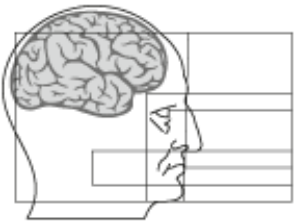


MAX PLANCK UCL CENTRE
for Computational Psychiatry and Ageing Research

UCL

Free energy minimization rests on belief updates that are precision-weighted prediction errors: a general proof.

Plus, an answer to the question: if precisions are crucial to belief updating, how are they themselves updated?

Christoph Mathys

Wellcome Trust Centre for Neuroimaging at UCL, London, UK
Max Planck UCL Centre for Computational Psychiatry and Ageing Research, London, UK

Free Energy Workshop, London, March 17, 2015

Belief building: a shamelessly artificial example

Imagine the following situation:

You're on a boat, you're lost in a storm and trying to get back to shore. A lighthouse has just appeared on the horizon, but you can only see it when you're at the peak of a wave. Your GPS etc., has all been washed overboard, but what you can still do to get an idea of your position is to measure the angle between north and the lighthouse. These are your measurements (in degrees):

76, 73, 75, 72, 77

What number are you going to base your calculation on?

Right. The mean: 74.6. How do you calculate that?

Updates to the mean

The usual way to calculate the mean $\bar{x}$ of $x_1, x_2, \ldots, x_n$ is to take

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

This requires you to remember all x_i, which can become inefficient. Since the measurements arrive sequentially, we would like to update $\bar{x}$ sequentially as the x_i come in – without having to remember them.

It turns out that this is possible. After some algebra (see next slide), we get

$$\bar{x}_{n+1} = \bar{x}_n + \frac{1}{n+1}(x_{n+1} - \bar{x}_n)$$

Updates to the mean

Proof of sequential update formula:

$$\bar{x}_{n+1} = \frac{1}{n+1}\sum_{i=1}^{n+1} x_i = \frac{x_{n+1}}{n+1} + \frac{1}{n+1}\sum_{i=1}^{n} x_i = \frac{x_{n+1}}{n+1} + \frac{n}{n+1}\underbrace{\frac{1}{n}\sum_{i=1}^{n} x_i}_{=\bar{x}_n} =$$

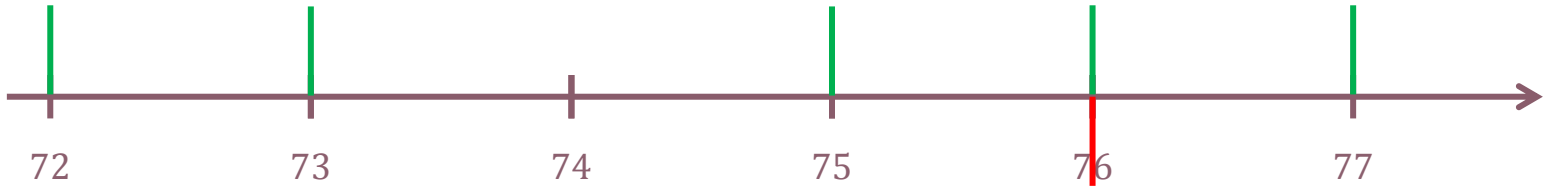
$$= \frac{x_{n+1}}{n+1} + \frac{n}{n+1}\bar{x}_n = \bar{x}_n + \frac{x_{n+1}}{n+1} + \frac{n}{n+1}\bar{x}_n - \frac{n+1}{n+1}\bar{x}_n =$$

$$= \bar{x}_n + \frac{1}{n+1}(x_{n+1} + (n - n - 1)\bar{x}_n) = \bar{x}_n + \frac{1}{n+1}(x_{n+1} - \bar{x}_n)$$

q.e.d.

Updates to the mean

The seqential updates in our example now look like this:



$$\bar{x}_1 = 76$$

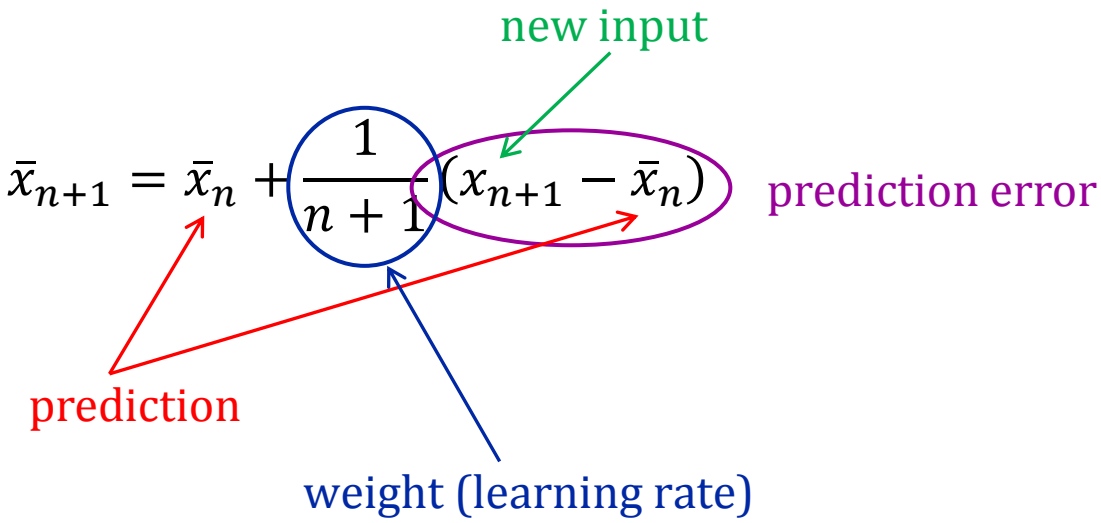
$$\bar{x}_2 = 76 + \frac{1}{2}(73 - 76) = 74.5$$

$$\bar{x}_3 = 74.5 + \frac{1}{3}(75 - 74.5) = 74.\overline{6}$$

$$\bar{x}_4 = 74.\overline{6} + \frac{1}{4}(72 - 74.\overline{6}) = 74$$

$$\bar{x}_5 = 74 + \frac{1}{5}(77 - 74) = 74.6$$

What are the building blocks of the updates we've just seen?



Is this a general pattern?

- More specifically, does it generalize to Bayesian inference?
- «Bayesian inference» simply means inference on uncertain quantities according to the rules of probability theory (i.e., according to logic).
- Crucially, Bayesian inference can be implemented by biological agents by minimizing the variational free energy of a generative model of their sensory input.
- Agents who use Bayesian inference will make better predictions (provided they have a good model of their environment), which will give them an evolutionary advantage.
- We may therefore assume that evolved biological agents use Bayesian inference, or a close approximation to it.
- So is Bayesian inference – **and with it free energy minimization** – based on predictions that are updated using uncertainty-weighted prediction errors?

Updates in a simple Gaussian model

- Think boat, lighthouse, etc., again, but now we're doing Bayesian inference.

- Before we make the next observation, our belief about the true angle ϑ can be described by a Gaussian prior:

$$p(\vartheta) \sim \mathcal{N}(\mu_\vartheta, \pi_\vartheta^{-1})$$

- The likelihood of our observation is also Gaussian, with precision π_ε:

$$p(x|\vartheta) \sim \mathcal{N}(\vartheta, \pi_\varepsilon^{-1})$$

- Bayes' rule now tells us that the posterior is Gaussian again:

$$p(\vartheta|x) = \frac{p(x|\vartheta)p(\vartheta)}{\int p(x|\vartheta')p(\vartheta')\mathrm{d}\vartheta'} \sim \mathcal{N}\left(\mu_{\vartheta|y}, \pi_{\vartheta|y}^{-1}\right)$$

Updates in a simple Gaussian model

- Here's how the updates to the sufficent statistics μ and π describing our belief look like:

$$\pi_{\vartheta|x} = \pi_\vartheta + \pi_\varepsilon$$

$$\mu_{\vartheta|x} = \mu_\vartheta + \frac{\pi_\varepsilon}{\pi_{\vartheta|x}}(x - \mu_\vartheta)$$

prediction error

prediction

weight (learning rate)$=\frac{\text{how much we're learning here}}{\text{how much we already know}}$

- So it's the same story all over again: the mean is updated by an uncertainty-weighted (more specifically: prediction-weighted) prediction error.

- The size of the update is proportional to the likelihood precision and inversely proportional to the posterior precision.

- This pattern is not specific to the univariate Gaussian case, but generalizes to Bayesian updates for all exponential families of likelihood distributions with conjugate priors (i.e., to all formal descriptions of inference you are ever likely to need).

The analogy with simple mean updating goes further

- Reminder (Gaussian update):

$$\mu_{\vartheta|x} = \mu_\vartheta + \frac{\pi_\varepsilon}{\pi_{\vartheta|x}}(x - \mu_\vartheta) = \mu_\vartheta + \frac{\pi_\varepsilon}{\pi_\vartheta + \pi_\varepsilon}(x - \mu_\vartheta)$$

- Reducing by $\pi_\varepsilon$the fraction of precisions that make the learning rate, we get

$$\mu_{\vartheta|x} = \mu_\vartheta + \frac{1}{\frac{\pi_\vartheta}{\pi_\varepsilon} + 1}(x - \mu_\vartheta)$$

- This is again our equation for updating an arithmetic mean, but with n replaced by $\frac{\pi_\vartheta}{\pi_\varepsilon}$.

- This shows that Bayesian inference on the mean of a Gaussian distribution entails nothing more than updating the arithmetic mean of observations with $\frac{\pi_\vartheta}{\pi_\varepsilon} =: \nu$ as a proxy for the number of prior observations, i.e. for the **weight of the prior relative to the observation**.

Generalization to all exponential families of distributions

- Many of the most widely used probability distributions are families of exponential distributions.

