outlined in the following two chapters.

Note that we have placed the **PREDICTION step** in the end of a trial. This is because usually, we think about the beginning of a trial as starting with receiving a new input, and of a prediction as being present before that input is received. However, in some variants of the HGF the prediction also depends on the time that has passed in between trials, which is something that can only be evaluated once the new input arrives - hence the additional computation of the (current) prediction in the beginning of the trial. Conceptually, it makes most sense to think of the prediction as happening continuously between trials. For implementational purposes, it is however most convenient to only compute the prediction once the new input (and with it its arrival time) enters. This ensures both that the posterior means of parent nodes have had enough time to be sent back to their children for preparation for the new input, and that the arrival time of the new input can be taken into account appropriately.

## 2 Computations for VAPE coupling

The exact computations of the **UPDATE step** depend on the nature of the coupling with the child node(s), while both the **PE step** and the **PREDICTION step** depend on the coupling with the parent node(s).

### 2.1 Update Step

If Node  $i$  is the value parent of Node  $i - 1$ , then the following update equations apply to Node  $i$ :

$$\begin{aligned}\pi_i^{(k)} &= \hat{\pi}_i^{(k)} + \alpha_{i-1,i}^2 \hat{\pi}_{i-1}^{(k)} \\ \mu_i^{(k)} &= \hat{\mu}_i^{(k)} + \frac{\alpha_{i-1,i}^2 \hat{\pi}_{i-1}^{(k)}}{\alpha_{i-1,i}^2 \hat{\pi}_{i-1}^{(k)} + \hat{\pi}_i^{(k)}} \delta_{i-1}^{(k)}\end{aligned}$$

We note here that we can let the update of the precision happen first, and therefore use it for the update of the mean:

$$\pi_i^{(k)} = \hat{\pi}_i^{(k)} + \alpha_{i-1,i}^2 \hat{\pi}_{i-1}^{(k)} \tag{1}$$

$$\mu_i^{(k)} = \hat{\mu}_i^{(k)} + \frac{\alpha_{i-1,i}^2 \hat{\pi}_{i-1}^{(k)}}{\pi_i^{(k)}} \delta_{i-1}^{(k)} \tag{2}$$

In sum, at the time of the update, Node  $i$  needs to have access to the following quantities:

**Its own predictions:**  $\hat{\mu}_i^{(k)}, \hat{\pi}_i^{(k)}$

**Coupling strength:**  $\alpha_{i-1,i}$

**From level below:**  $\delta_{i-1}^{(k)}, \hat{\pi}_{i-1}^{(k)}$

All of these are available at the time of the update. Node  $i$  therefore only needs to receive the PE and the predicted precision from the level below to perform its update.

## 2.2 Prediction Error Step

We will assume in the following, that Node  $i$  is the value child of Node  $i+1$ . Then the following quantities have to be sent up to Node  $i+1$  (cf. necessary information from level below in a value parent):

**Predicted precision:**  $\hat{\pi}_i^{(k)}$

**Prediction error:**  $\delta_i^{(k)}$

Node  $i$  has already performed the PREDICTION step on the previous trial, so it has already computed the predicted precision of the current trial,  $\hat{\pi}_i^{(k)}$ . Hence, in the PE step, it needs to perform only the following calculation:

$$\delta_i^{(k)} = \mu_i^{(k)} - \hat{\mu}_i^{(k)} \quad (3)$$

## 2.3 Prediction Step

Still assuming that Node  $i$  is the value child of Node  $i+1$ , the PREDICTION step consists of the following computations:

$$\hat{\mu}_i^{(k+1)} = \mu_i^{(k)} + \alpha_{i,i+1} \mu_{i+1}^{(k)} \quad (4)$$

$$\hat{\pi}_i^{(k+1)} = \frac{1}{\frac{1}{\pi_i^{(k)}} + \nu_i^{(k+1)}} \quad (5)$$

with

$$\nu_i^{(k+1)} = \exp(\omega_i). \quad (6)$$

Note that if Node  $i$  additionally has a VOPE parent node, the estimated volatility  $\nu_i^{(k+1)}$  that enters the precision update would also depend on the posterior mean of that volatility parent (cf. PREDICTION step for VOPE coupling).

In general, the prediction of the mean will depend only on whether Node  $i$  has a value parent or not, whereas the prediction of the precision only depends on whether Node  $i$  has a volatility parent or not.

Thus, the PREDICTION step only depends on knowing the node's own posteriors and receiving the value parent's posterior in time before the new input arrives.

### 3 Computations for VOPE coupling

As in the case of VAPE coupling, the exact computations of the UPDATE step depend on the nature of the coupling with the child node(s), while both the PE step and the PREDICTION step depend on the coupling with the parent node(s).

To describe the computations entailed by VOPE coupling, we will introduce two changes to the notation. First of all, we will express the volatility PE, or VOPE, as a function of the previously defined value PE, or VAPE. That means from now on, we will use the character  $\delta_i$  only for VAPES:

$$\delta_i^{(k)} \equiv \delta_i^{(k,VAPE)} = \mu_i^{(k)} - \hat{\mu}_i^{(k)}, \quad (7)$$

and introduce a new character  $\Delta_i$  for VOPEs, which we define as

$$\begin{aligned} \Delta_i^{(k)} \equiv \delta_i^{(k,VOPE)} &= \frac{\frac{1}{\pi_i^{(k)}} + (\mu_i^{(k)} - \hat{\mu}_i^{(k)})^2}{\frac{1}{\pi_i^{(k-1)}} + \nu_i^{(k)}} - 1 \\ &= \hat{\pi}_i^{(k)} \left( \frac{1}{\pi_i^{(k)}} + (\mu_i^{(k)} - \hat{\mu}_i^{(k)})^2 \right) - 1 \\ &= \hat{\pi}_i^{(k)} \left( \frac{1}{\pi_i^{(k)}} + (\delta_i^{(k)})^2 \right) - 1 \\ &= \frac{\hat{\pi}_i^{(k)}}{\pi_i^{(k)}} + \hat{\pi}_i^{(k)} (\delta_i^{(k)})^2 - 1. \end{aligned} \quad (8)$$

Note that from the first to the second line, we have used the following definition:

$$\hat{\pi}_{i-1}^{(k)} = \frac{1}{\frac{1}{\pi_{i-1}^{(k-1)}} + \nu_{i-1}^{(k)}}.$$

This ensures that a given node does not need to have access to the posterior precision from the level below:  $\pi_{i-1}^{(k-1)}$ , which facilitates implementation.

In sum, we are introducing a second prediction error unit  $\Delta_i$  which is concerned with deviations from predicted uncertainty and is informed by value prediction errors and other estimates of uncertainty. It is this prediction error - a function of the unweighted (squared) value prediction error with a new precision weight - which communicates between a level's nodes and a level's volatility parent's nodes.

Second, we will introduce another quantity, which we term the (auxiliary) expected precision

$$\gamma_i^{(k)} = \nu_i^{(k)} \hat{\pi}_i^{(k)}, \quad (9)$$

which will be computed as part of the PREDICTION step and only serves to simplify the equations and the corresponding message passing.

