

Generative models of aberrant cognition

Christoph Mathys



Colloquium

Department of Cognition, Behaviour, and Research Methods

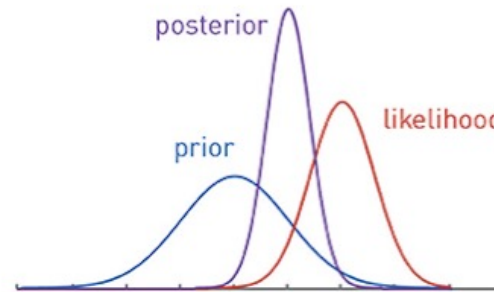
University of Vienna

17 June 2021

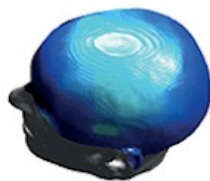
Computational modelling and the inferential brain

Bayes' Theorem

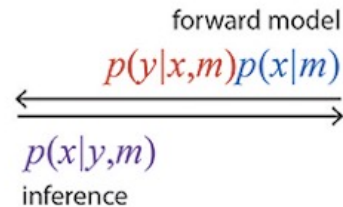
$$p(x|y,m) = \frac{p(y|x,m)p(x|m)}{p(y|m)}$$



Generative models as computational assays

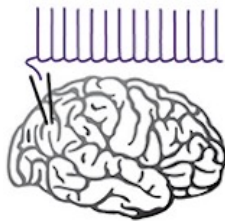


measurement y

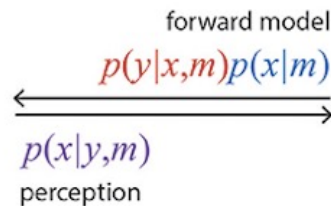


hidden system state x

Perception as the inversion of a generative model



neuronal activity y

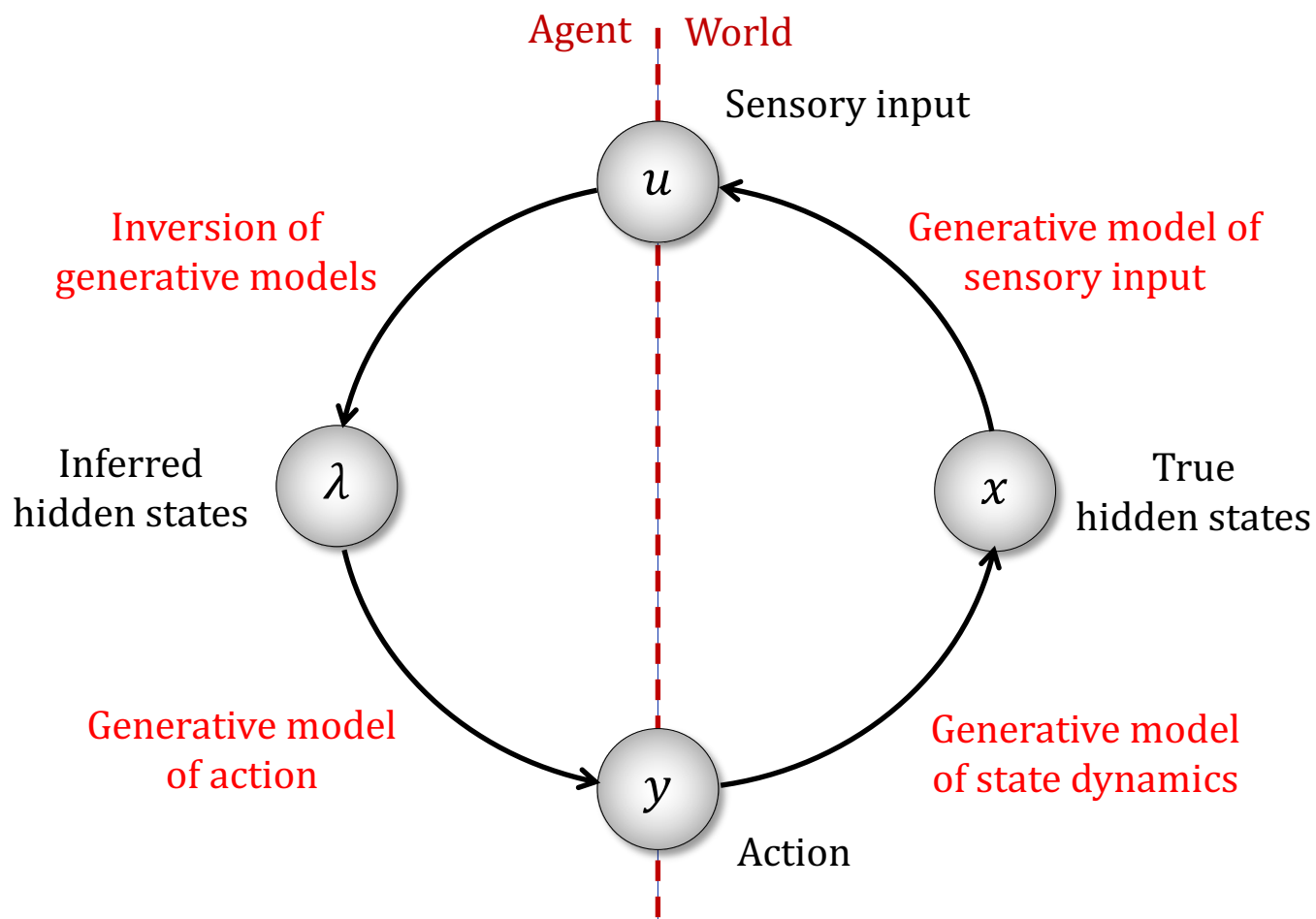


environmental state x

Generative models

- *Generative models* contain all the information needed to **generate observations**.
- This allows for model-based **simulation** of (synthetic) data.
- Advantage 1: A **formal** (i.e. mathematical), **quantitative**, and **consistent** description of the phenomenon of interest
- Advantage 2: Can be calibrated by **measurement**, which in principle allows for consistent and reliable **diagnosis**
- Advantage 3: Allows for **predictive simulation**
- Disadvantage: **Reductionist**. The model will not be as rich as the real phenomenon. (“You want to measure the soul?!”)
- Alternative 1: *Qualitative description*. Captures the richness of the phenomenon much better, but has **no guarantee of consistency or reliability**.
- Alternative 2: *Discriminative modelling*. Often has better predictive performance than generative modelling but **doesn’t give the mechanistic insight that’s needed to identify targets for intervention**.