- For example, the Gaussian distribution is an exponential family of distributions (and so are the beta, gamma, binomial, Bernoulli, multinomial, categorical, Dirichlet, Wishart, Gaussian-gamma, log-Gaussian, multivariate Gaussian, Poisson, and exponential distributions, among others). This means it can be written the following way:

$$p(\boldsymbol{x}|\boldsymbol{\vartheta}) = h(\boldsymbol{x}) \exp(\boldsymbol{\eta}(\boldsymbol{\vartheta}) \cdot \boldsymbol{T}(\boldsymbol{x}) - A(\boldsymbol{\vartheta})) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-\mu)^2}{2\sigma}\right)$$

with

$$\boldsymbol{x} = x, \qquad \boldsymbol{\vartheta} = (\mu, \sigma)^{\mathrm{T}}, \qquad h(\boldsymbol{x}) = \frac{1}{\sqrt{2\pi}}, \qquad \boldsymbol{\eta}(\boldsymbol{\vartheta}) = \left(\frac{\mu}{\sigma}, -\frac{1}{2\sigma}\right)^{\mathrm{T}}, \qquad \boldsymbol{T}(\boldsymbol{x}) = (x, x^2)^{\mathrm{T}}, \qquad A(\boldsymbol{\vartheta}) = \frac{\mu^2}{\sigma} + \frac{\ln \sigma}{2}$$

- This allows us to look at Bayesian belief updating in a very general way for all exponential families of distributions.

Generalization to all exponential families of distributions

- Our likelihood is an exponential family in its general form:

$$p(\boldsymbol{x}|\boldsymbol{\vartheta}) = h(\boldsymbol{x}) \exp(\boldsymbol{\eta}(\boldsymbol{\vartheta}) \cdot \boldsymbol{T}(\boldsymbol{x}) - A(\boldsymbol{\vartheta}))$$

- The vector $\boldsymbol{T}(\boldsymbol{x})$ (a function of the observation $\boldsymbol{x}$) is called the sufficient statistic.

- For the prior, we may assume that we have made ν observations with sufficient statistic $\boldsymbol{\xi}$:

$$p(\boldsymbol{\vartheta}|\boldsymbol{\xi}, \nu) = z(\boldsymbol{\xi}, \nu) \exp\big(\nu(\boldsymbol{\eta}(\boldsymbol{\vartheta}) \cdot \boldsymbol{\xi} - A(\boldsymbol{\vartheta}))\big) \quad \text{(where } z(\xi, \nu) \text{ is a normlization constant)}$$

- It then turns out that the posterior has the same form, but with an updated $\boldsymbol{\xi}$ and ν replaced with $\nu + 1$:

$$p(\boldsymbol{\vartheta}|\boldsymbol{x}, \boldsymbol{\xi}, \nu) = z(\boldsymbol{\xi}', \nu + 1) \exp\big((\nu + 1)(\boldsymbol{\eta}(\boldsymbol{\vartheta}) \cdot \boldsymbol{\xi}' - A(\boldsymbol{\vartheta}))\big)$$

$$\boldsymbol{\xi}' = \boldsymbol{\xi} + \frac{1}{\nu + 1}(\boldsymbol{T}(\boldsymbol{x}) - \boldsymbol{\xi})$$

Proof of the update equation

$$\overbrace{p(\boldsymbol{\vartheta}|\boldsymbol{x},\boldsymbol{\xi},\nu)}^{\text{posterior}} \propto \overbrace{p(\boldsymbol{x}|\boldsymbol{\vartheta})}^{\text{likelihood}}\ \overbrace{p(\boldsymbol{\vartheta}|\boldsymbol{\xi},\nu)}^{\text{prior}}$$

$$= h(\boldsymbol{x}) \exp(\boldsymbol{\eta}(\boldsymbol{\vartheta}) \cdot \boldsymbol{T}(\boldsymbol{x}) - A(\boldsymbol{\vartheta}))z(\boldsymbol{\xi},\nu) \exp\big(\nu(\boldsymbol{\eta}(\boldsymbol{\vartheta}) \cdot \boldsymbol{\xi} - A(\boldsymbol{\vartheta}))\big)$$

$$\propto \exp\big(\boldsymbol{\eta}(\boldsymbol{\vartheta}) \cdot (\boldsymbol{T}(\boldsymbol{x}) + \nu\boldsymbol{\xi}) - (\nu+1)A(\boldsymbol{\vartheta})\big)$$

$$= \exp\left((\nu+1)\left(\boldsymbol{\eta}(\boldsymbol{\vartheta}) \cdot \frac{1}{\nu+1}(\boldsymbol{T}(\boldsymbol{x}) + \nu\boldsymbol{\xi}) - A(\boldsymbol{\vartheta})\right)\right)$$

$$= \exp\left((\nu+1)\left(\boldsymbol{\eta}(\boldsymbol{\vartheta}) \cdot \left(\boldsymbol{\xi} + \frac{1}{\nu+1}(\boldsymbol{T}(\boldsymbol{x}) + \nu\boldsymbol{\xi} - (\nu+1)\boldsymbol{\xi})\right) - A(\boldsymbol{\vartheta})\right)\right)$$

$$= \exp\left((\nu+1)\left(\boldsymbol{\eta}(\boldsymbol{\vartheta}) \cdot \left(\underbrace{\boldsymbol{\xi} + \frac{1}{\nu+1}(\boldsymbol{T}(\boldsymbol{x}) - \boldsymbol{\xi})}_{=:\boldsymbol{\xi}'}\right) - A(\boldsymbol{\vartheta})\right)\right)$$

$$\Longrightarrow \quad p(\boldsymbol{\vartheta}|\boldsymbol{x},\boldsymbol{\xi},\nu) = z(\boldsymbol{\xi}',\nu') \exp\Big(\nu'\big(\boldsymbol{\eta}(\boldsymbol{\vartheta}) \cdot \boldsymbol{\xi}' - A(\boldsymbol{\vartheta})\big)\Big)$$

$$\text{with } \nu' := \nu + 1, \qquad \boldsymbol{\xi}' := \boldsymbol{\xi} + \frac{1}{\nu+1}(\boldsymbol{T}(\boldsymbol{x}) - \boldsymbol{\xi})$$

q.e.d.

Some examples

- **Univariate Gaussian** model with unkown mean but **known precision** (our example from the beginning):

$$\boldsymbol{T}(x) = x$$

- This means updating beliefs about the mean simply requires tracking the mean of observations

- **Univariate Gaussian** model with unkown mean and unkown precision:

$$\boldsymbol{T}(x) = (x, x^2)^{\mathrm{T}}$$

- Updating beliefs about both mean and precision of a Gaussian requires tracking the means of observations and squared observations; this amounts to the first and second moments by which a Gaussian distribution is fully characterized.

- In the **multivariate Gaussian** case we have $\boldsymbol{T}(\boldsymbol{x}) = \left(\boldsymbol{x}, \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\right)^{\mathrm{T}}$

Some examples

- **Bernoulli** model (one out of two possible outcomes, coded as 0 and 1; e.g., coin flipping):

$$\boldsymbol{T}(x) = x$$

- The prior here turns out to be a **beta distribution** corresponding to ν pseudo-observations with mean ξ. All we need to do to get the posterior (i.e., to update our belief) is to update the mean as new observations come in.

- **Categorical** model (one out of several possible outcomes, with the observed outcome coded as 1, the rest as 0)

$$\boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{x}$$

- The prior and posterior here are **Dirichlet distributions**, and again, all we need to do to update beliefs that have a Dirichlet form is to track the means of observed succeses (1) and failures (0).

Some examples

- **Beta** model (an outcome bounded between 0 and 1):

$$\boldsymbol{T}(x) = (\ln x\,, \ln(1-x))^{\mathrm{T}}$$

- **Gamma** model (an outcome bounded below at 0):

$$\boldsymbol{T}(x) = (\ln x\,, x)^{\mathrm{T}}$$

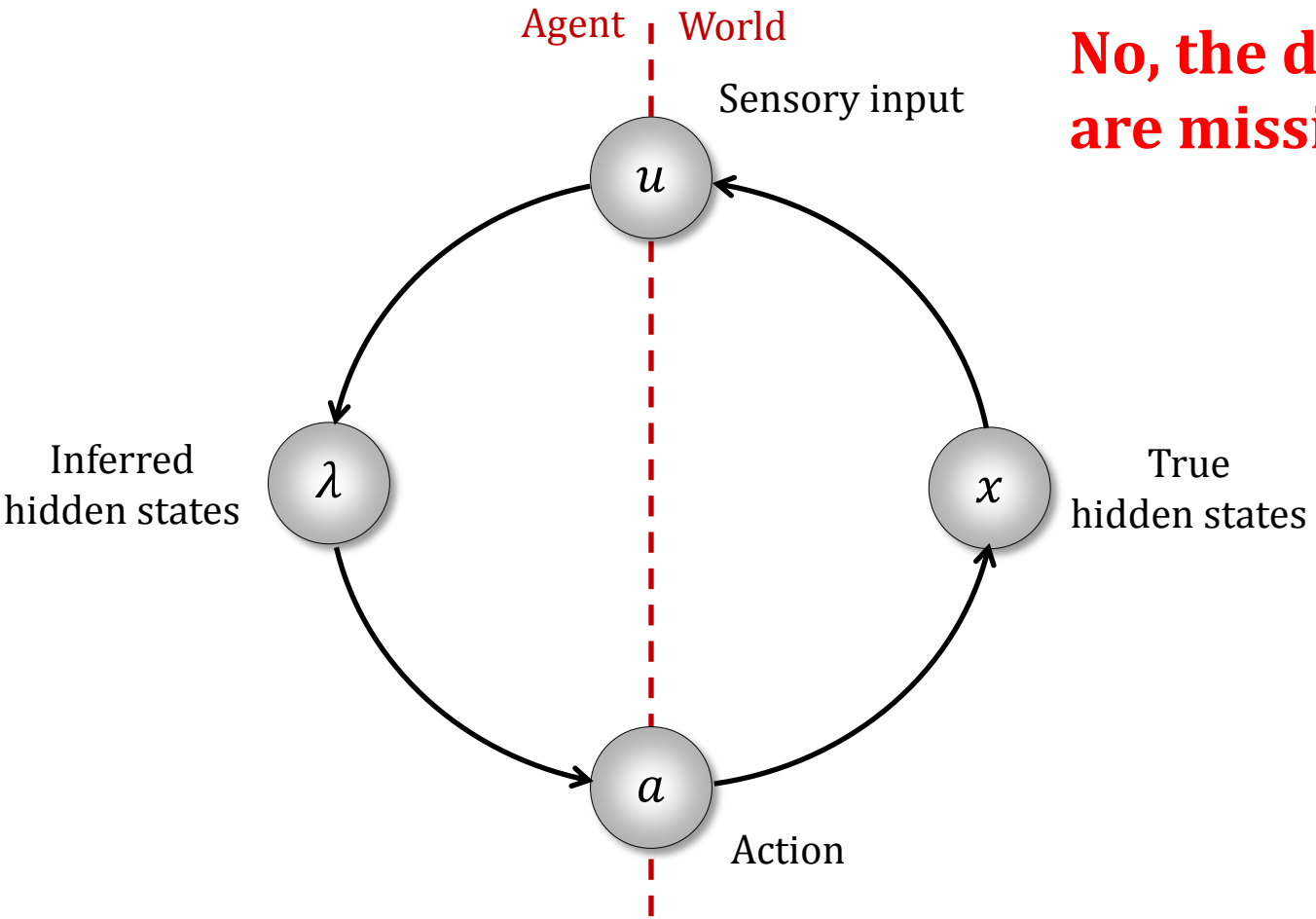
- Now that we have dealt with beliefs about states that are binary (Bernoulli), categorical, bounded on both sides (beta), bounded on one side (gamma), and unbounded (Gaussian), we have covered a very large proportion of states we may have beliefs about.