### 3.1 Update Step

If Node  $i$  is the volatility parent of Node  $i - 1$ , then the following update equations apply to Node  $i$ :

$$\begin{aligned} \pi_i^{(k)} &= \hat{\pi}_i^{(k)} + \frac{1}{2}(\kappa_{i-1} \nu_{i-1}^{(k)} \hat{\pi}_{i-1}^{(k)})^2 * (1 + (1 - \frac{1}{\pi_{i-1}^{(k-1)} \nu_{i-1}^{(k)}}) \delta_{i-1}^{(k)}) \\ &= \hat{\pi}_i^{(k)} + \frac{1}{2}(\kappa_{i-1} \nu_{i-1}^{(k)} \hat{\pi}_{i-1}^{(k)})^2 * (1 + (2 - \frac{1}{\hat{\pi}_{i-1}^{(k)} \nu_{i-1}^{(k)}}) \delta_{i-1}^{(k)}) \\ \mu_i^{(k)} &= \hat{\mu}_i^{(k)} + \frac{1}{2} \kappa_{i-1} \nu_{i-1}^{(k)} \frac{\hat{\pi}_{i-1}^{(k)}}{\pi_i^{(k)}} \delta_{i-1}^{(k)}, \end{aligned}$$

where we have again used the definition of the predicted precision  $\hat{\pi}_{i-1}^{(k)}$  to derive an expression for the posterior precision from the previous trial  $\pi_{i-1}^{(k-1)}$ :

$$\hat{\pi}_{i-1}^{(k)} = \frac{1}{\frac{1}{\pi_{i-1}^{(k-1)}} + \nu_{i-1}^{(k)}}$$

$$\Leftrightarrow \pi_{i-1}^{(k-1)} = \frac{1}{\frac{1}{\hat{\pi}_{i-1}^{(k)}} - \nu_{i-1}^{(k)}}.$$

With the changes from above, namely the definitions of the VOPE  $\Delta_i$  and the expected precision  $\gamma_i^{(k)}$ , the update equations for the precision and the mean in volatility coupling simplify to:

$$\pi_i^{(k)} = \hat{\pi}_i^{(k)} + \frac{1}{2}(\kappa_{i,i-1}\gamma_{i-1}^{(k)})^2 + (\kappa_{i,i-1}\gamma_{i-1}^{(k)})^2\Delta_{i-1}^{(k)} - \frac{1}{2}\kappa_{i,i-1}\gamma_{i-1}^{(k)}\Delta_{i-1}^{(k)} \quad (10)$$

$$\mu_i^{(k)} = \hat{\mu}_i^{(k)} + \frac{1}{2}\frac{\kappa_{i,i-1}\gamma_{i-1}^{(k)}}{\pi_i^{(k)}}\Delta_{i-1}^{(k)} \quad (11)$$

Therefore, at the time of the update, Node  $i$  needs to have access to the following quantities:

**Its own predictions:**  $\hat{\mu}_i^{(k)}, \hat{\pi}_i^{(k)}$

**Coupling strength:**  $\kappa_{i,i-1}$

**From level below:**  $\Delta_{i-1}^{(k)}, \hat{\pi}_{i-1}^{(k)}, \nu_{i-1}^{(k)}, \gamma_{i-1}^{(k)}$

### 3.2 Prediction Error Step

The exact computation of the prediction error depends, like the computation of the new prediction, on the nature of the coupling with the parent nodes. We will therefore assume in the following, that Node  $i$  is the volatility child of Node  $i+1$ . Then the following quantities have to be sent up to Node  $i+1$  (see also necessary information from level below in a volatility parent):

**Volatility estimate:**  $\nu_i^{(k)}$

**Predicted precision:**  $\hat{\pi}_i^{(k)}$

**Expected precision:**  $\gamma_i^{(k)}$

**Prediction error:**  $\Delta_i^{(k)}$

Node  $i$  has already performed the **PREDICTION step** on the previous trial, so it has already computed the predicted precision,  $\hat{\pi}_i^{(k)}$ , the expected precision,  $\gamma_i^{(k)}$ , and the volatility estimate,  $\nu_i^{(k)}$ , for the current trial. Hence, in the **PE step**, it needs to perform only the following calculations:

$$\delta_i^{(k)} = \mu_i^{(k)} - \hat{\mu}_i^{(k)} \quad (12)$$

$$\Delta_i^{(k)} = \frac{\hat{\pi}_i^{(k)}}{\pi_i^{(k)}} + \hat{\pi}_i^{(k)} (\delta_i^{(k)})^2 - 1. \quad (13)$$

### 3.3 Prediction Step

Still assuming that Node  $i$  is the volatility child of Node  $i + 1$ , the **PREDICTION step** consists of the following simple computations:

$$\hat{\mu}_i^{(k+1)} = \mu_i^{(k)} \quad (14)$$

$$\nu_i^{(k+1)} = \exp(\kappa_i \mu_{i+1}^{(k)} + \omega_i) \quad (15)$$

$$\hat{\pi}_i^{(k+1)} = \frac{1}{\frac{1}{\pi_i^{(k)}} + \nu_i^{(k+1)}} \quad (16)$$

$$\gamma_i^{(k+1)} = \nu_i^{(k+1)} \hat{\pi}_i^{(k+1)} \quad (17)$$

Thus, the prediction for trial  $k + 1$  depends again only on receiving the posterior mean of Node  $i + 1$  on trial  $k$ , and knowing the Node's own posteriors.

Note that if Node  $i$  additionally has a **VAPE** parent node, the prediction of the new mean,  $\hat{\mu}_i^{k+1}$  would also depend on the posterior mean of that value parent (cf. **PREDICTION step** for **VAPE** coupling).

## 4 Implementation HGF coupling

If we assume that a cortical column consists of units (i.e., sub-populations of neurons), one for each computational step, then the computations for **VAPE** coupling as well as for **VOPE** coupling could be implemented as sketched in Figure 1.

A few points to note here:

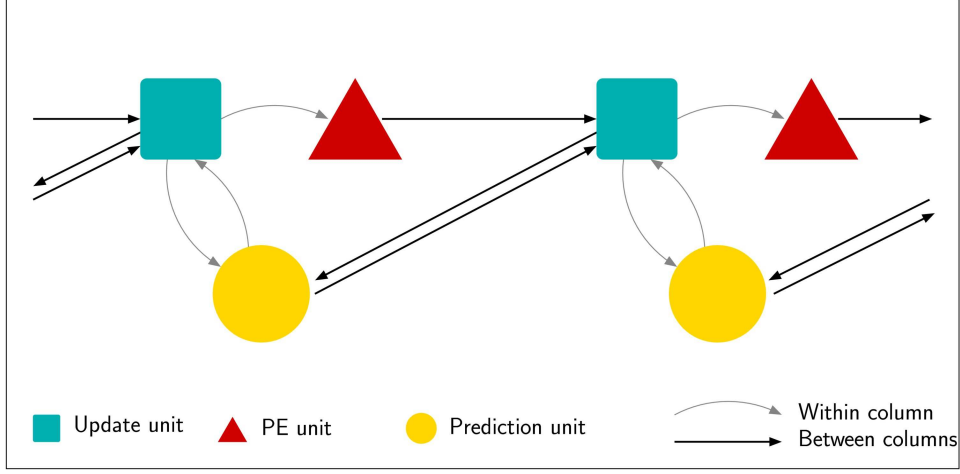


Figure 1: Example implementation of the coupling between cortical columns in the HGF. Figure shows two cortical columns that could be coupled to each other via VAPE or VOPE coupling. In each case, the messages passed along the connections and the computations within the nodes would differ according to the equations presented in the previous chapters.