Generative models of an agent's interaction with its environment



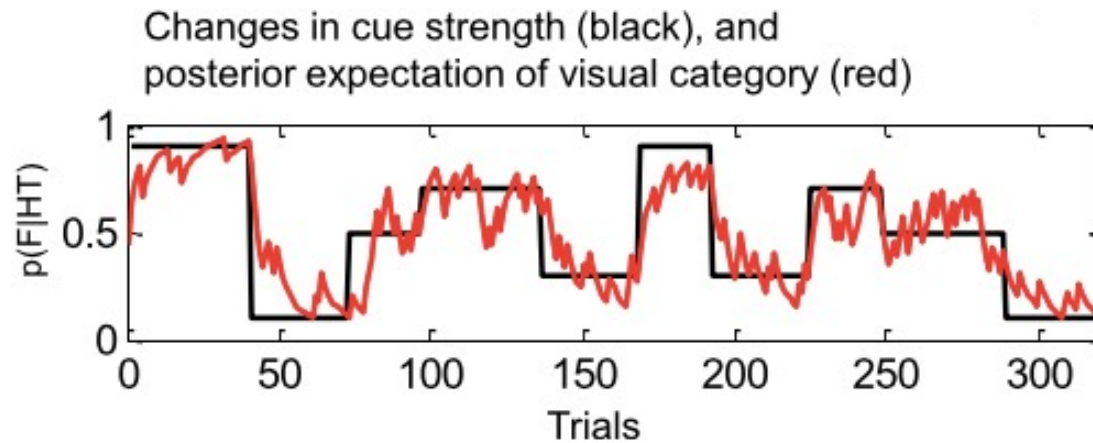
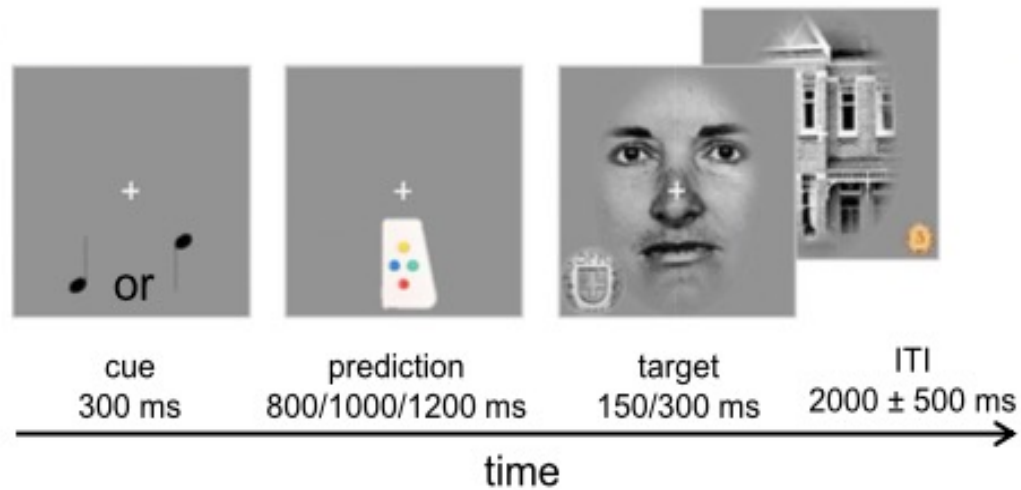
Hierarchical Prediction Errors in Midbrain and Basal Forebrain during Sensory Learning

Sandra Iglesias,^{1,2,*} Christoph Mathys,^{1,2} Kay H. Brodersen,^{1,2} Lars Kasper,^{1,2} Marco Piccirelli,² Hanneke E.M. den Ouden,³ and Klaas E. Stephan^{1,2,4}

In Bayesian brain theories, hierarchically related prediction errors (PEs) play a central role for predicting sensory inputs and inferring their underlying causes, e.g., the probabilistic structure of the environment and its volatility. Notably, PEs at different hierarchical levels may be encoded by different neuromodulatory transmitters. Here, we tested this possibility in computational fMRI studies of audio-visual learning. Using a hierarchical Bayesian model, we found that low-level PEs about visual stimulus outcome were reflected by widespread activity in visual and supra-modal areas but also in the midbrain. In contrast, high-level PEs about stimulus probabilities were encoded by the basal forebrain. These findings were replicated in two groups of healthy volunteers. While our fMRI measures do not reveal the exact neuron types activated in midbrain and basal forebrain, they suggest a dichotomy between neuromodulatory systems, linking dopamine to low-level PEs about stimulus outcome and acetylcholine to more abstract PEs about stimulus probabilities.

Neuron 80, 519–530, October 16, 2013

Task

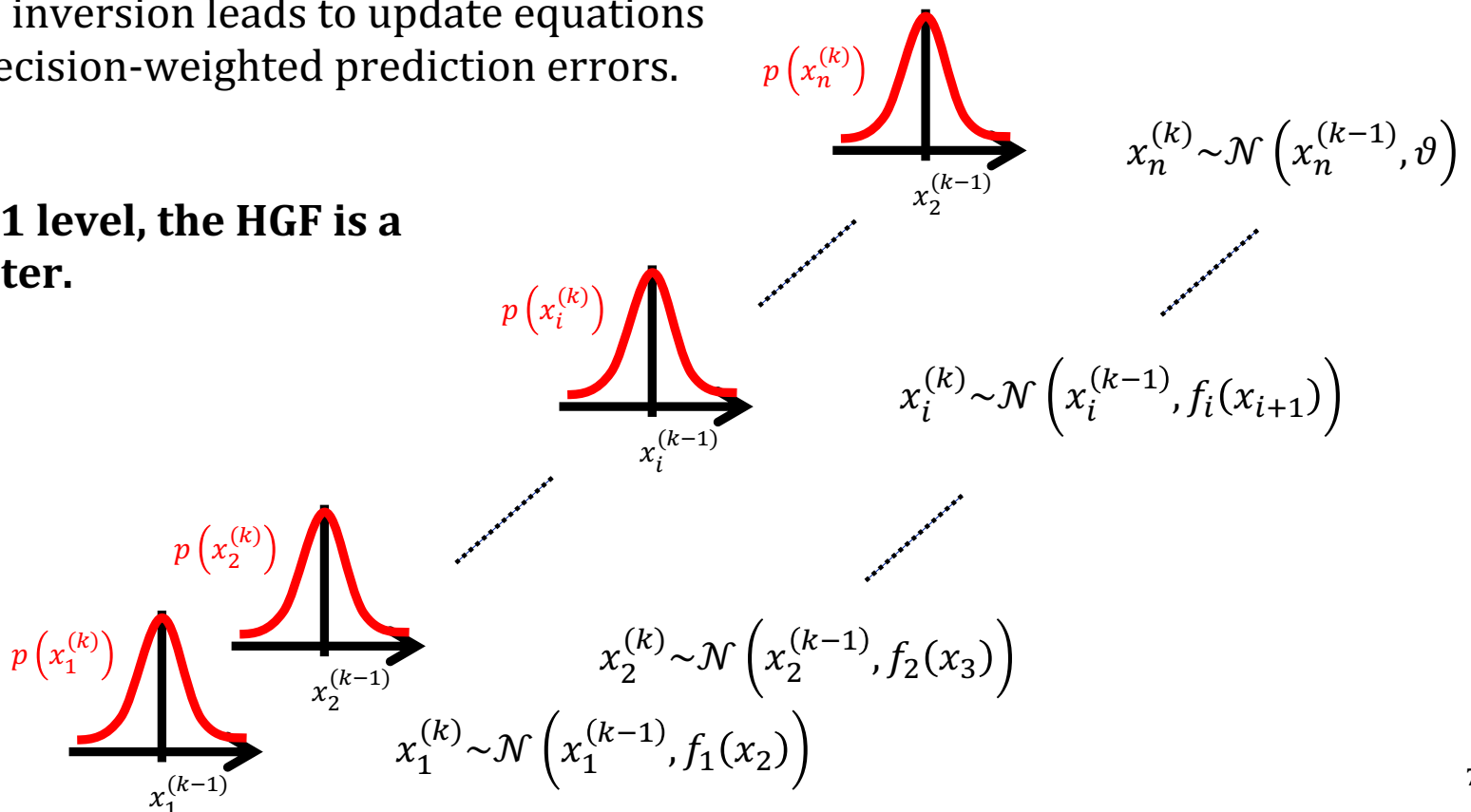


Example model: hierarchical Gaussian filter (HGF, Mathys et al., 2011; 2014)

The HGF provides a generic solution to the problem of adapting one's learning rate in a volatile environment.