- **All Bayesian (i.e., probabilistic, rational) updates of such beliefs take the form of precision-weighted prediction errors.**

Limitations

- Examples of distributions that are not exponential families: Student's *t*, Cauchy
- These distributions are popular because of their «fat tails». However, fat tails can also be achieved with appropriate hierarchies of Gaussians (cf. the hierarchical Gaussian filter, HGF)
- A further kind of distributions that are not exponential families are found in mixture models.
- Such models are popular because of they provide multimodal distributions. But again, appropriate hierarchies of distributions may save the day.

Does inference as we've described it adequately describe the situation of actual biological agents?



No, the dynamics are missing!

What about dynamics?

- Up to now, we've only looked at inference on static quantities, but biological agents live in a continually changing world.
- In our example, the boat's position changes and with it the angle to the lighthouse.
- How can we take into account that old information becomes obsolete? If we don't, our learning rate becomes smaller and smaller because our eqations were derived under the assumption that we're accumulating information about a stable quantity.

What's the simplest way to keep the learning rate from going too low?

- Keep it constant!
- So, taking the update equation for the mean of our observations as our point of departure...

$$\bar{x}_n = \bar{x}_{n-1} + \frac{1}{n}(x_n - \bar{x}_{n-1}),$$

- ... we simply replace $\frac{1}{n}$ with a constant α:

$$\mu_n = \mu_{n-1} + \alpha(x_n - \mu_{n-1}).$$

- This is called *Rescorla-Wagner learning* [although it wasn't this line of reasoning that led Rescorla & Wagner (1972) to their formulation].

Does a constant learning rate solve our problems?

- Partly: it implies a certain rate of forgetting because it amounts to taking only the $n = \frac{1}{\alpha}$ last data points into account. But...
- ... if the learning rate is supposed to reflect uncertainty in Bayesian inference, then how do we
- (a) know that α reflects the right level of uncertainty at any one time, and
- (b) account for changes in uncertainty if α is constant?
- What we really need is an adaptive learning that accurately reflects uncertainty.

An adaptive learning rate that accurately reflects uncertainty

- This requires us to think a bit more about what kinds of uncertainty we are dealing with.

- A possible taxonomy of uncertainty is (cf. Yu & Dayan, 2003; Payzan-LeNestour & Bossaerts, 2011):

- (a) **outcome uncertainty** that remains unaccounted for by the model, called *risk* by economists (π_ε in our Bayesian example); this uncertainty remains even when we know all parameters exactly,

- (b) **informational** or *expected* uncertainty about the value of model parameters ($\pi_{\vartheta|x}$ in the Bayesian example),

- (c) **environmental** or *unexpected* uncertainty owing to changes in model parameters (not accounted for in our Bayesian example, hence unexpected).

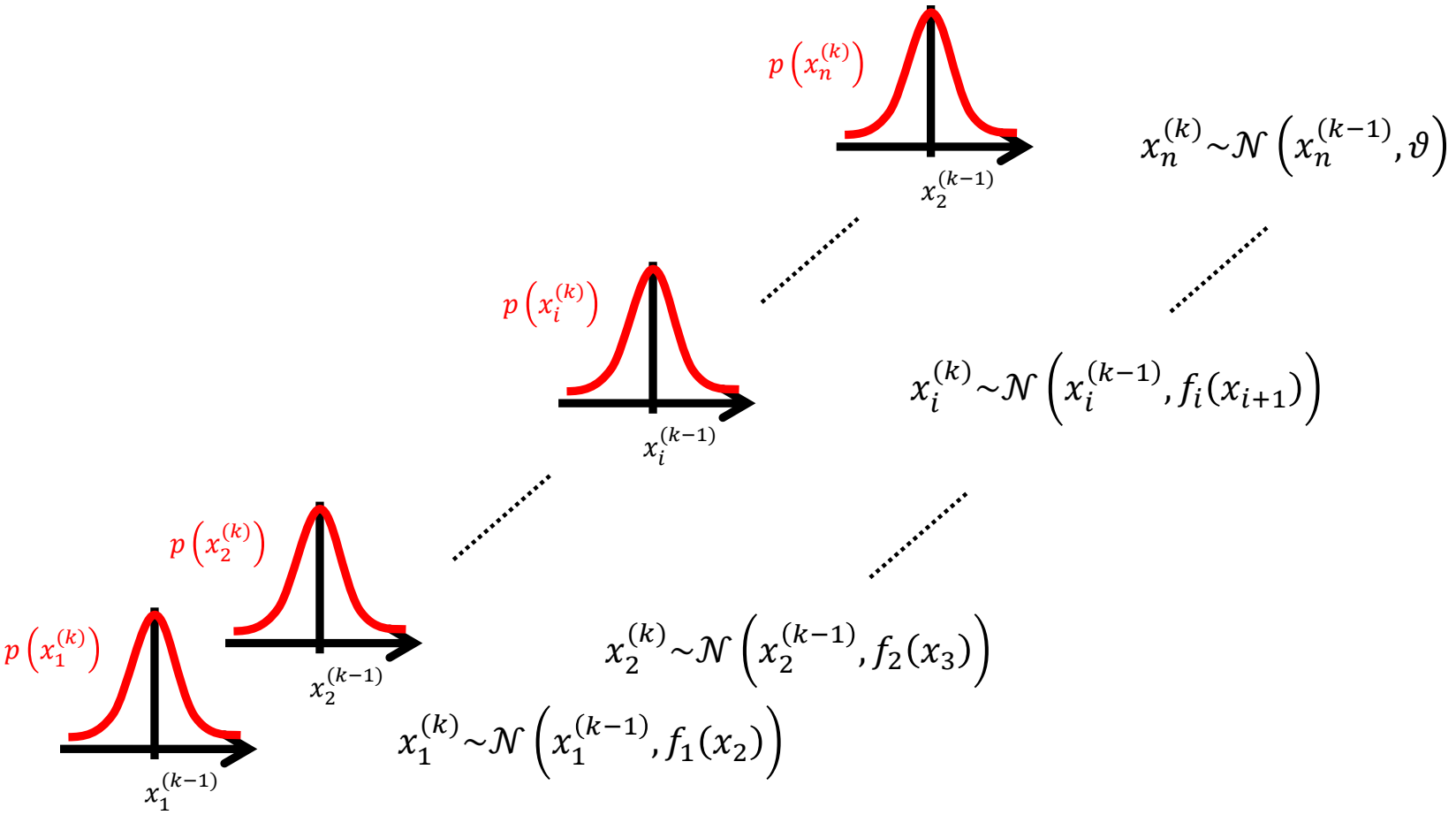
An adaptive learning rate that accurately reflects uncertainty

- Various efforts have been made to come up with an adaptive learning rate:
  - Kalman (1960)
  - Sutton (1992)
  - Nassar et al. (2010)
  - Payzan-LeNestour & Bossaerts (2011)
  - Mathys et al. (2011)
  - Wilson et al. (2013)

- The Kalman filter is optimal for linear dynamical systems, but realistic data usually require non-linear models.

- Mathys et al. use a generic non-linear hierarchical Bayesian model that allows us to derive update equations that are optimal in the sense that they minimize surprise.