- In this implementation, we assume that the estimated volatility,  $\nu_i^k$ , is computed within the **Prediction** unit. This is motivated by the equations presented above and the fact that this unit receives information about the updated quantities (posterior means and precisions) from the level above.
- Similarly, the only unit that needs to know the value of a node's tonic learning rate,  $\omega_i$ , is the **Prediction** unit. Therefore, this parameter could be implemented in terms of self-connections or interneurons within that neuronal population.
- A noteworthy arrow in this picture is the one from the **Prediction** unit of one node to the **Update** unit of its parent node. This seems to violate the classical assumptions that while top-down connections signal predictions, bottom-up connections would only signal prediction errors. Due to the update equations in the HGF, however, any given parent node needs access to the predicted precision from its child node to perform an update. To avoid this arrow, we could also assume that the **Prediction** unit sends the predicted precision to the **PE** unit of the same node, which then sends this quantity up to the **Update** unit of the parent node. However, this seems more complicated for two reasons: First, the **PE** unit doesn't actually need this quantity to perform its computation. Secondly, the **PE** is computed only somewhat after the prediction of the current trial (which is already computed after the





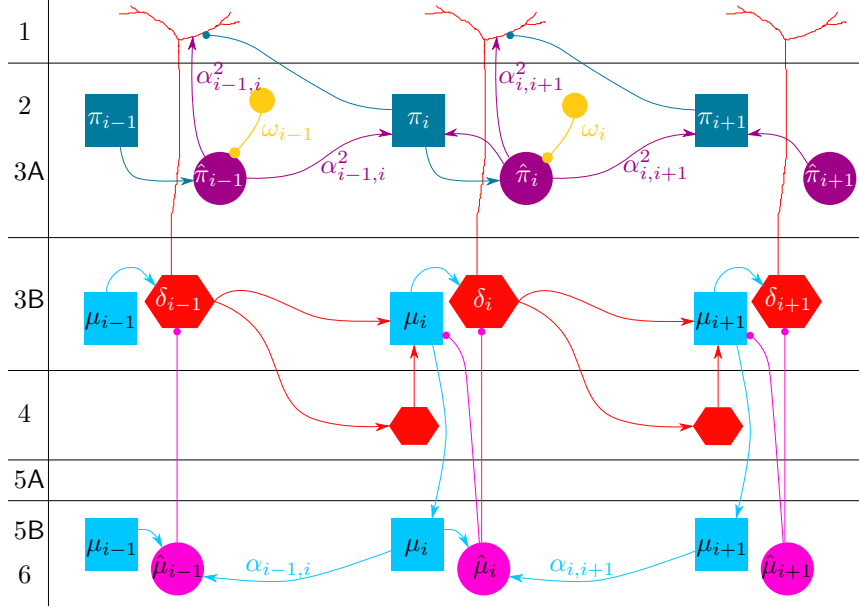


Figure 3: Overview of possible layer-specific message passing in VAPE coupling. Assignment of nodes to layers is loosely based on figure 3 in [3], reproduced in this document in figure 2. Connections with arrowheads indicate positive influences, or excitatory connections, while connections with circular heads indicate inhibitory influences.

PC architecture, while the VOPE coupling comes with more challenges, simply because it hasn’t been addressed in the same detail in existing PC proposals on neural architecture (but see [2]).

### 5.1 VAPE coupling

Figure 3 sketches a possible architecture implementing all computations involved in VAPE coupling for three example cortical regions (levels). The assignment of neural elements to cortical layers follows the proposal in figure 3 of [3], reproduced here in figure 2. For example, we’ve placed all precision-related nodes into the upper layers, while expectations and predictions of the mean live in intermediate and deep layers, respectively. In the following, we go through the different computational steps and note differences to the proposal for predictive coding.

For the PE step, we have the following computation:

$$\delta_i^{(k)} = \mu_i^{(k)} - \hat{\mu}_i^{(k)}. \quad (18)$$

The message passing implicated is depicted in figure 4 and is in line with PE computations as suggested by [3].

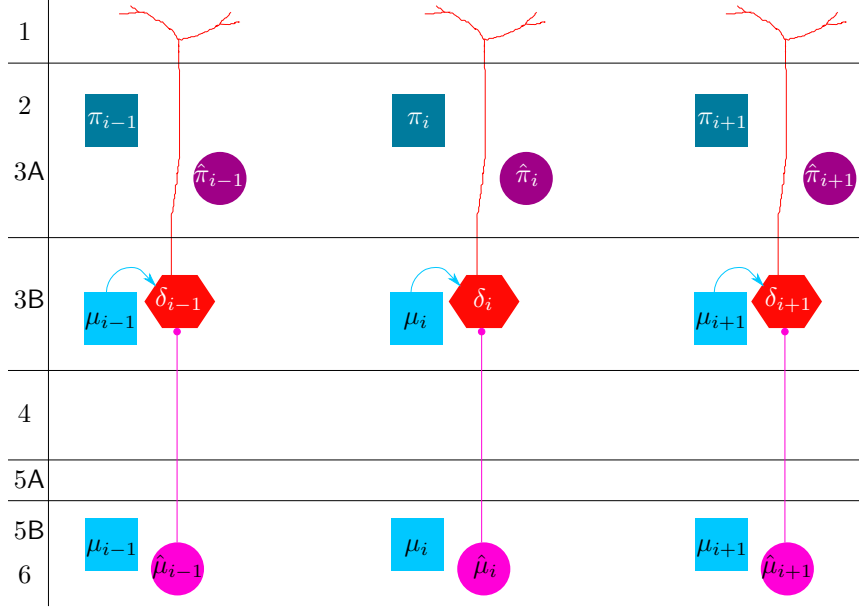


Figure 4: Implementation of VAPE PE computation on three cortical levels:  $\delta_i = \mu_i - \hat{\mu}_i$ .

In the **Update** step, the posterior estimates of mean and variance are given by:

$$\mu_i^{(k)} = \hat{\mu}_i^{(k)} + \frac{\alpha_{i-1,i}^2 \hat{\pi}_{i-1}^{(k)}}{\pi_i^{(k)}} \delta_{i-1}^{(k)} \quad (19)$$

and

$$\pi_i^{(k)} = \hat{\pi}_i^{(k)} + \alpha_{i-1,i}^2 \hat{\pi}_{i-1}^{(k)}. \quad (20)$$

A possible message passing implementation for these updates is shown in figure 5. There are two important differences to [3] to note here.