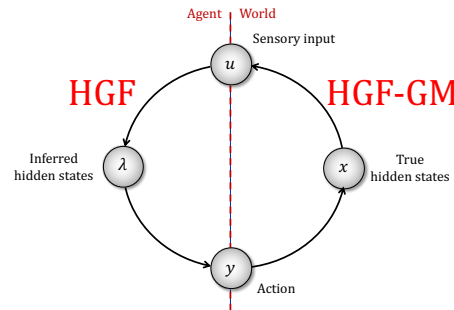
Variational inversion leads to update equations that are precision-weighted prediction errors.

With only 1 level, the HGF is a Kalman filter.



Variational inversion and update equations

- Important distinction: **generative model** (HGF-GM) vs its inversion, **inference model** (HGF proper)



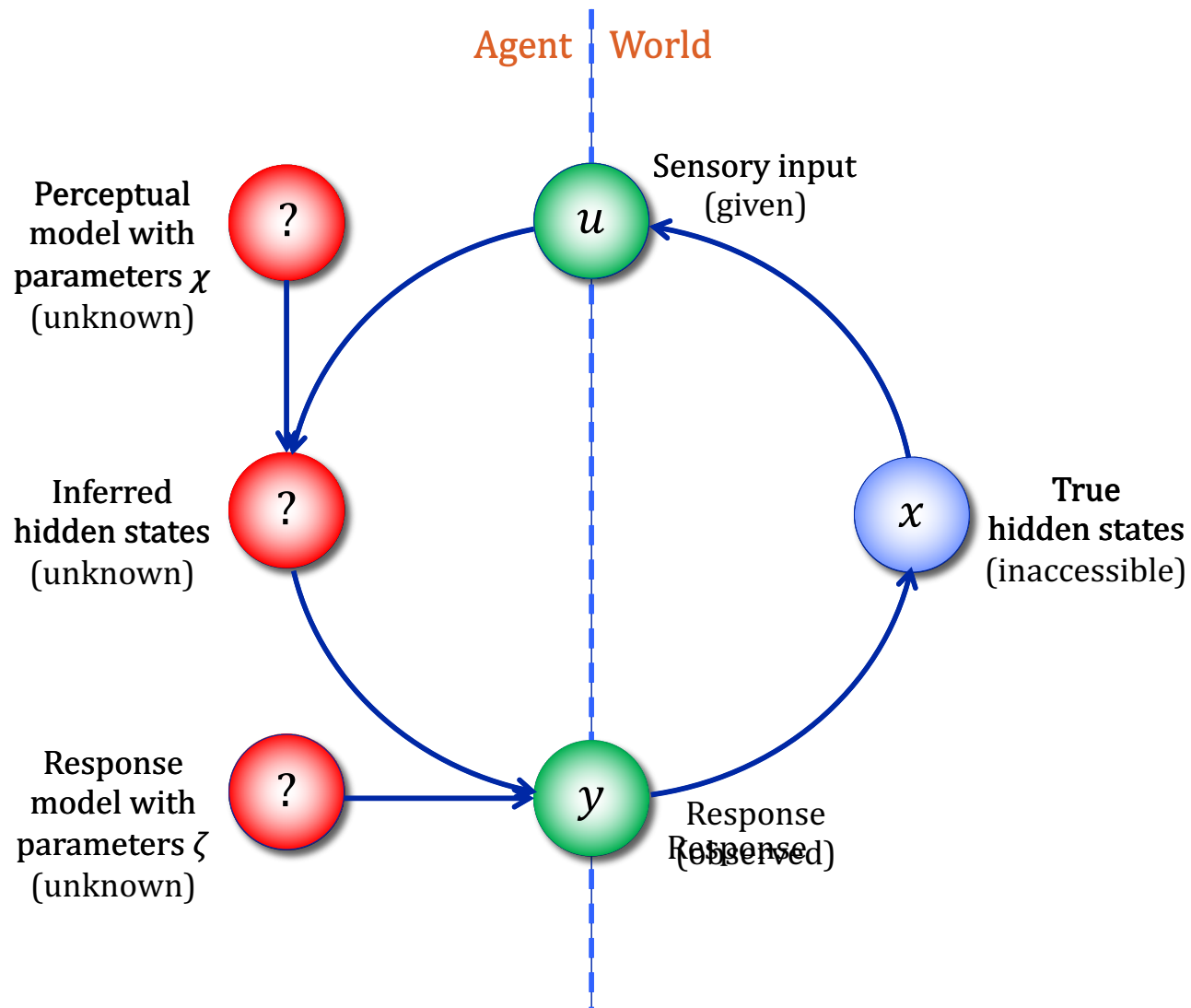
- Inversion of HGF-GM proceeds by introducing a mean field approximation and fitting quadratic approximations to the resulting variational energies (Mathys et al., 2011).
- This leads to **simple one-step update equations** (HGF proper) for the sufficient statistics (mean and precision) of the approximate Gaussian posteriors of the states x_i .
- The updates of the means have the same structure as value updates in Rescorla-Wagner learning:

$$\Delta\mu_i \propto \frac{\hat{\pi}_{i-1}}{\pi_i} \delta_{i-1}$$

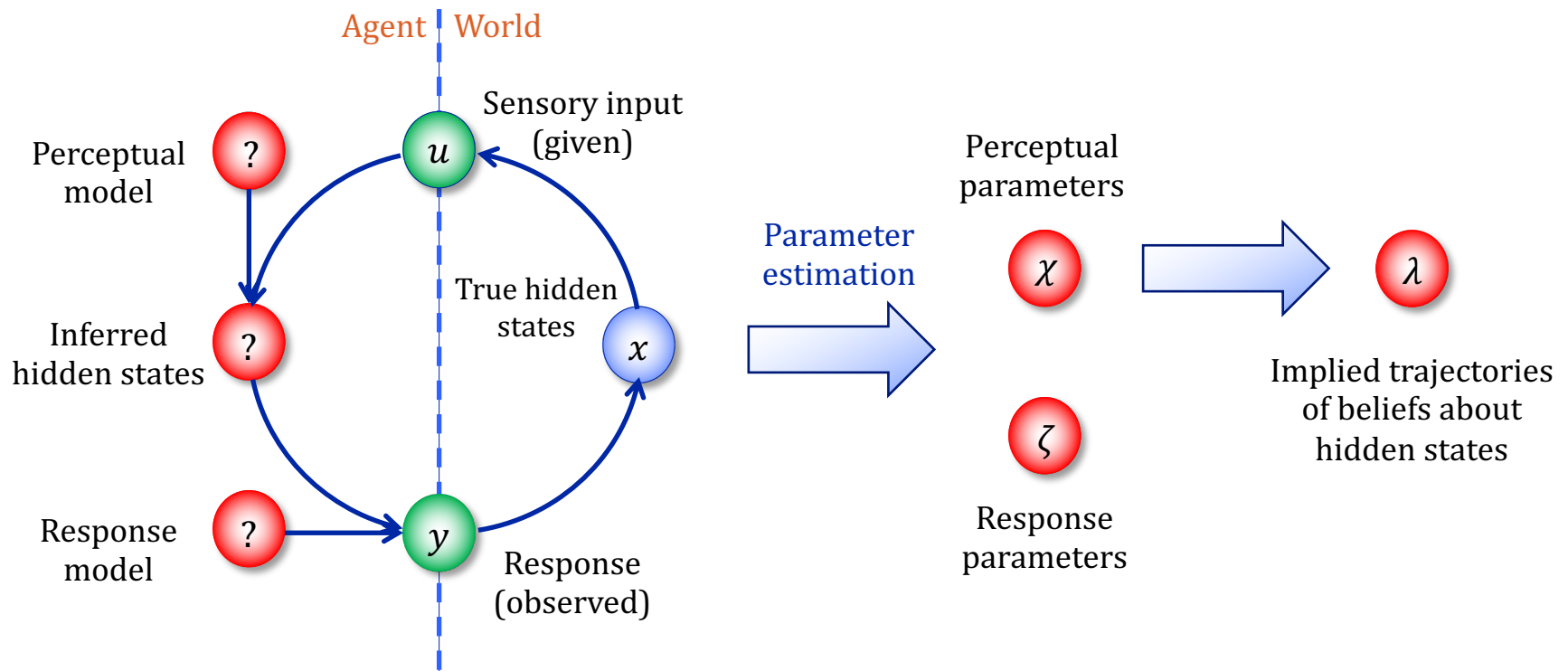
Precisions determine learning rate
Prediction error

- The updates are **precision-weighted prediction errors**.

Application to experimental data

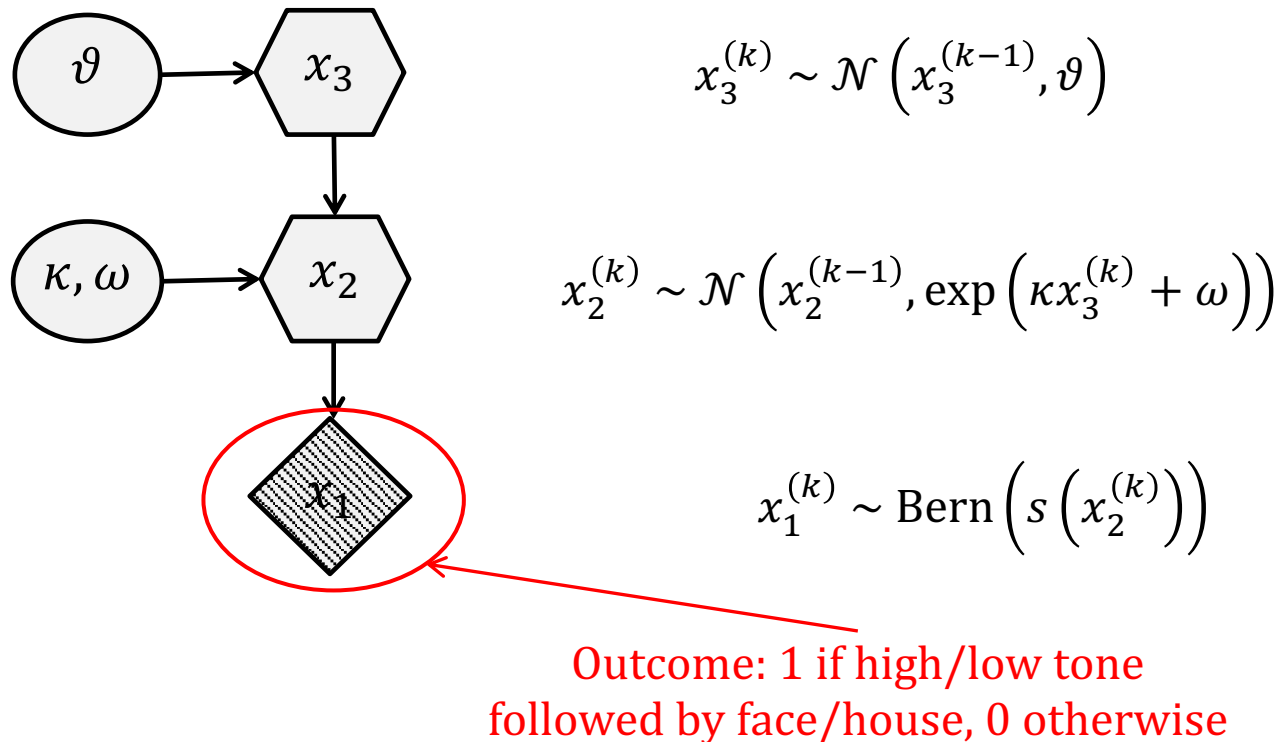


Application to experimental data: parameter estimation



Generative model for trial outcomes in the Iglesias et al. task:

3-level HGF for binary observations



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

Generative model, perceptual model, and decision model

- The *generative model* describes (probabilistically) how the states x_i evolve in time: $x_i^{(k-1)} \rightarrow x_i^{(k)}$
- Variationally inverting the generative gives us the *perceptual model* (also called *inference model* or *recognition model*)
- The perceptual model describes (deterministically, via update equations) how *beliefs* $\{\mu_i, \pi_i\}$ about states evolve in time: $\{\mu_i^{(k-1)}, \pi_i^{(k-1)}\} \rightarrow \{\mu_i^{(k)}, \pi_i^{(k)}\}$
- The decision model (also called observation model or response model) describes (probabilistically) how beliefs are translated into observed actions:
$$\left\{ \mu_i^{(k)}, \pi_i^{(k)} \right\}_{i=1, \dots, l} \rightarrow y^{(k)}$$
- In the process of applying the HGF to data, we only need the perceptual and decision models. The generative model has already done its work: it has supplied the update equations of the perceptual model
- Both the perceptual and decision models have parameters that can be estimated individually for each dataset. Doing so requires defining priors for these parameters.