The hierarchical Gaussian filter (HGF)



The hierarchical Gaussian filter (HGF)

- At the outcome level (i.e., at the very bottom of the hierarchy), we have

$$u^{(k)} \sim \mathcal{N}\left(x_1^{(k)}, \hat{\pi}_u^{-1}\right)$$

- This gives us the following update for our belief on x_1 (our quantity of interest):

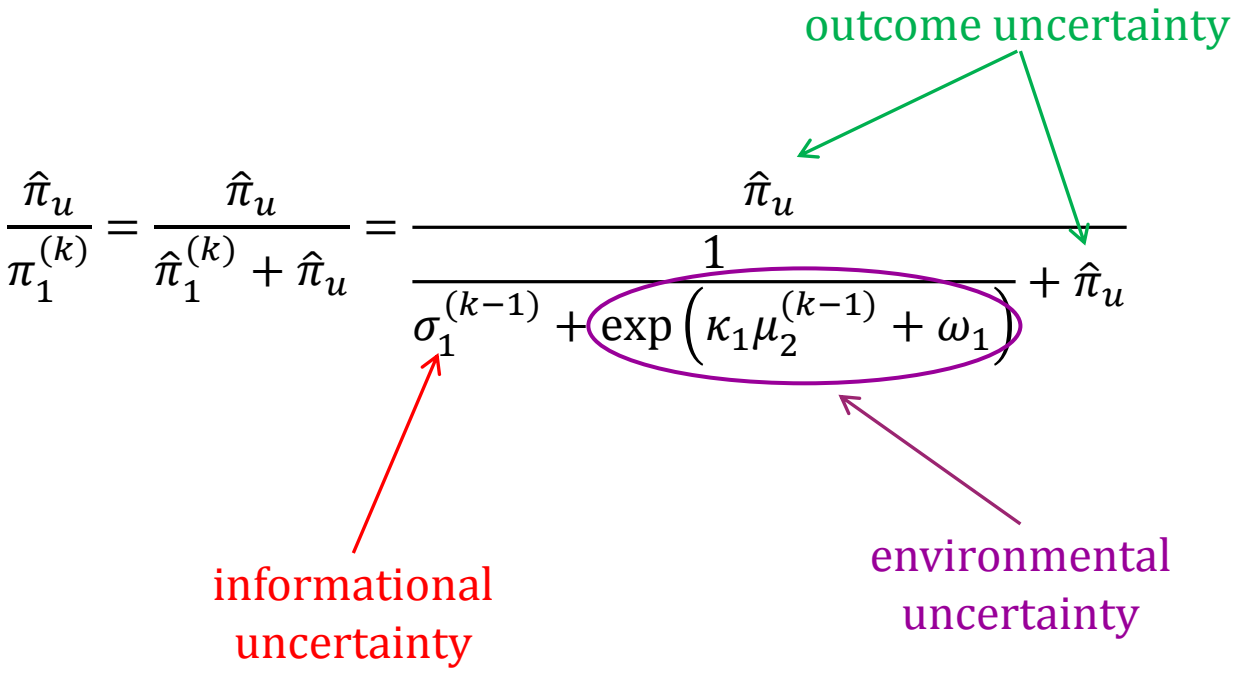
$$\pi_1^{(k)} = \hat{\pi}_1^{(k)} + \hat{\pi}_u$$

$$\mu_1^{(k)} = \mu_1^{(k-1)} + \frac{\hat{\pi}_u}{\pi_1^{(k)}}\left(u^{(k)} - \mu_1^{(k-1)}\right)$$

- The familiar structure again – but now with a learning rate that is responsive to all kinds of uncertainty, including environmental (unexpected) uncertainty.

The learning rate in the HGF

Unpacking the learning rate, we see:



VAPEs and VOPEs

The updates of the belief on x_1 are driven by value prediction errors (VAPEs)

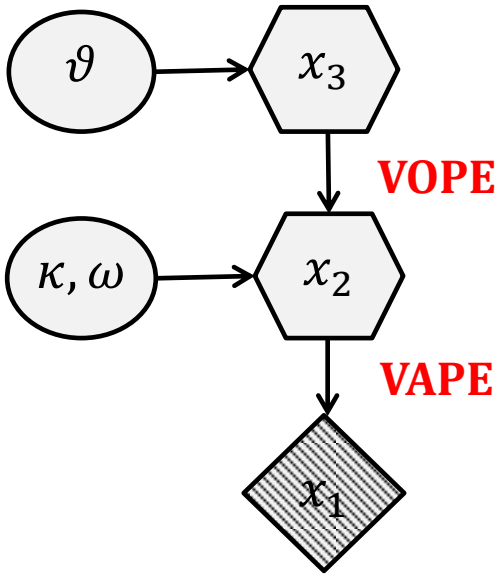
$$\mu_1^{(k)} = \mu_1^{(k-1)} + \frac{\hat{\pi}_u}{\pi_1^{(k)}}\left(u^{(k)} - \mu_1^{(k-1)}\right), \quad \textbf{VAPE}$$

while the x_2-updates are driven by volatility prediction errors (VOPEs)

$$\mu_2^{(k)} = \mu_2^{(k-1)} + \frac{1}{2}\kappa_1 v_1^{(k)} \frac{\hat{\pi}_1^{(k)}}{\pi_2^{(k)}} \delta_1^{(k)} \quad \textbf{VOPE}$$

$$\delta_1^{(k)} \stackrel{\text{def}}{=} \frac{\sigma_1^{(k)} + \left(\mu_1^{(k)} - \mu_1^{(k-1)}\right)^2}{\sigma_1^{(k-1)} + \exp\left(\kappa_1 \mu_2^{(k-1)} + \omega_1\right)} - 1$$

3-level HGF for binary observations



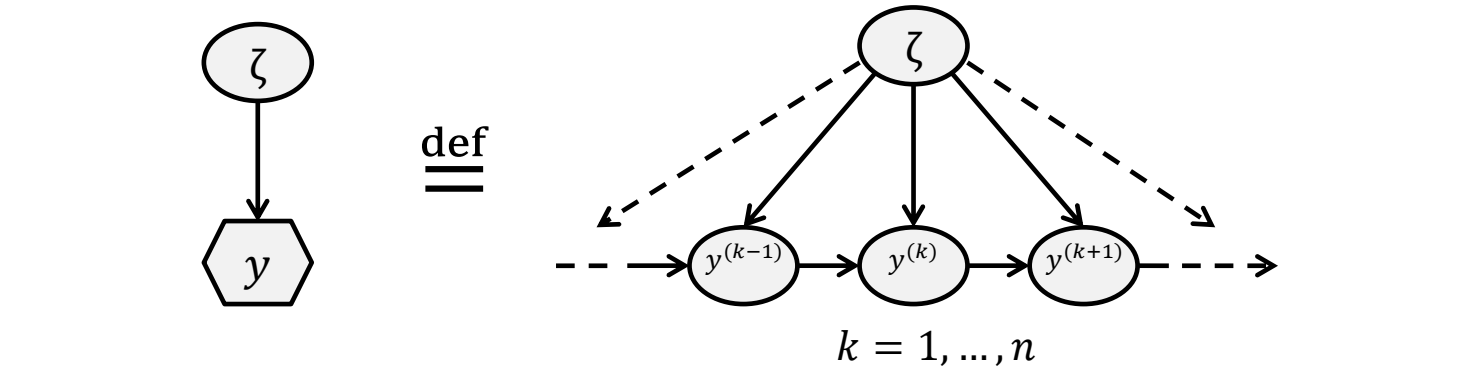
$$x_3^{(k)} \sim \mathcal{N}\left(x_3^{(k-1)}, \vartheta\right)$$

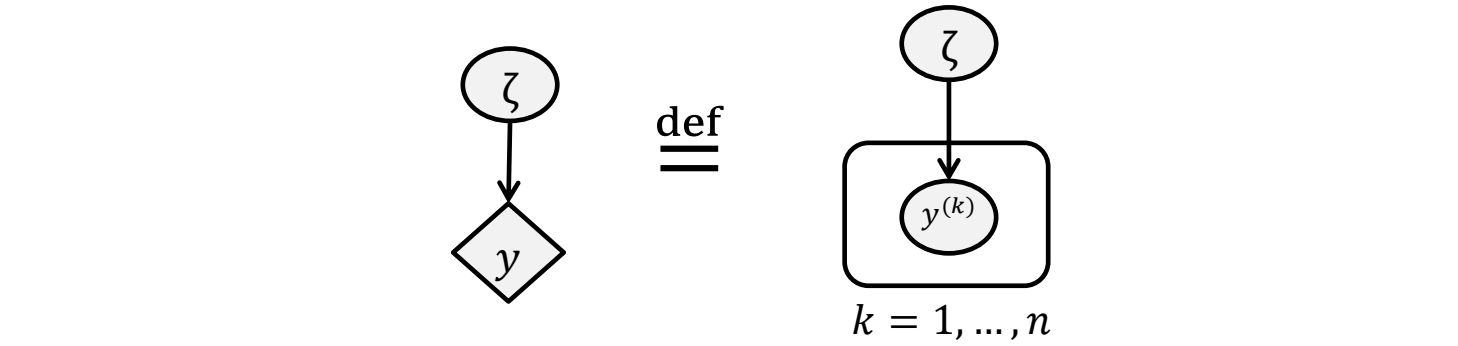
$$x_2^{(k)} \sim \mathcal{N}\left(x_2^{(k-1)}, \exp\left(\kappa x_3^{(k)} + \omega\right)\right)$$

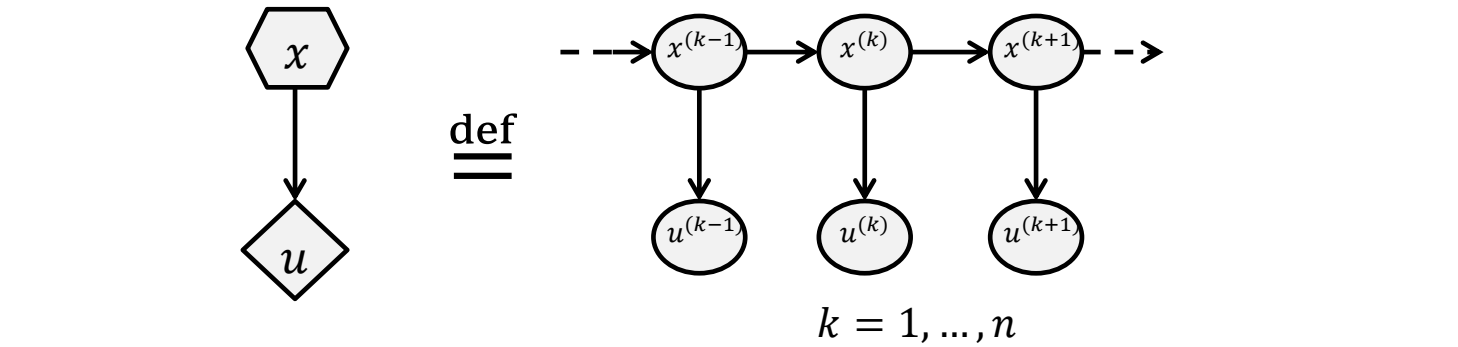
$$x_1^{(k)} \sim \text{Bern}\left(x_2^{(k)}\right)$$

Mathys et al., 2011; Iglesias et al., 2013; Vossel et al., 2014a; Hauser et al., 2014; Diaconescu et al., 2014; Vossel et al., 2014b; ...