First, in the PC proposal, the update of the mean  $\mu_i$  is driven both by the lower-level PE  $\delta_{i-1}$  and the PE on the level itself,  $\delta_i$ , which mediates the influence of the empirical prior, or the prediction, on the mean.

In the HGF equations, we do not have this negative feedback loop from  $\delta_i$  to  $\mu_i$ . Instead, the prediction  $\hat{\mu}_i$  has a direct influence on  $\mu_i$  (see equations above), hence the arrow from the  $\hat{\mu}_i$  node to the  $\mu_i$  node in figure 5. This difference is due to the discrete nature of the update equations in the HGF, and can be resolved by considering the potential (PC-like) within-trial temporal evolution of the  $\mu_i$  node. Noting that  $\mu_i$  as computed in the above equation will be the endpoint of the update, i.e., the equilibrium value to

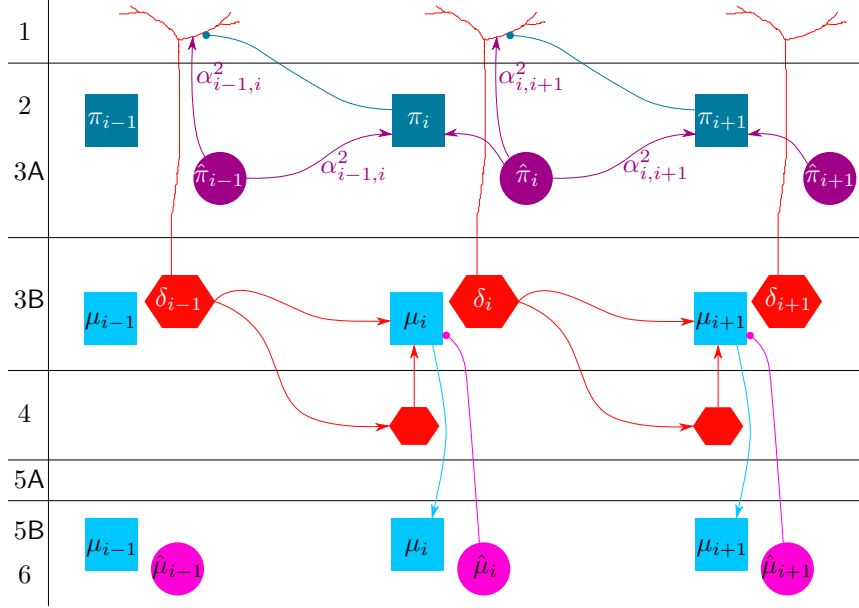


Figure 5: Message passing entailed by the Update step in VAPE coupling.

which the node should stabilize at the end of trial  $k$ , we can, in the style of [1], propose the following temporal evolution:

$$\dot{\mu}_i^{(k)} = \hat{\mu}_i^{(k)} + \frac{\alpha_{i-1,i}^2 \hat{\pi}_{i-1}^{(k)}}{\pi_i^{(k)}} \delta_{i-1}^{(k)} - \mu_i^{(k)} \quad (21)$$

A node with this dynamic will converge to the required posterior value, which can be seen by setting  $\dot{\mu}_i$  to zero. Now we note that

$$\dot{\mu}_i^{(k)} = \frac{\alpha_{i-1,i}^2 \hat{\pi}_{i-1}^{(k)}}{\pi_i^{(k)}} \delta_{i-1}^{(k)} - \mu_i^{(k)} + \hat{\mu}_i^{(k)} \quad (22)$$

can be summarized as

$$\dot{\mu}_i^{(k)} = \frac{\alpha_{i-1,i}^2 \hat{\pi}_{i-1}^{(k)}}{\pi_i^{(k)}} \delta_{i-1}^{(k)} - \delta_i^{(k)}. \quad (23)$$

Following this differential equation for the temporal evolution of node  $\mu_i$ , we could also place the arrow between  $\delta_i$  and  $\mu_i$ , resulting in the same feedback loop as in the PC proposal.<sup>1</sup> The exercise of writing down potential

<sup>1</sup>Note that in the implementation proposed in this chapter, this would not be possible, as the prediction error is always automatically precision-weighted via the modulation of its dendrites in the superficial layers. For the feedback loop however, one would need the prediction error in its unweighted form, a requirement that will be met in the next chapter.

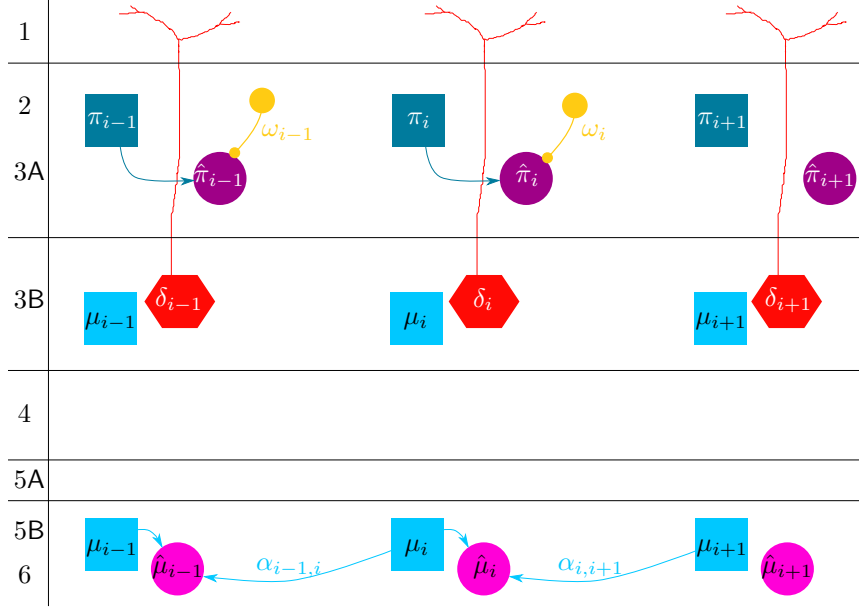


Figure 6: Message passing entailed by the Prediction step in VAPE coupling.

within-trial temporal dynamics for all the HGF nodes will be performed and discussed in the next chapter, where we will see that the resulting equations for continuous time network dynamics result in a more detailed and more biologically plausible implementation scheme.

The second difference to the PC proposal is rather irreducible. As already mentioned in the previous section, the messages sent up to parent nodes do not only consist of prediction errors or their precision weights. Instead, to update the precision node  $\pi_i$ , node  $i$  also needs access to the prediction precision  $\hat{\pi}_{i-1}$  from the level below. This is reflected in the violet arrow travelling from superficial layers of level  $i$  to level  $i + 1$ . However, given that in most PC proposals the update equations were concerned only with hierarchies for estimating the mean and not the precision (but see [2] for an exception), it does not seem too surprising that the explicit treatment of precision updates in the HGF results in new predictions for cortical architecture.