Parameter estimation with the HGF Toolbox

- Available at

www.translationalneuromodeling.org/tapas

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

```
est2 = tapas_fitModel(sim2.y,...  
                      usdchf,...  
                      'tapas_hgf_config',...  
                      'tapas_gaussian_obs_config',...  
                      'tapas_quasineutron_optim_config');
```

Parameter estimates for the perceptual model:

```
mu_0: [1.0352 1]  
sa_0: [3.7101e-05 0.0996]  
rho: [0 0]  
ka: 1  
om: [-12.8680 -1.8689]  
pi_u: 9.8449e+03
```

Parameter estimates for the observation model:

```
ze: 2.3970e-05
```

Parameter estimation with the HGF Toolbox

Activations by Precision-Weighted Visual Outcome Prediction Error ε_2

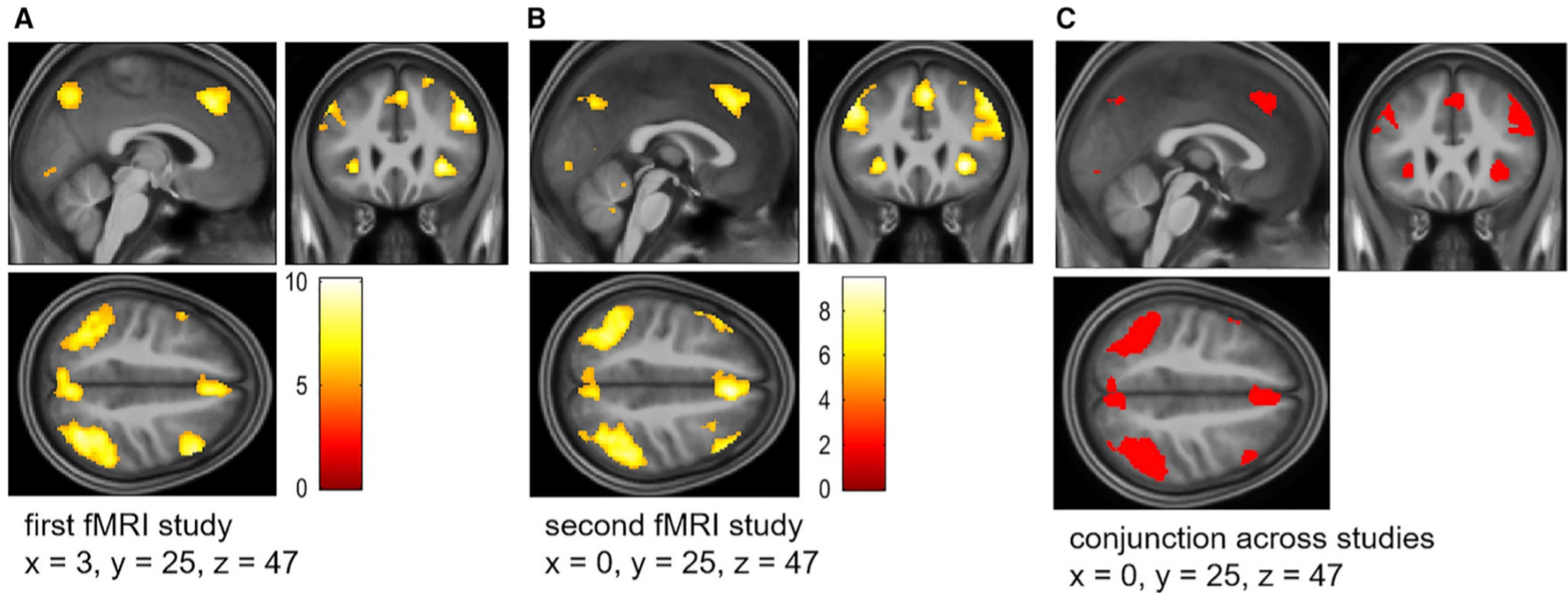


Figure 2. Whole-Brain Activations by ε_2

Activations by precision-weighted prediction errors 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 peak-level 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.

Iglesias et al, 2019 (Correction to Iglesias et al., 2013)

Parameter estimation with the HGF Toolbox

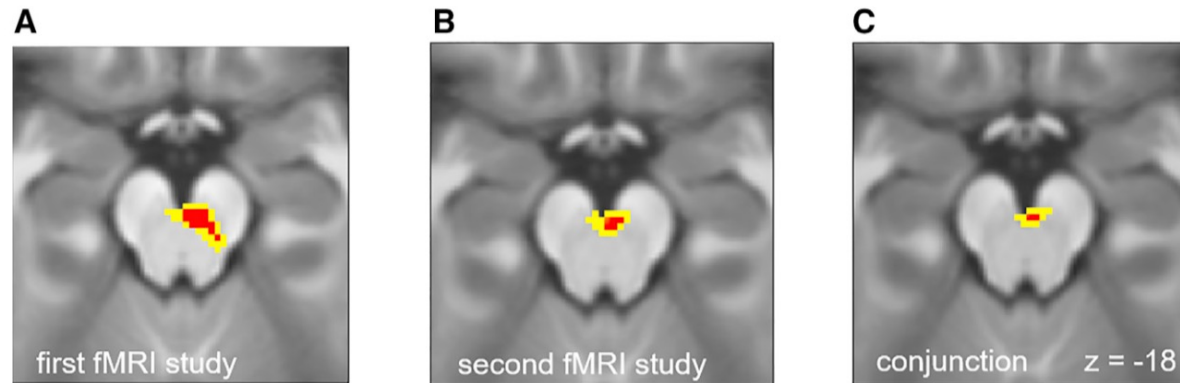


Figure 3. Midbrain Activation by ε_2

Activation of the dopaminergic VTA/SN by precision-weighted prediction errors about visual outcome, ε_2 . The activation at $p < 0.05$ FWE peak-level corrected for the volume of our anatomical mask (comprising both dopaminergic and cholinergic brain structures: VTA/SN, PPT/LDT, and basal forebrain) is shown in red. The activation thresholded at $p < 0.001$ uncorrected is shown in yellow.

(A) Results from the first fMRI study. (B) Second fMRI study. (C) Conjunction (logical AND) across both studies.

Parameter estimation with the HGF Toolbox

Activations by Precision-Weighted Prediction Error about Stimulus Probabilities ε_3