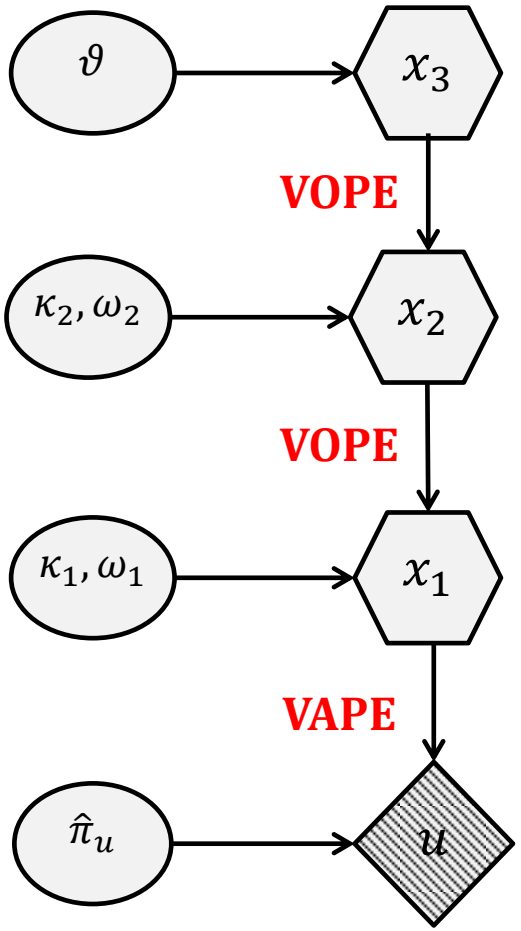
Taking it all together: notation







3-level HGF for continuous observations



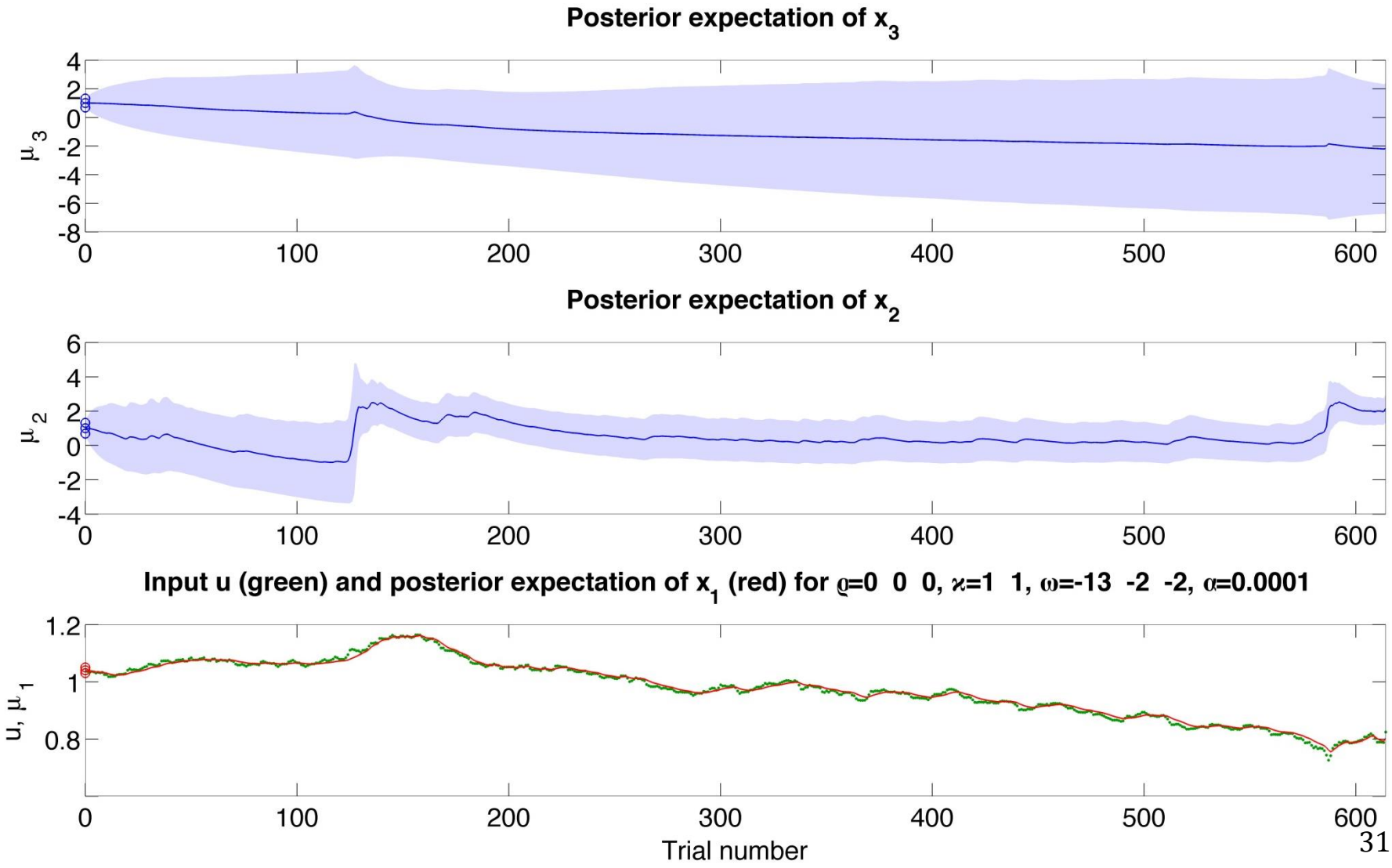
$$x_3^{(k)} \sim \mathcal{N}\left(x_3^{(k-1)}, \vartheta\right)$$

$$x_2^{(k)} \sim \mathcal{N}\left(x_2^{(k-1)}, \exp\left(\kappa_2 x_3^{(k)} + \omega_2\right)\right)$$

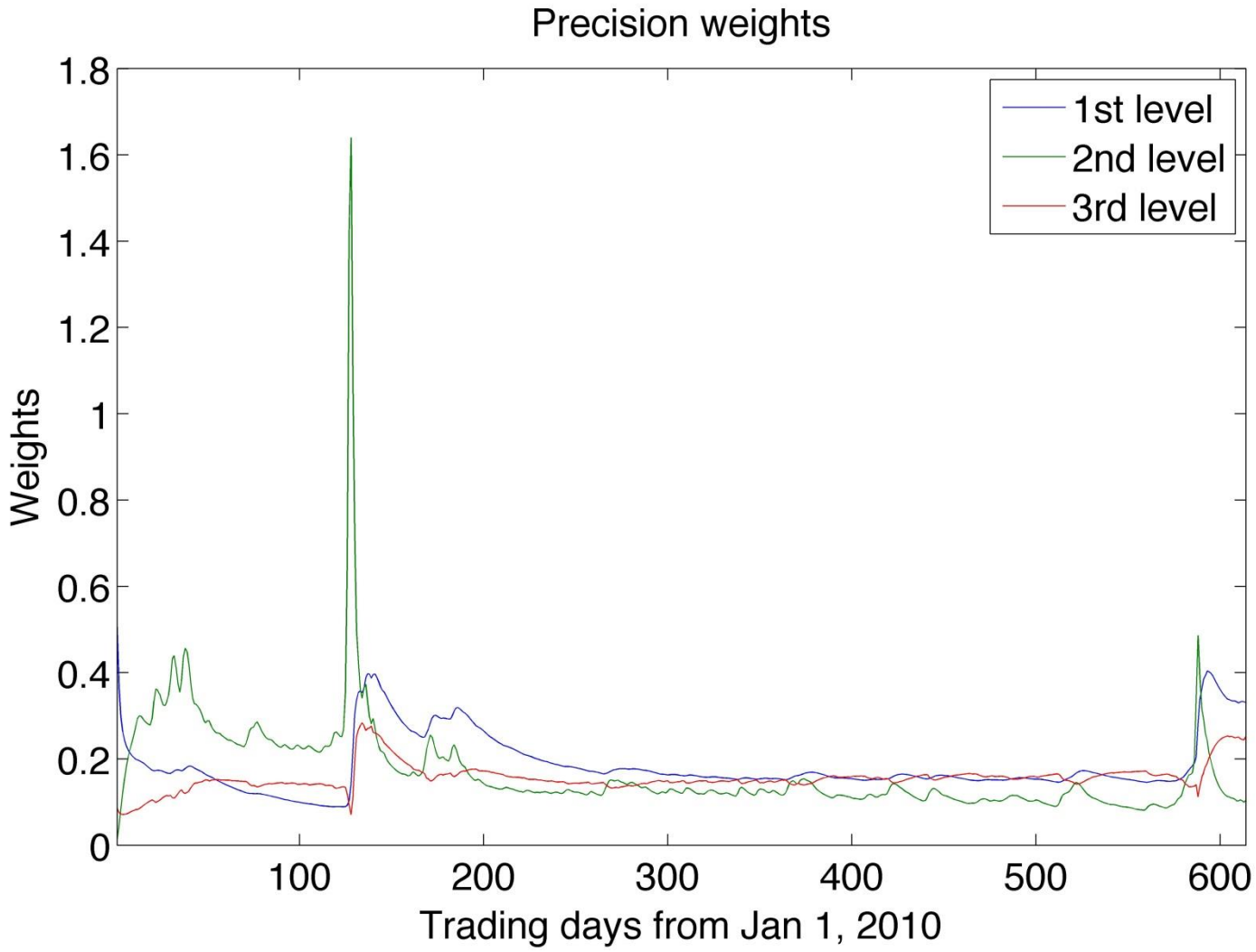
$$x_1^{(k)} \sim \mathcal{N}\left(x_1^{(k-1)}, \exp\left(\kappa_1 x_2^{(k)} + \omega_1\right)\right)$$