Finally, in the Prediction step, we have:

$$\hat{\mu}_i^{(k+1)} = \mu_i^{(k)} + \alpha_{i,i+1} \mu_{i+1}^{(k)} \quad (24)$$

and

$$\hat{\pi}_i^{(k+1)} = \frac{1}{\frac{1}{\pi_i^{(k)}} + \exp(\omega_i)}. \quad (25)$$

This again seems unproblematic, although the arrow from the posterior  $\mu_i$  to the prediction  $\hat{\mu}_i$  would not appear in the classical PC architecture, where the influence of the posterior on the new prediction is again mediated by the prediction error node. This can be resolved by moving to a continuous time representation in the same way as seen before and will be discussed below.

Finally, we have placed the predicted precision  $\hat{\pi}$  in the superficial layers here, for convenience, as it interacts only with the posterior precision  $\pi$  and acts as a weight on the prediction error, and also in accordance with [3], where precision signals arise from layers 2 and 3A.

## 5.2 VOPE coupling

For volatility parents, we propose the following: Each level of a cortical hierarchy implements its own volatility parent - or volatility estimation - in subpopulations of neurons within the same cortical column. Coupling across cortical levels, on the other hand, is exclusively value-related. This means that having a volatility parent is nothing more but a more sophisticated way of predicting and updating a level's precision estimates.

This setup deliberately excludes volatility parents of volatility parents. However, it allows for a cortical hierarchy where higher levels predict either the mean, or the volatility (or both), of lower levels. In other words, any given level receives top-down predictions not only about its mean, but also about its volatility, the current estimation of which is represented in the mean of its volatility parent. Higher-level predictions of volatility (top-down) and deviations from these predictions - value prediction errors about the volatility mean (bottom-up) are communicated between cortical levels via value (VAPE) coupling in the same way as higher-level predictions of the mean and deviations (value prediction errors about the state mean).

!Include schematic visualization of the different scenarios here!

Note that this setup also allows for a centralized or global volatility estimation. For example, a cortical or subcortical region which monitors the overall environmental volatility could be coupled to all local volatility units via value coupling. Our proposal is more flexible than previous ones [2] though, in that it also enables differential volatility estimation, e.g. within different sensory channels.

Moreover, in principle, both the mean  $\mu_i$  as well as the volatility estimate  $\tilde{\mu}_i$  of a given level could have (one or more) value parents in higher cortical areas. This allows for both separate and combined higher-level prediction

of both the precision and the mean of level  $i$ .

Due to our proposal, we slightly change the notation for volatility coupling from here on. Level  $i$  from now on refers to the hierarchical level within a cortical hierarchy. Nodes on this level comprise the 'original' nodes  $\mu_i$ ,  $\pi_i$ ,  $\hat{\mu}_i$ ,  $\hat{\pi}_i$ , and  $\delta_i$ , which are concerned with value or state estimation, but also all the nodes belonging to a volatility parent of these nodes, which we now denote as  $\check{\mu}_i$ ,  $\check{\pi}_i$ ,  $\hat{\check{\mu}}_i$ ,  $\hat{\check{\pi}}_i$ , and  $\check{\delta}_i$ . Here,  $\check{\delta}_i$  refers to the value prediction error about the mean of the volatility,  $\check{\mu}_i$ .

The VOPE  $\Delta_i$ , which we have written as a function of the VAPE  $\delta_i$ , of level  $i$ , is what communicates between a level's nodes and a level's volatility parent's nodes. Value estimation and volatility estimation both happen within a cortical level or column. For the ensuing equations (simple reformulations of the computations in VOPE coupling as presented before) and visualizations of this proposal, please refer to the appendix.

!Discuss similarities with and differences to [2] here!

## 6 Within-trial time dynamics

In this section, we derive differential equations which lay out the within-trial update dynamics entailed by the HGF. To this end, we consider the posterior values of all our nodes (quantities) as given by the HGF update equations as the equilibrium point towards which all dynamics must converge.

We will note that these equations imply a slightly different coupling between nodes compared to the previous section. On first glance, this complicates the picture, as we're introducing more and more nodes. On the other hand, we increase biological plausibility by making the computations of a single node or connection simpler and simpler.

We divide the equations into value estimation and volatility estimation, noting that VAPE and VOPE coupling are not useful divisions anymore (the mean of the value estimation of a level  $i$ ,  $\mu_i$ , will be VAPE-coupled to its parent  $\mu_{i+1}$ , while the precision  $\pi_i$  of the value estimation will be VOPE-coupled to the volatility parent  $\check{\mu}_i$  on level  $i$  itself; on the other hand, the mean of the volatility estimation,  $\check{\mu}_i$ , can again have a VAPE coupling to a higher level, e.g.  $\mu_{i+2}$ , and the precision of volatility,  $\check{\pi}_i$ , will, in the current framework, not be coupled to any parent, i.e., its prediction will depend only on the parameter  $\check{\omega}_i$ ).

## 6.1 Value estimation

For the PE about the mean, we want to reach the following posterior:

$$\delta_i^{(k)} = \mu_i^{(k)} - \hat{\mu}_i^{(k)}. \quad (26)$$

This can be easily achieved by a node with these dynamics:

$$\dot{\delta}_i^{(k)} = \mu_i^{(k)} - \hat{\mu}_i^{(k)} - \delta_i^{(k)}. \quad (27)$$

From this it follows that we require an additional inhibitory self-connection for the PE node.

We now consider an alternative implementation for the PE computation, which will turn out to be very useful.

Instead of considering the unweighted  $\delta_i$ , we can also directly model a node which corresponds to the precision-weighted prediction error  $\varepsilon_i$ :

$$\varepsilon_i = \frac{\alpha^2 \hat{\pi}_i}{\pi_{i+1}} (\mu_i - \hat{\mu}_i) = \frac{\alpha^2 \hat{\pi}_i}{\pi_{i+1}} \delta_i. \quad (28)$$

This node could evolve according to

$$\dot{\varepsilon}_i = \mu_i - \hat{\mu}_i - \frac{\pi_{i+1}}{\alpha^2 \hat{\pi}_i} \varepsilon_i, \quad (29)$$

which allows us to re-introduce the unweighted PE  $\delta_i$ :

$$\begin{aligned} \delta_i &= \frac{\pi_{i+1}}{\alpha^2 \hat{\pi}_i} \varepsilon_i \\ &= \frac{\pi_{i+1} (\mu_i - \hat{\mu}_i) \alpha^2 \hat{\pi}_i}{\alpha^2 \hat{\pi}_i \pi_{i+1}} \\ &= \mu_i - \hat{\mu}_i, \end{aligned} \quad (30)$$

such that

$$\dot{\varepsilon}_i = \mu_i - \hat{\mu}_i - \delta_i \quad (31)$$

and

$$\dot{\delta}_i = \varepsilon_i - \alpha^2 \frac{\hat{\pi}_i}{\pi_{i+1}} \delta_i. \quad (32)$$

This has the advantage that the weighted PE  $\varepsilon_i$  can travel up the hierarchy to update higher levels, while the unweighted PE  $\delta_i$  can be used to feedback to the mean on the same level, an element that will be introduced next.