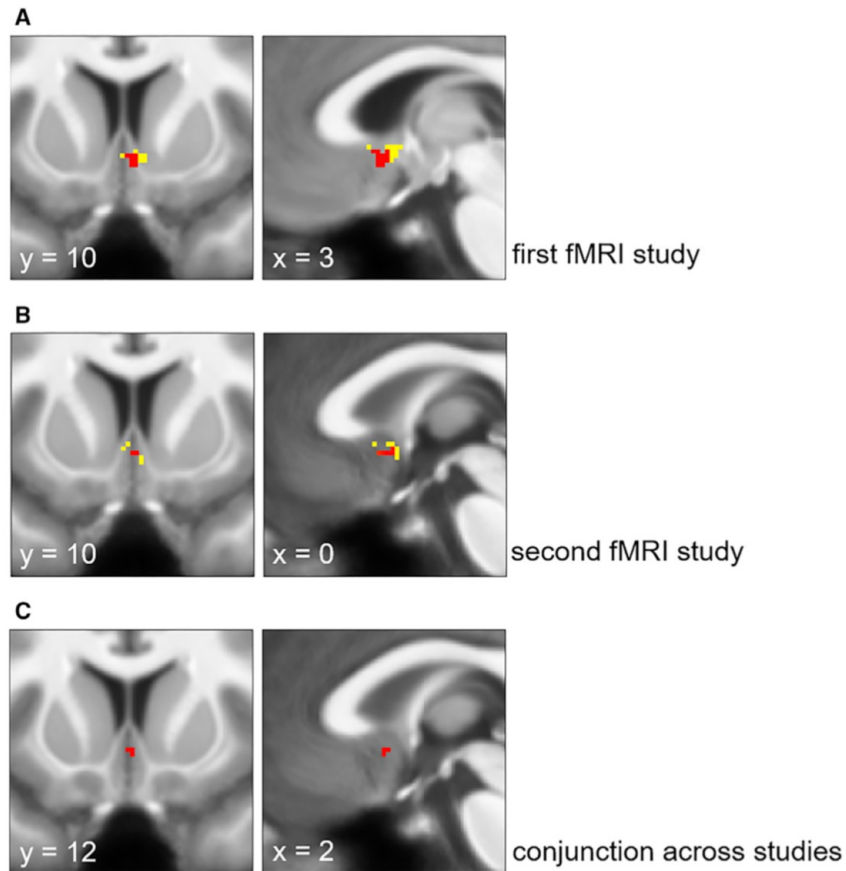
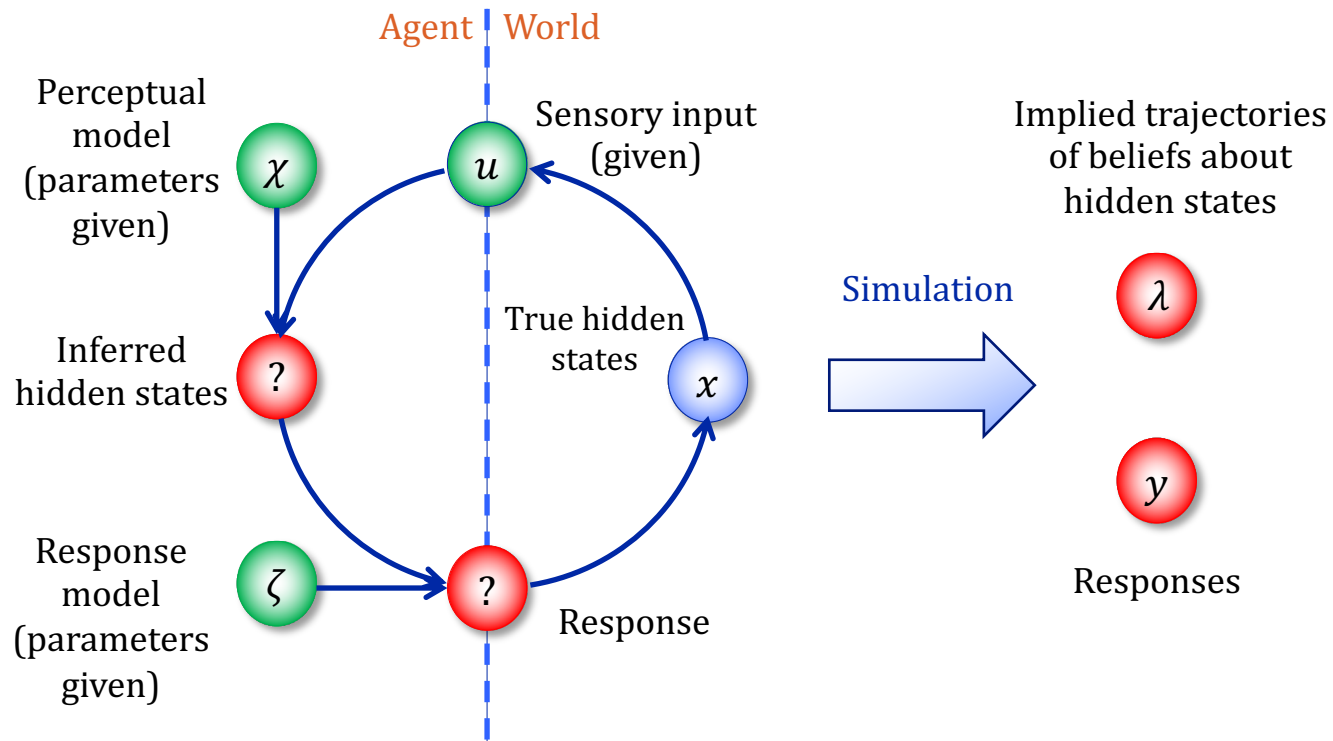


Figure 6. Basal Forebrain Activations by ε_3

Activation of the basal forebrain by 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 peak-level corrected for the entire anatomical mask (red) on the results thresholded at $p < 0.001$ (yellow; the yellow cluster also survives $p < 0.05$ FWE cluster-level correction for the entire anatomical mask). The anatomical mask comprised both dopaminergic and cholinergic brain structures: VTA/SN, PPT/LDT, and basal forebrain. (A) and (B) show results from the first (A: local maximum at $x = 4$, $y = 12$, $z = -11$, $t = 4.71$) and the second fMRI study (B: local maximum at $x = 0$, $y = 10$, $z = -8$, $t = 5.09$). (C) shows the conjunction analysis ("logical AND") across both studies. To ease visual comparison with Iglesias et al. (2013), the figure sections (x and y coordinates are indicated on each panel) are not located at the local maxima but correspond closely to those in Iglesias et al. (2013).

Iglesias et al, 2019
(Correction to Iglesias et al.,
2013)