$$u^{(k)} \sim \mathcal{N}\left(x_1^{(k)}, \hat{\pi}_u^{-1}\right)$$

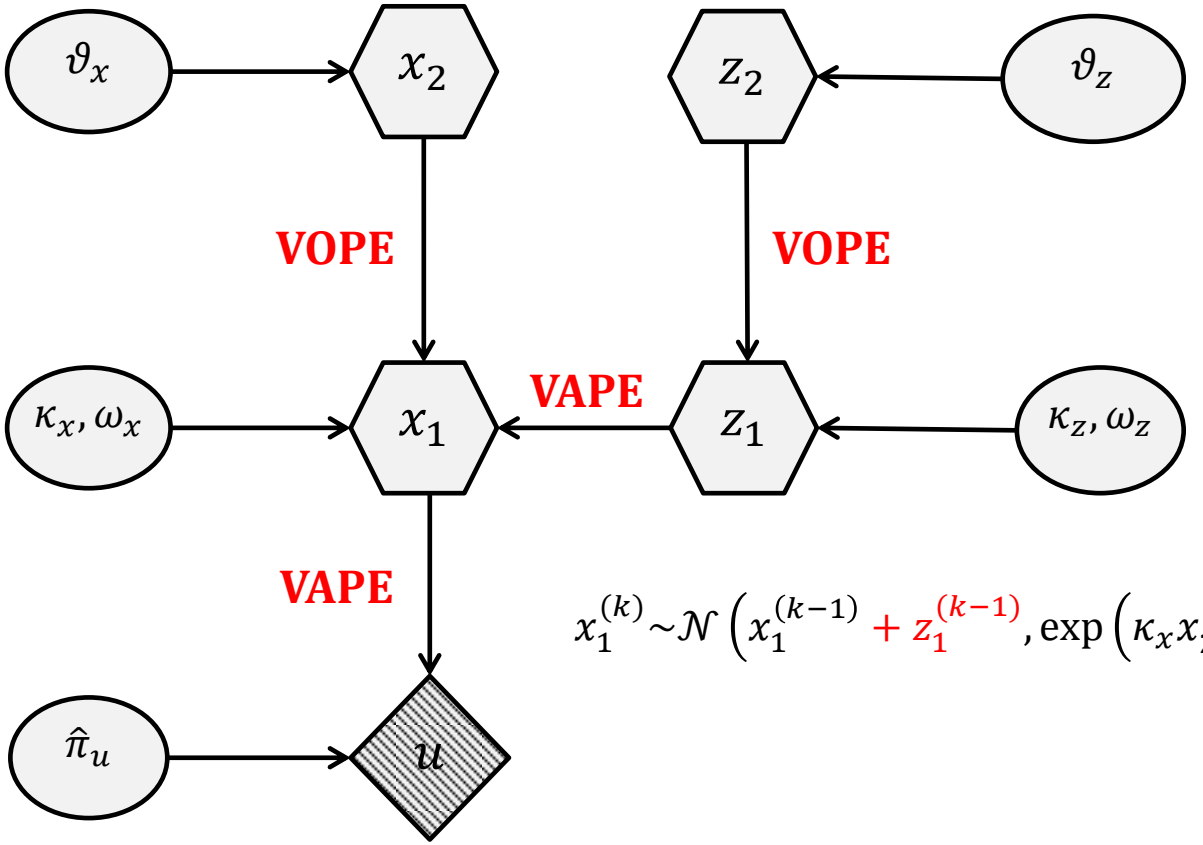
3-level HGF for continuous observations



3-level HGF for continuous observations

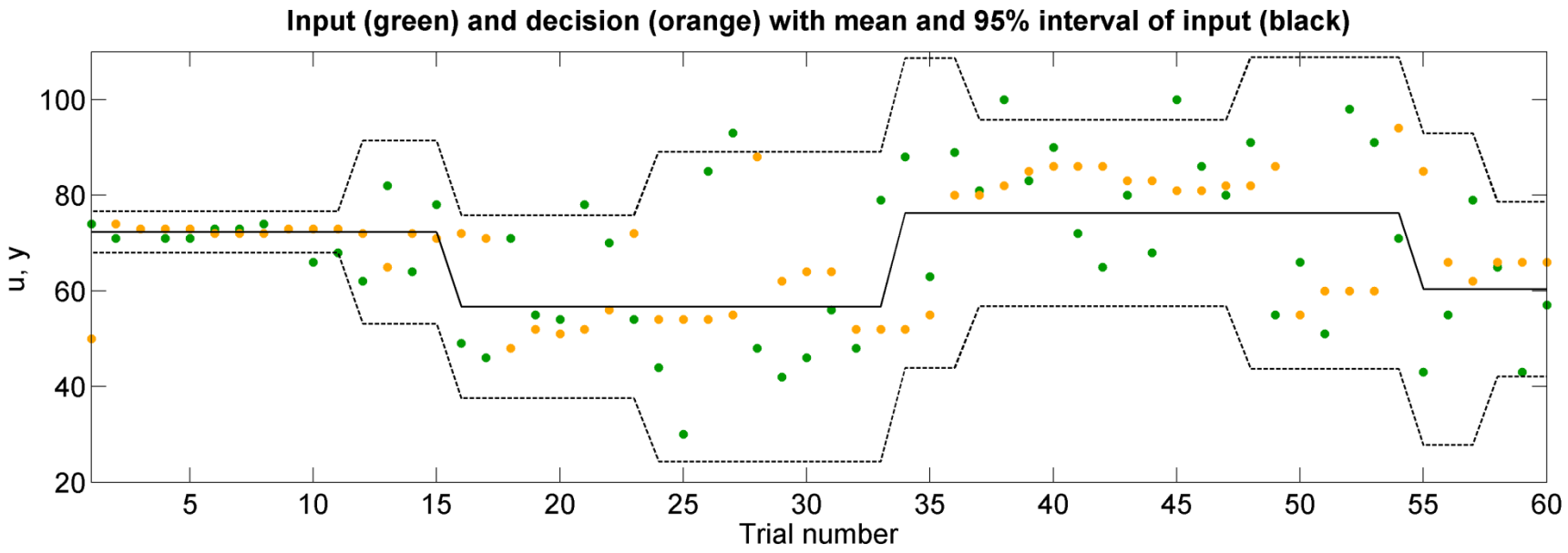


Variable drift



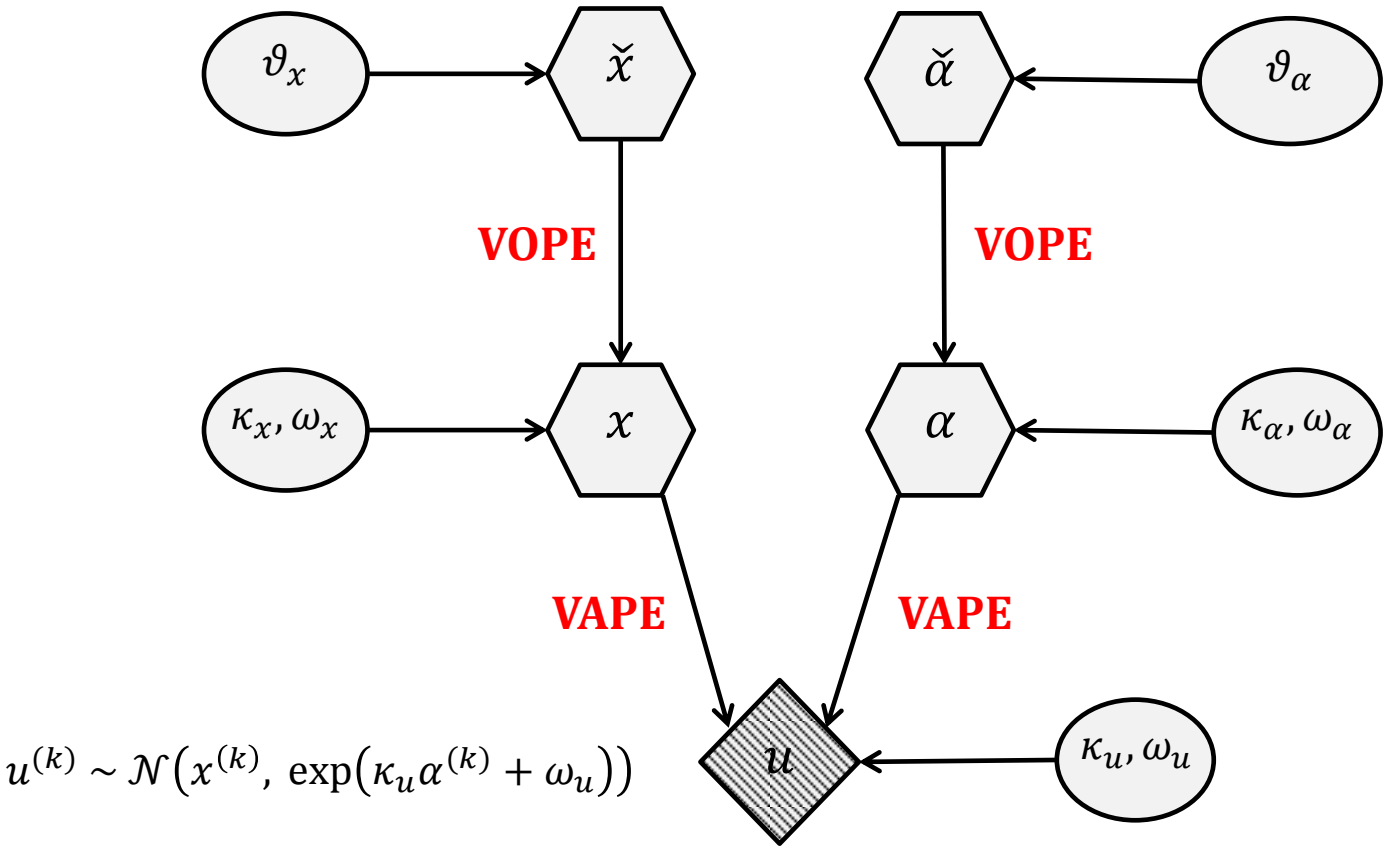
$$x_1^{(k)} \sim \mathcal{N}\left(x_1^{(k-1)} + z_1^{(k-1)}, \exp\left(\kappa_x x_2^{(k)} + \omega_x\right)\right)$$