In the **Update** step, the posterior estimate of the mean is given by:

$$\mu_i^{(k)} = \hat{\mu}_i^{(k)} + \frac{\alpha_{i-1,i}^2 \hat{\pi}_{i-1}^{(k)}}{\pi_i^{(k)}} \delta_{i-1}^{(k)}. \quad (33)$$

Here, we propose the following temporal evolution:

$$\dot{\mu}_i^{(k)} = \hat{\mu}_i^{(k)} + \frac{\alpha_{i-1,i}^2 \hat{\pi}_{i-1}^{(k)}}{\pi_i^{(k)}} \delta_{i-1}^{(k)} - \mu_i^{(k)} \quad (34)$$

Clearly, a node with this dynamic converges to the required posterior value, which can be seen by setting  $\dot{\mu}_i$  to zero. Now we note that

$$\dot{\mu}_i^{(k)} = \frac{\alpha_{i-1,i}^2 \hat{\pi}_{i-1}^{(k)}}{\pi_i^{(k)}} \delta_{i-1}^{(k)} - \mu_i^{(k)} + \hat{\mu}_i^{(k)} \quad (35)$$

can be summarized as

$$\dot{\mu}_i^{(k)} = \varepsilon_{i-1}^{(k)} - \delta_i^{(k)}. \quad (36)$$

We will therefore consider a connection between  $\delta_i$  and  $\mu_i$ , instead of connecting the prediction node  $\hat{\mu}_i$  with the mean  $\mu_i$ . As already outlined in the previous section, this also resolves one of the apparent differences between our message passing scheme and classical PC proposals.

For the posterior precision, we have

$$\pi_i^{(k)} = \hat{\pi}_i^{(k)} + \alpha_{i-1,i}^2 \hat{\pi}_{i-1}^{(k)}. \quad (37)$$

Here, we stick with the same approach as for the PE node, and simply add a self-inhibitory connection to implement the dynamics:

$$\dot{\pi}_i^{(k)} = \hat{\pi}_i^{(k)} + \alpha_{i-1,i}^2 \hat{\pi}_{i-1}^{(k)} - \pi_i^{(k)}. \quad (38)$$

Finally, in the **Prediction** step, we want to reach the following prediction of the mean:

$$\hat{\mu}_i^{(k+1)} = \mu_i^{(k)} + \alpha_{i,i+1} \mu_{i+1}^{(k)}. \quad (39)$$

Rewriting this as a differential equation

$$\dot{\hat{\mu}}_i^{(k+1)} = \alpha_{i,i+1} \mu_{i+1}^{(k)} + \mu_i^{(k)} - \hat{\mu}_i^{(k+1)}, \quad (40)$$

we again see the PE node appear:

$$\dot{\hat{\mu}}_i^{(k+1)} = \alpha_{i,i+1} \mu_{i+1}^{(k)} + \delta_i^{(k)}. \quad (41)$$

Note that  $\delta_i^{(k)}$  is actually defined in terms of  $\hat{\mu}_i^{(k)}$ , and not  $\hat{\mu}_i^{(k+1)}$ , but given that we are in continuous time now, we suspect that this is probably equivalent here or can be resolved by introducing a delay between the computation of the prediction and the prediction error. For simplicity, we will skip the trial indices from now on.

The precision of this prediction should converge to:

$$\hat{\pi}_i = \frac{1}{\frac{1}{\pi_i} + \exp(\omega_i)}. \quad (42)$$

We can write the dynamics as:

$$\begin{aligned} \dot{\hat{\pi}}_i &= 1 - \left(\frac{1}{\pi_i} + \exp(\omega_i)\right)\hat{\pi}_i \\ &= 1 - \frac{\hat{\pi}_i}{\pi_i} - \exp(\omega_i)\hat{\pi}_i. \end{aligned} \quad (43)$$

To simplify the implementation, we introduce an additional node  $p_{i1}$ :

$$p_{i1} = \frac{\hat{\pi}_i}{\pi_i}. \quad (44)$$

This new node can have the following dynamics:

$$\dot{p}_{i1} = \hat{\pi}_i - \pi_i p_{i1}, \quad (45)$$

and the dynamics for the precision of the prediction  $\hat{\pi}_i$  become

$$\dot{\hat{\pi}}_i = 1 - p_{i1} - \exp(\omega_i)\hat{\pi}_i, \quad (46)$$

where the last term can be implemented as a self-inhibition with a connection weight of  $\exp(\omega_i)$ . However, to facilitate the implementation of the influence of a volatility parent on the precision estimate, we introduce another additional node,  $p_{i2}$ , here and also already consider a node that estimates volatility,  $\nu_i$ . In the case of value estimation without a volatility node, this estimate will only depend on the parameter value  $\omega_i$ , i.e., the tonic volatility.

The dynamics of  $\hat{\pi}_i$  then become:

$$\dot{\hat{\pi}}_i = 1 - p_{i1} - p_{i2}, \quad (47)$$

with the node  $p_{i2}$  defined as

$$p_{i2} = \hat{\pi}_i \exp(\omega_i), \quad (48)$$

with dynamics described by

$$\dot{p}_{i2} = \exp(\omega_i)\hat{\pi}_i - p_{i2} \quad (49)$$

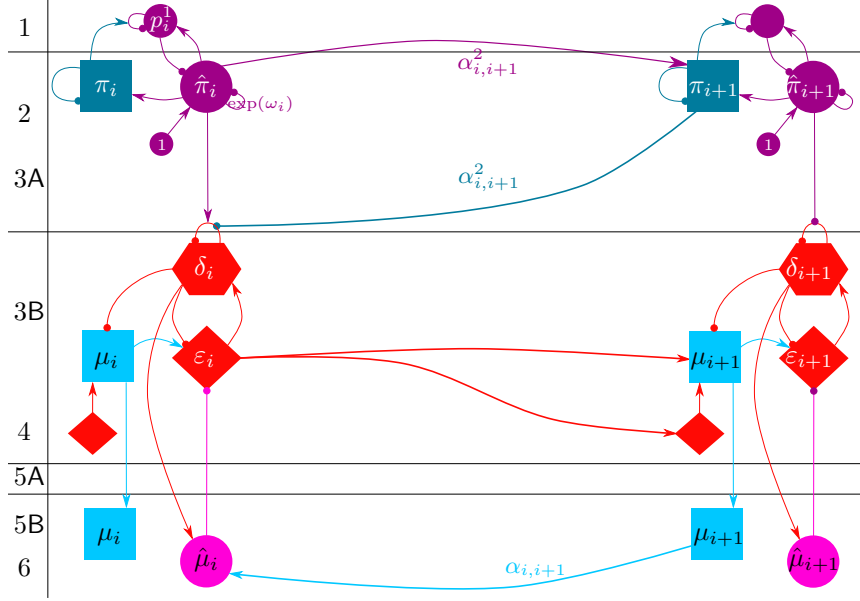


Figure 7: The coupling between a hierarchical level  $i$  with its parent  $i + 1$  in the case of VAPE coupling: implementing the differential equations that describe within-trial dynamics of the nodes.

in the absence of a volatility parent (see below for the extension to volatility estimation). We introduce the volatility estimate  $\nu_i$  as

$$\nu_i = \exp(\omega_i), \quad (50)$$

evolving as

$$\dot{\nu}_i = \exp(\omega_i) - \nu_i, \quad (51)$$

leading to

$$\dot{p}_{i2} = \nu_i \hat{\pi}_i - p_{i2}. \quad (52)$$

A visualization of the network entailed by the differential equations so far - in the absence of volatility estimation - is depicted in figure 7.