Generation of new data: simulation

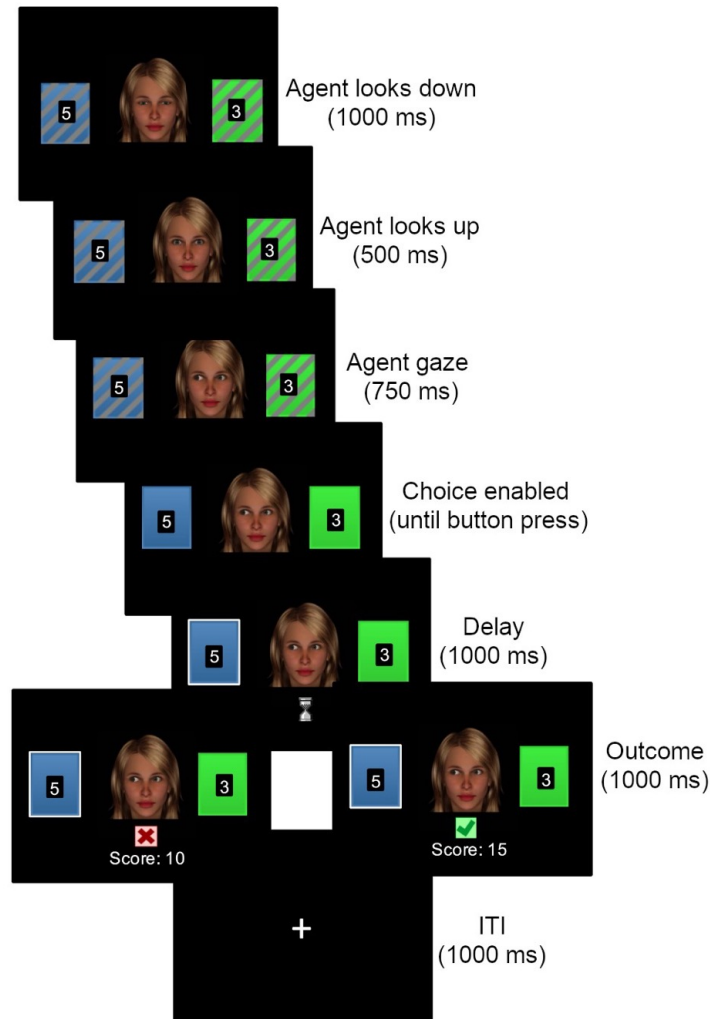


RESEARCH ARTICLE

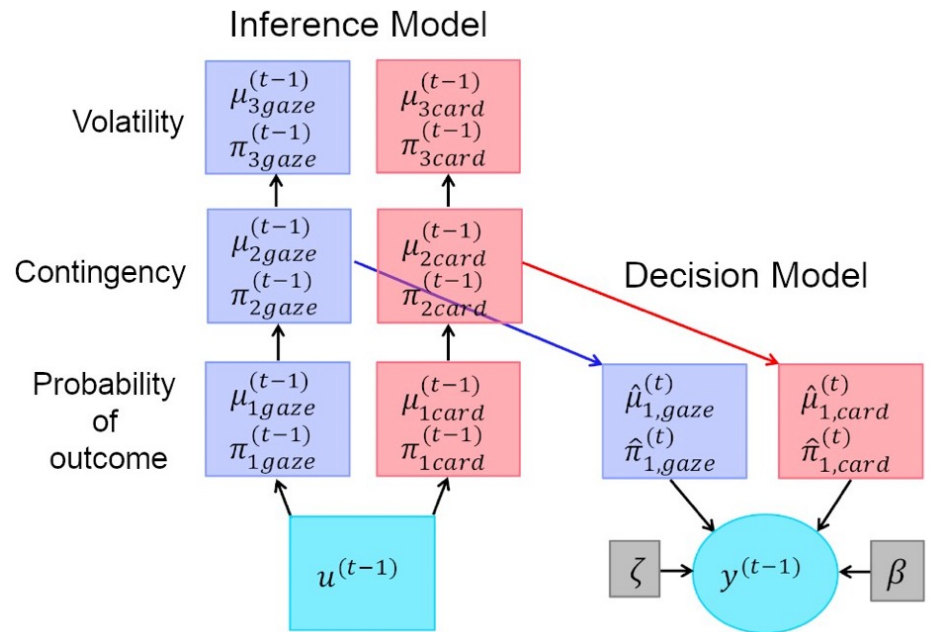
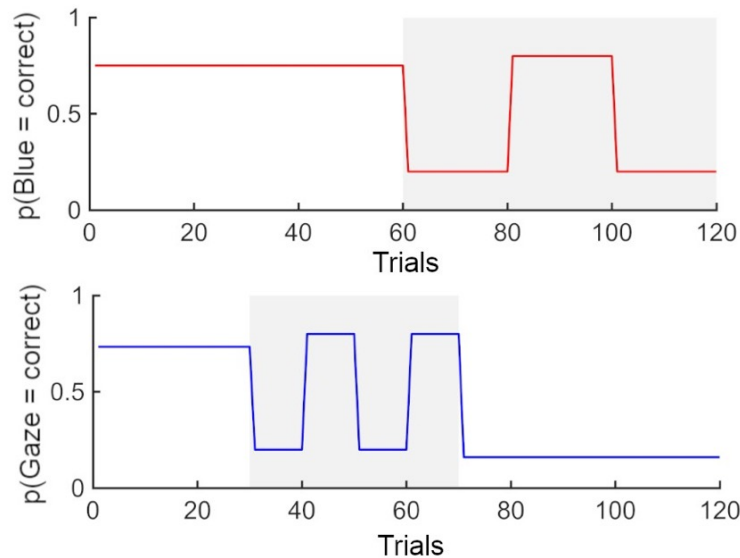
Aberrant computational mechanisms of social learning and decision-making in schizophrenia and borderline personality disorder

Lara Henco ^{1,2*}, Andreea O. Diaconescu ^{3,4,5}, Juha M. Lahnakoski^{1,6,7}, Marie-Luise Brandi ¹, Sophia Hörmann ¹, Johannes Hennings ⁸, Alkomiet Hasan^{9,10}, Irina Papazova ⁹, Wolfgang Strube⁹, Dimitris Bolis ^{1,11}, Leonhard Schilbach ^{1,2,11,12} , Christoph Mathys ^{4,13,14} 

Social inference in borderline personality disorder, schizophrenia, and major depression.



Social inference in borderline personality disorder, schizophrenia, and major depression.



Social inference in borderline personality disorder, schizophrenia, and major depression.