Jumping Gaussian estimation task

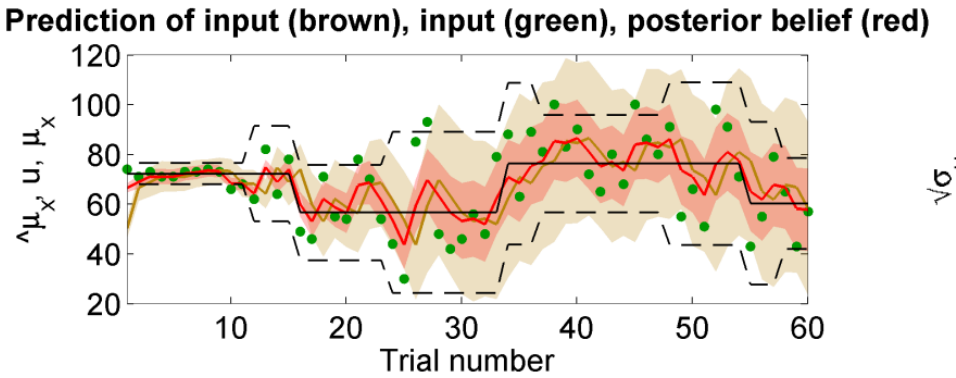


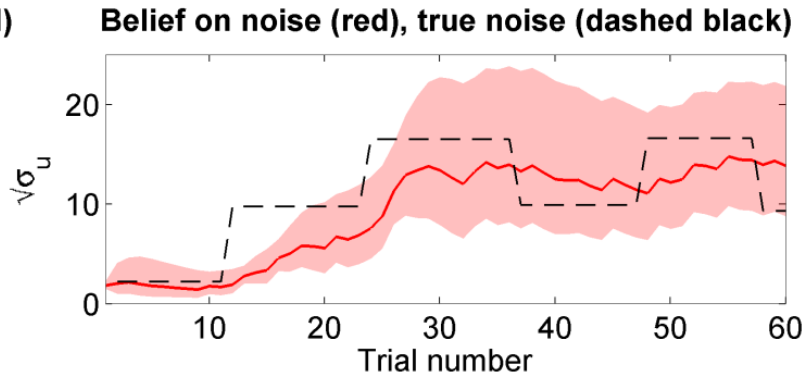
Chaohui Guo

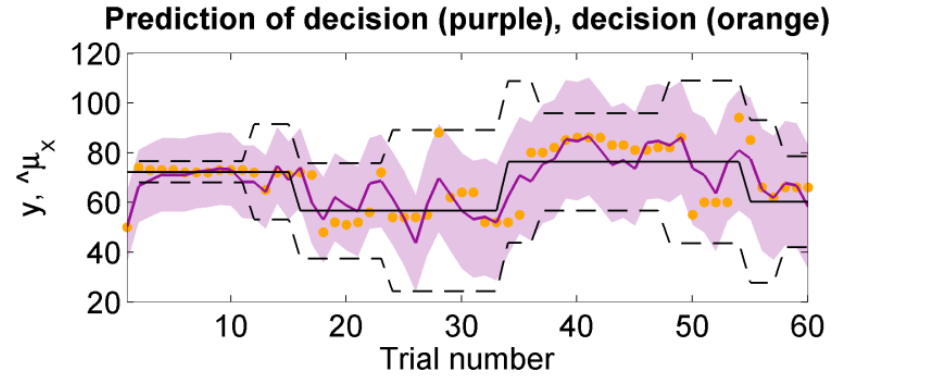
Independent mean and variance model

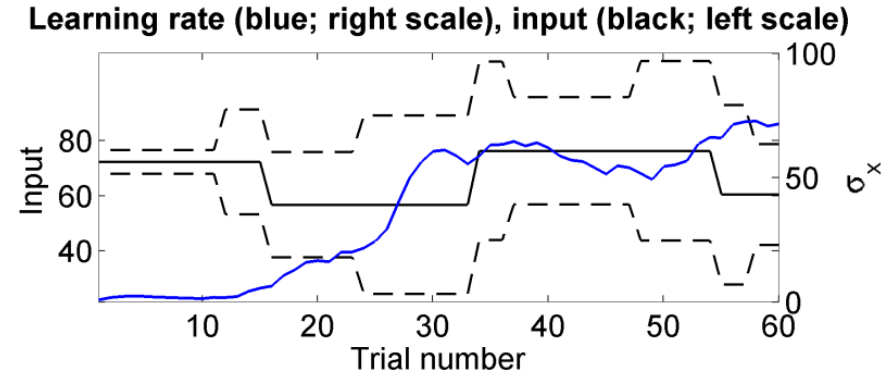


Jumping Gaussian estimation task

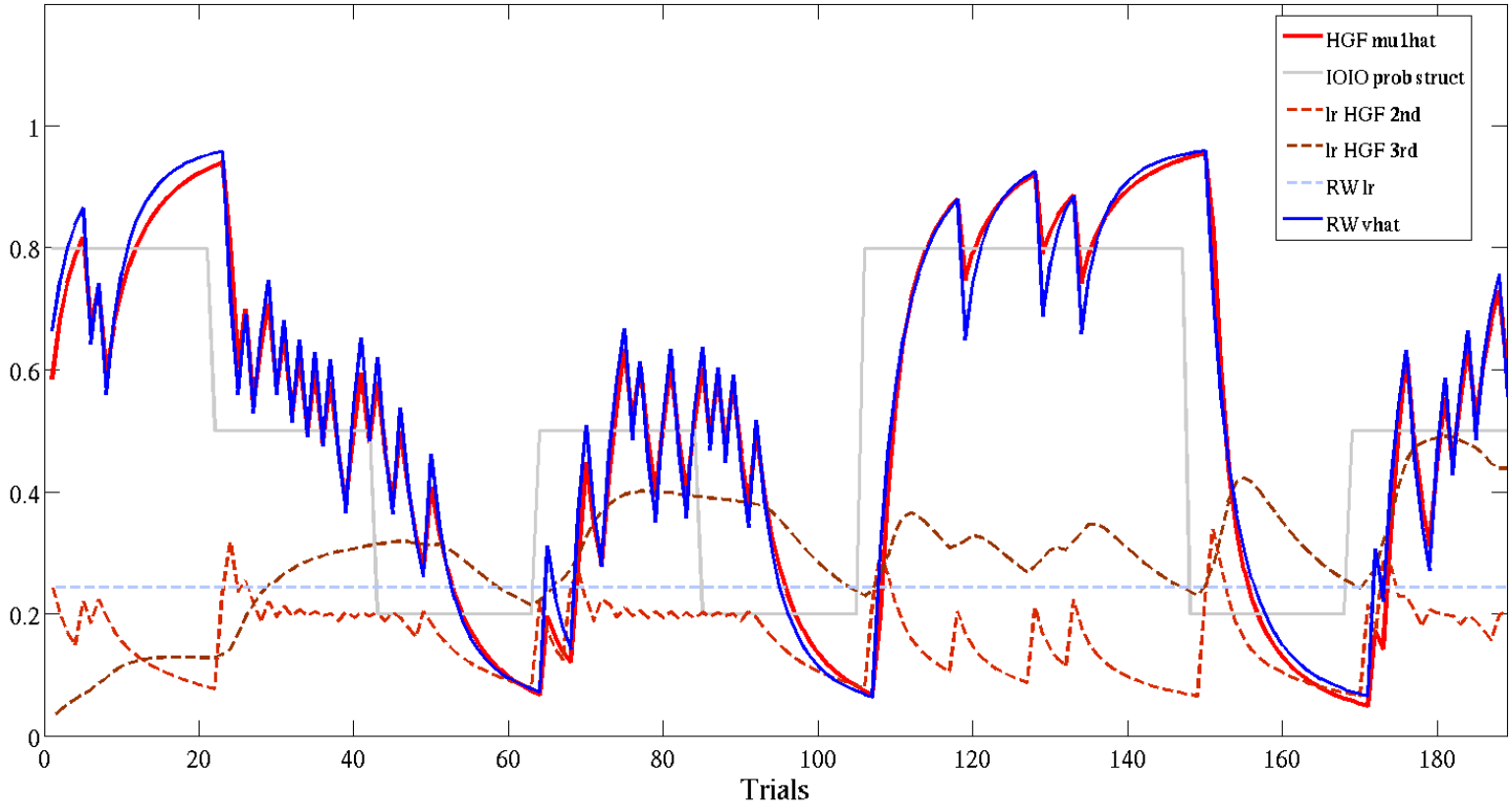








The learning rate in the HGF



Andreea Diaconescu

HGF: empirical evidence (Iglesias et al., 2013)

Model comparison:

	Behavioral study		fMRI study 1		fMRI study 2	
BMS results	PP	XP	PP	XP	PP	XP
HGF1	0.8435	1	0.7422	1	0.7166	1
HGF2	0.0259	0	0.0200	0	-	-
HGF3	0.0361	0	0.1404	0	0.1304	0
Sutton	0.0685	0	0.0710	0	0.0761	0
RW	0.0260	0	0.0264	0	0.0769	0