Note that the dynamics of  $\nu_i$  can be implemented by a self-inhibition and a positive influence from a tonically active neuron (with activity 1) with a connection weight of  $\exp(\omega_i)$ . The advantage of this description is that if we add a volatility parent, we simply have to let this tonically active neuron be driven by the volatility estimate  $\check{\mu}_i$ , weighted by the coupling parameter  $\kappa_i$ , to arrive at

$$\nu_i = \exp(\kappa_i \check{\mu}_i) \exp(\omega_i). \quad (53)$$



$\hat{\pi}_i$  on  $p_{i2}$ , lead to the desired influence:

$$\dot{\nu}_i = \exp(\kappa_i \check{\mu}_i) \exp(\omega_i) - \nu_i. \quad (55)$$

Additionally, to simplify the precision weighting of the volatility prediction error, we use the auxiliary expected precision node  $\gamma_i$  as introduced in the previous section:

$$\gamma_i = \kappa_i \nu_i \hat{\pi}_i \quad (56)$$

with, for example,

$$\dot{\gamma}_i = \kappa_i \nu_i \hat{\pi}_i - \gamma_i. \quad (57)$$

For the PE about the level of volatility  $\Delta_i$ , we have previously seen that we want to reach the following posterior:

$$\Delta_i^{(k)} = \frac{\hat{\pi}_i^{(k)}}{\pi_i^{(k)}} + \hat{\pi}_i^{(k)} (\delta_i^{(k)})^2 - 1. \quad (58)$$

This corresponds to the unweighted volatility prediction error. Again, we start by considering a node which in turn reflects the precision-weighted (VO)PE, which we will call  $E_i$ :

$$E_i^{(k)} = \frac{1}{2} \frac{\gamma_i^{(k)}}{\check{\pi}_i^{(k)}} \Delta_i^{(k)}. \quad (59)$$

This could evolve according to

$$\dot{E}_i^{(k)} = \Delta_i^{(k)} - \frac{2\check{\pi}_i^{(k)}}{\gamma_i^{(k)}} E_i^{(k)}, \quad (60)$$

which can also be written as

$$\dot{E}_i^{(k)} = \frac{\hat{\pi}_i^{(k)}}{\pi_i^{(k)}} + \hat{\pi}_i^{(k)} (\delta_i^{(k)})^2 - 1 - \Delta_i^{(k)}, \quad (61)$$

with the unweighted PE node  $\Delta_i$

$$\Delta_i^{(k)} = \frac{2\check{\pi}_i^{(k)}}{\gamma_i^{(k)}} E_i^{(k)} \quad (62)$$

evolving as

$$\dot{\Delta}_i^{(k)} = E_i^{(k)} - \frac{\gamma_i^{(k)}}{2\check{\pi}_i^{(k)}} \Delta_i^{(k)}. \quad (63)$$



to communicate with higher cortical levels  $\mu_{i+1}$ , completely parallel to the precision-weighted value prediction about the mean,  $\varepsilon_i$ .

Third, we note that in volatility estimation, we have to distinguish between the (weighted and unweighted) prediction errors that are used to *update* the volatility estimate,  $E$  and  $\Delta$ , and the (weighted and unweighted) prediction errors *about* the volatility estimate,  $\check{\varepsilon}_i$  and  $\check{\delta}_i$ , which will be used to update any higher cortical level  $i + 1$  predicting this volatility estimate.

Finally, although the unweighted volatility PE  $\Delta_i$  is not needed for a feedback to  $\mu_i$  or  $\check{\mu}$ , it is still very useful as it is used in the update of the precision of the volatility estimation  $\check{\pi}_i$ :

$$\check{\pi}_i^{(k)} = \hat{\pi}_i^{(k)} + \frac{1}{2}(\gamma_i^{(k)})^2 + (\gamma_i^{(k)})^2 \Delta_i^{(k)} - \frac{1}{2}\gamma_i^{(k)} \Delta_i^{(k)}, \quad (66)$$

which can be summarized as

$$\check{\pi}_i^{(k)} = \hat{\pi}_i^{(k)} + \frac{1}{2}(\gamma_i^{(k)})^2 + ((\gamma_i^{(k)})^2 - \frac{1}{2}\gamma_i^{(k)})\Delta_i^{(k)}. \quad (67)$$

This could have the temporal evolution

$$\dot{\check{\pi}}_i^{(k)} = \hat{\pi}_i^{(k)} + \frac{1}{2}(\gamma_i^{(k)})^2 + ((\gamma_i^{(k)})^2 - \frac{1}{2}\gamma_i^{(k)})\Delta_i^{(k)} - \check{\pi}_i^{(k)}. \quad (68)$$

Finally, turning towards the Prediction step in the volatility estimation, this is again very simple, as we only consider value parents for volatility nodes. For the prediction of the mean, we have

$$\hat{\mu}_i^{(k+1)} = \check{\mu}_i^{(k)} + \alpha_{i,i+1}\mu_{i+1}^{(k)}, \quad (69)$$

which could evolve as

$$\dot{\hat{\mu}}_i^{(k+1)} = \check{\mu}_i^{(k)} - \hat{\mu}_i^{(k+1)} + \alpha_{i,i+1}\mu_{i+1}^{(k)} = \check{\delta}_i + \alpha_{i,i+1}\mu_{i+1}^{(k)}, \quad (70)$$

and for the precision of that prediction, we get

$$\hat{\pi}_i^{(k+1)} = \frac{1}{\frac{1}{\hat{\pi}_i^{(k)}} + \exp(\check{\omega}_i)}, \quad (71)$$

which could evolve as shown above in the value estimation (using the auxiliary nodes  $\check{p}_{i1}$ ):

$$\dot{\hat{\pi}}_i = 1 - \check{p}_{i1} - \hat{\pi}_i \exp(\check{\omega}_i) \quad (72)$$

with

$$\dot{\check{p}}_{i1} = \hat{\pi}_i - \check{\pi}_i \check{p}_{i1}. \quad (73)$$

Note that we do not bother to introduce  $\check{p}_{i2}$  or  $\check{\nu}_i$  here, as we exclude the possibility of a volatility parent  $\check{\mu}_i$  having a volatility parent.

The differential equations can also be used to simulate within-trial responses of all nodes, examining network stability, and predicting neural dynamics which could for example be compared with electrophysiological data.