HGF: empirical evidence (Iglesias et al., 2013)

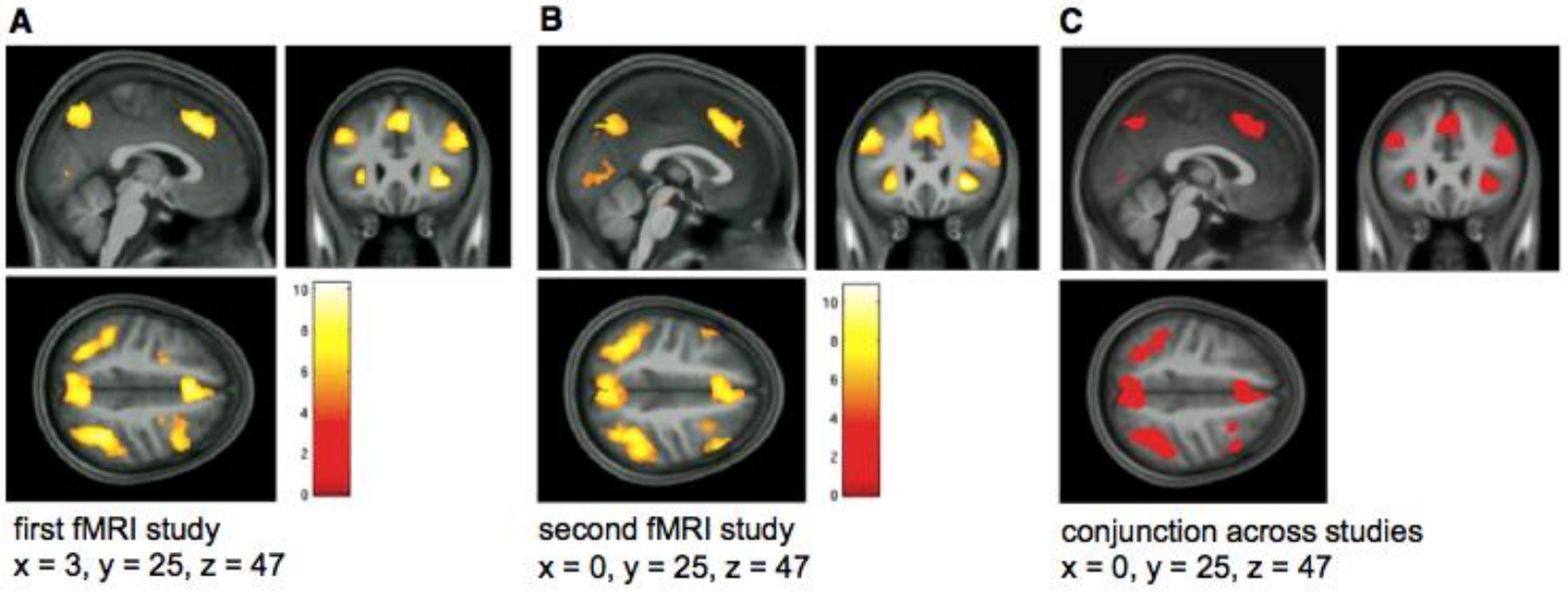


Figure 2. Whole-Brain Activations by ε_2

Activations by precision-weighted prediction error about visual stimulus outcome, ε_2, in the first fMRI study (A) and the second fMRI study (B). Both activation maps are shown at a threshold of $p < 0.05$, FWE corrected for multiple comparisons across the whole brain. To highlight replication across studies, (C) shows the results of a "logical AND" conjunction, illustrating voxels that were significantly activated in both studies.

HGF: empirical evidence (Iglesias et al., 2013)

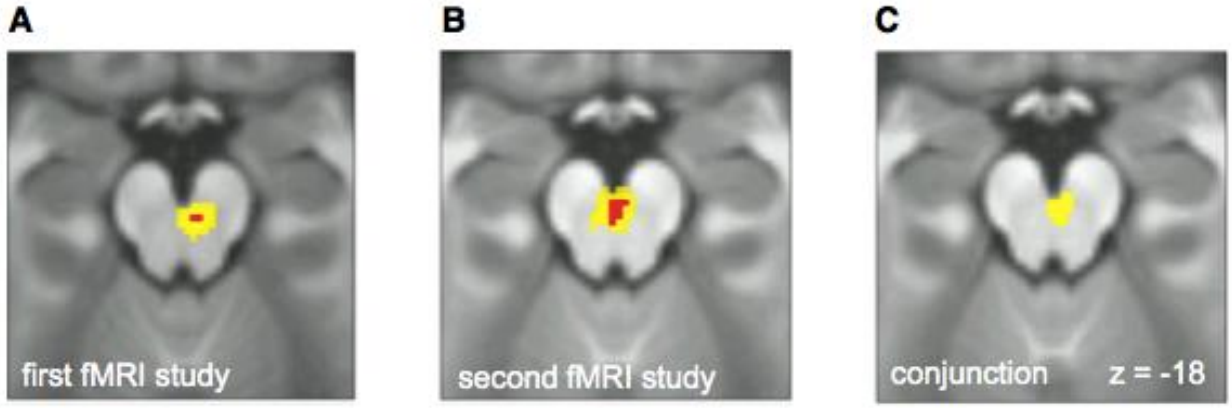


Figure 3. Midbrain Activation by ε_2
Activation of the dopaminergic VTA/SN associated with precision-weighted prediction error about stimulus category, ε_2. This activation is shown both at $p < 0.05$ FWE whole-brain corrected (red) and $p < 0.05$ FWE corrected for the volume of our anatomical mask comprising both dopaminergic and cholinergic nuclei (yellow).
(A) Results from the first fMRI study.
(B) Second fMRI study.
(C) Conjunction (logical AND) across both studies.

HGF: empirical evidence (Iglesias et al., 2013)

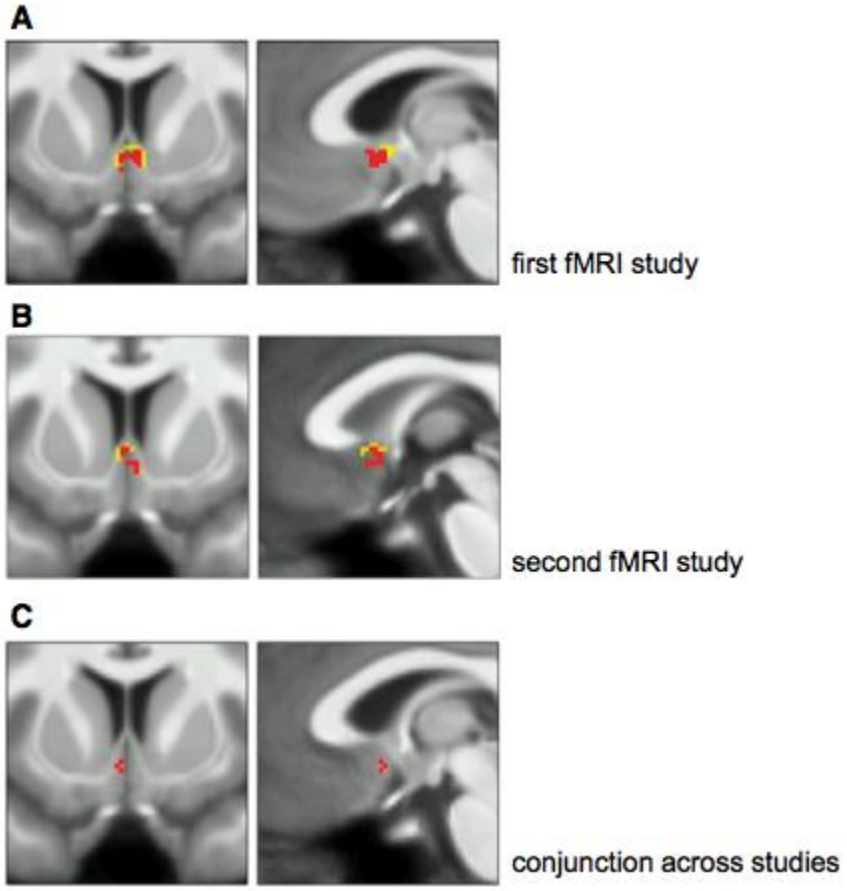


Figure 6. Basal Forebrain Activations by ε_3
Activation of the cholinergic basal forebrain associated with precision-weighted prediction error about stimulus probabilities ε_3 within the anatomically defined mask. For visualization of the activation area we overlay the results thresholded at $p < 0.05$ FWE corrected for the entire anatomical mask (red) on the results thresholded at $p < 0.001$ uncorrected (yellow) in the first (A: $x = 3, y = 9, z = -8$) and the second fMRI study (B: $x = 0, y = 10, z = -8$). (C) The conjunction analysis ("logical AND") across both studies ($x = 2, y = 11, z = -8$).

How to estimate and compare models: the HGF Toolbox

- Available at

  http://www.translationalneuromodeling.org/tapas

- Start with README, manual, and interactive demo
- Modular, extensible
- Matlab-based

Summary

- Bayesian inference (and free energy minimization) imply the application of a canonical kind of belief update: the uncertainty- (i.e., precision-) weighted prediction error.

- Precision-weighting has to take account of all forms of uncertainty, including about precision (or volatility) itself.

- Updates take two forms: VAPEs (value prediction error-driven) and VOPEs (volatility prediction error-driven)

- Evidence from neuroimaging indicates that the brain processes VOPEs as well as VAPEs.

Thanks

- Rick Adams
- Kay Brodersen
- Jean Daunizeau
- Andreea Diaconescu
- Chaohui Guo
- Karl Friston
- Sandra Iglesias
- Lars Kasper
- Ekaterina Lomakina
- Saee Paliwal
- Klaas Enno Stephan
- Simone Vossel
- Lilian Weber