## 7 Open Questions and Issues

- To Do:
  - Decide on the most useful visualizations for the paper.
  - Check which improvements (in terms of expressing the coupling in the equations) can be translated from the differential equations to the one-step update equations (e.g., introduction of the precision-weighted PE node).
  - Implement the differential equations, examine the stability of the network and present exemplary simulations.
  - Examine the computations for input nodes and their coupling (this is probably not needed for the theoretical treatment of VAPes and VOPEs, but relevant for the python implementation).
  - Discuss overlap and differences with Karls precision paper (Kanai et al. 2015).
- Equations: The coupling parameters  $\alpha$  and  $\kappa$  must be available to both the connection between a node's PE unit and the parent's UPDATE unit, as well as the connection between a node's PREDICTION unit and the parent's UPDATE unit. How can it be ensured that the coupling strength is the same for both connections? For this question we would actually have to consider the learning or update equations for the parameters of the model (see for example the approach in [1]). However, even without writing them down, we could assume that under 'healthy' conditions, the weights of these different connections automatically converge to the same (or at least very similar) values. This opens up new possibilities of modeling abnormal inference: Using the simulations presented above, we could find out how an agent's inference would be altered if these values were not the same, and what range of difference would actually lead to observable changes in the inference.
- Implementation: To whom does the knowledge about parameters, in particular coupling strengths, belong? Ideally, this would belong to the connections themselves and be accessible to both child and parent node. However, in terms of synaptic signalling, how can a node have direct access to a coupling strength other than via the signalled value (which is the product of the coupling strength and the signal)?



- Not covered here: Computations of input nodes and binary input nodes.

## 8 Appendix

### 8.1 Computations in VOPE coupling

In the following, we present the computations of volatility nodes according to the proposal that each cortical level of a hierarchy implements its own volatility estimation.

The **Update** steps in volatility coupling are:

$$\check{\mu}_i^{(k)} = \hat{\mu}_i^{(k)} + \frac{1}{2} \frac{\gamma_i^{(k)}}{\check{\pi}_i^{(k)}} \Delta_i^{(k)} \quad (74)$$

$$\check{\pi}_i^{(k)} = \hat{\pi}_i^{(k)} + \frac{1}{2} (\gamma_i^{(k)})^2 + (\gamma_i^{(k)})^2 \Delta_i^{(k)} - \frac{1}{2} \gamma_i^{(k)} \Delta_i^{(k)} \quad (75)$$

In the **PE** step, the node computes:

$$\Delta_i^{(k)} = \frac{\hat{\pi}_i^{(k)}}{\pi_i^{(k)}} + \hat{\pi}_i^{(k)} (\delta_i^{(k)})^2 - 1. \quad (76)$$

Finally, In the **Prediction** step, we now need to compute four nodes: the predicted (volatility) mean

$$\hat{\mu}_i^{(k+1)} = \check{\mu}_i^{(k)}, \quad (77)$$

the precision of that prediction

$$\hat{\pi}_i^{(k+1)} = \frac{1}{\frac{1}{\check{\pi}_i^{(k)}} + \nu_i^{(k+1)}}, \quad (78)$$

the predicted environmental uncertainty (as a function of the next higher level in the hierarchy,  $\mu_{i+1}$ )

$$\nu_i^{(k+1)} = \exp(\kappa_{i,i+1} \mu_{i+1}^{(k)} + \omega_i), \quad (79)$$

and the new (auxiliary) expected precision

$$\gamma_i^{(k+1)} = \kappa_{i+1,i} \nu_i^{(k+1)} \hat{\pi}_i^{(k+1)}. \quad (80)$$

The last node is only defined for convenience in terms of simplifying the equations and the corresponding message passing.



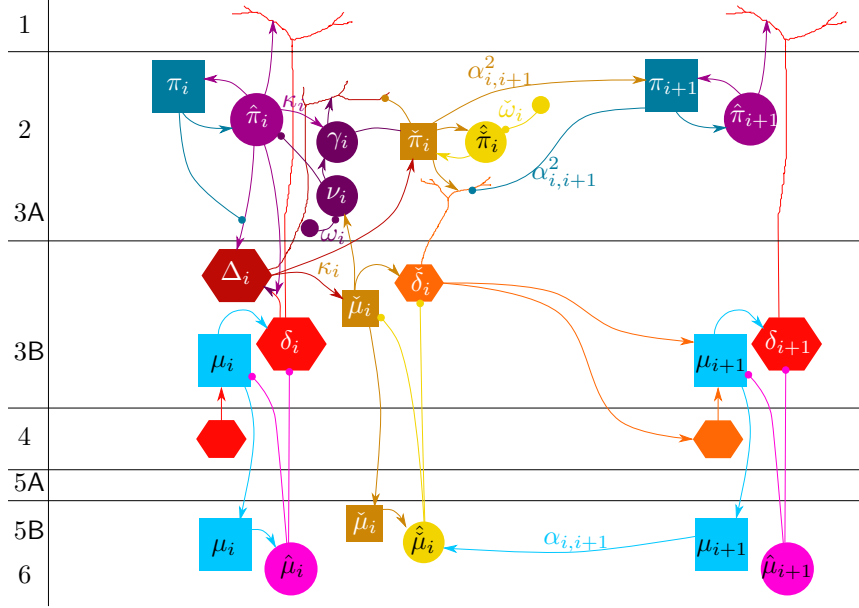


Figure 12: Zoomed-in version of the coupling between a hierarchical level  $i$  with its parent  $i + 1$  in the case of VOPE coupling. Note that the coupling is of VOPE type within level  $i$ , to update  $\check{\mu}_i$ , which is turn predicted by level  $i + 1$ , where the higher level now serves as a VAPE parent to the volatility estimate  $\check{\mu}_i$ .

## 8.2 Possible PC-like implementation of VOPE coupling

Figure 10 displays one proposal for within-column volatility estimation, where we've zoomed in to a level  $i$  of the cortical hierarchy (and a value parent  $i + 1$ ) and added its volatility parent to the superficial layers.

An alternative implementation of this idea is displayed in figure 11. Here, only the precision-related computations of the volatility parent are placed in the superficial layers, while the corresponding prediction errors,  $\Delta_i$  and  $\delta_i$  as well as the actual volatility estimate  $\check{\mu}_i$  live in middle layers, and the prediction of the volatility estimate,  $\hat{\mu}_i$ , lives in layer 6. This setup stresses the structural similarities to the message passing entailed by mean coupling. Moreover, using this setup, we can now easily depict how a higher cortical level  $i + 1$  would serve as a value parent to the volatility estimate in level  $i$  - instead of predicting the mean - by simply exchanging the arrows between the levels as shown in figure 12.

## References

- [1] Bogacz, R. (2017). A tutorial on the free-energy framework for modelling perception and learning. *Journal of Mathematical Psychology*, 76:198–211.
- [2] Kanai, R., Komura, Y., Shipp, S., and Friston, K. J. (2015). Cerebral hierarchies: predictive processing, precision and the pulvinar. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1668):20140169–20140169.
- [3] Shipp, S. (2016). Neural elements for predictive coding. *Frontiers in Psychology*, 7(NOV):1–21.