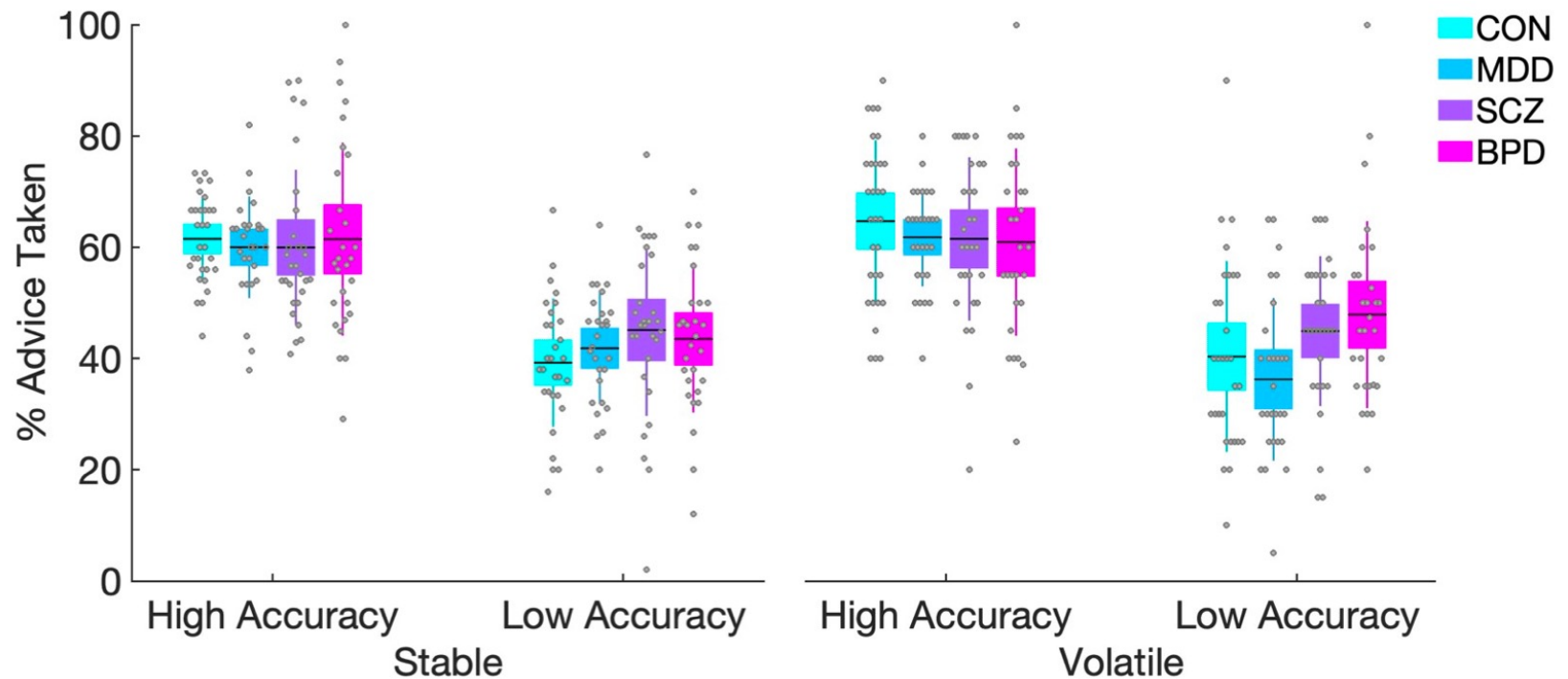


Fig 3. Behavioral results with regard to advice taking. All participants followed the advice significantly more during phases of high compared to low accuracy. The descriptive data shows a trend of BPD patients to follow the advice more during volatile phases of low accuracy (right most plot) but the interaction was not significant. Means are plotted with boxes marking 95% confidence intervals and vertical lines showing standard deviations.

Social inference in borderline personality disorder, schizophrenia, and major depression.

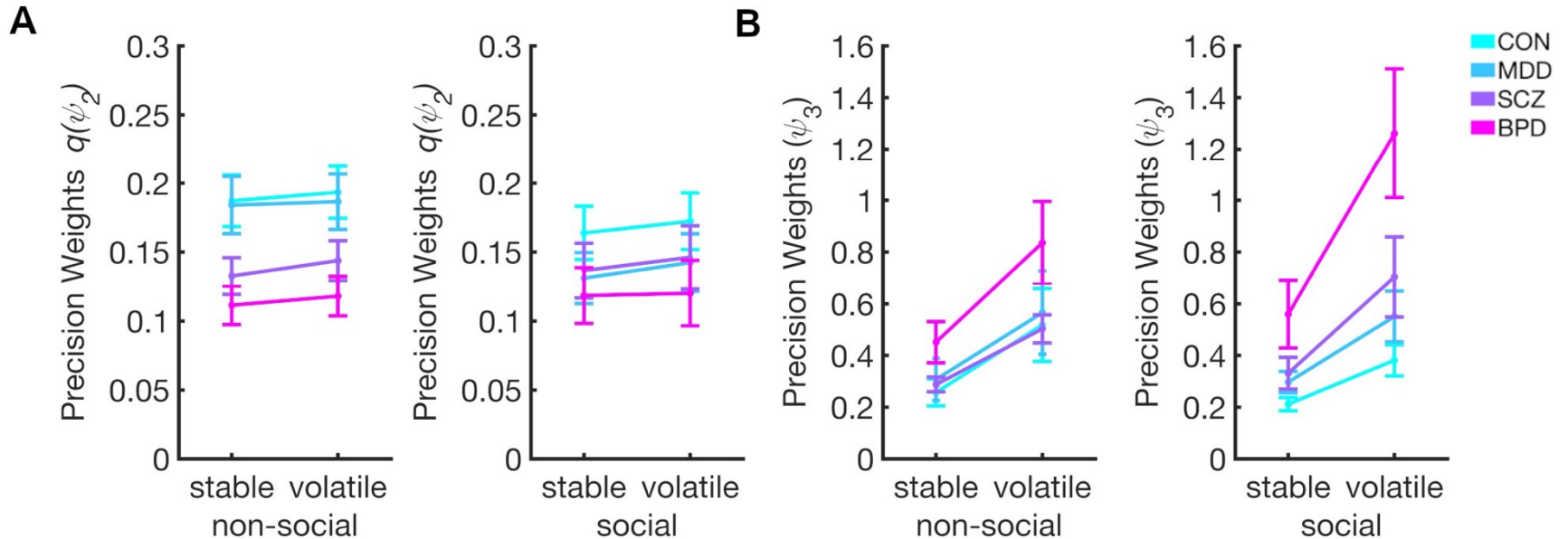


Fig 4. Results for mixed ANOVA using precision weights for updating beliefs about social and non-social contingency and volatility. (A) Precision weights $q(\psi_2)$ and (B) precision weights ψ_3 . Overall, $q(\psi_2)$ and ψ_3 increase when transitioning from stable to volatile phase. Patients with BPD show reduced overall $q(\psi_2)$. At same time, patients with BPD show higher ψ_3 compared to the other groups and a more pronounced increase in response to volatility. Bars indicate SEM. See also [S1 Fig](#), [S7 Table](#) and [S8 Table](#) for all results.

Henco et al. (2020): Summary

- Patients diagnosed with major depressive disorder (MDD, N = 29), schizophrenia (SCZ, N = 31), and borderline personality disorder (BPD, N = 31) as well as healthy controls (CTL, N = 34) performed a probabilistic reward learning task in which participants could learn from social and non-social information.
- SCZ and BPD performed more poorly on the task than CTL and MDD
- BPD performed better in the social compared to the non-social domain. In contrast, CTL and MDD showed the opposite pattern, and SCZ showed no difference between domains.
- In effect, BPD gave up a possible overall performance advantage by concentrating their learning in the social at the expense of the non-social domain.
- BPD showed slower learning from social and non-social information and an exaggerated sensitivity to changes in environmental volatility.
- Compared to CTL and MDD, BPD and SCZ showed a stronger reliance on social relative to non-social information when making choices.



ELSEVIER

Contents lists available at ScienceDirect

Schizophrenia Research

journal homepage: www.elsevier.com/locate/schres



A generative framework for the study of delusions

Tore Erdmann^a, Christoph Mathys^{b,c,*}

^a Scuola Internazionale Superiore di Studi Avanzati (SISSA), Via Bonomea 265, 34136 Trieste, Italy

^b Interacting Minds Centre, Aarhus University, Jens Chr. Skous Vej 4, 8000 Aarhus C, Denmark

^c Translational Neuromodeling Unit (TNU), Institute for Biomedical Engineering, University of Zurich and ETH Zurich, Wilfriedstrasse 6, 8032 Zurich, Switzerland

A generative framework for the study of delusions

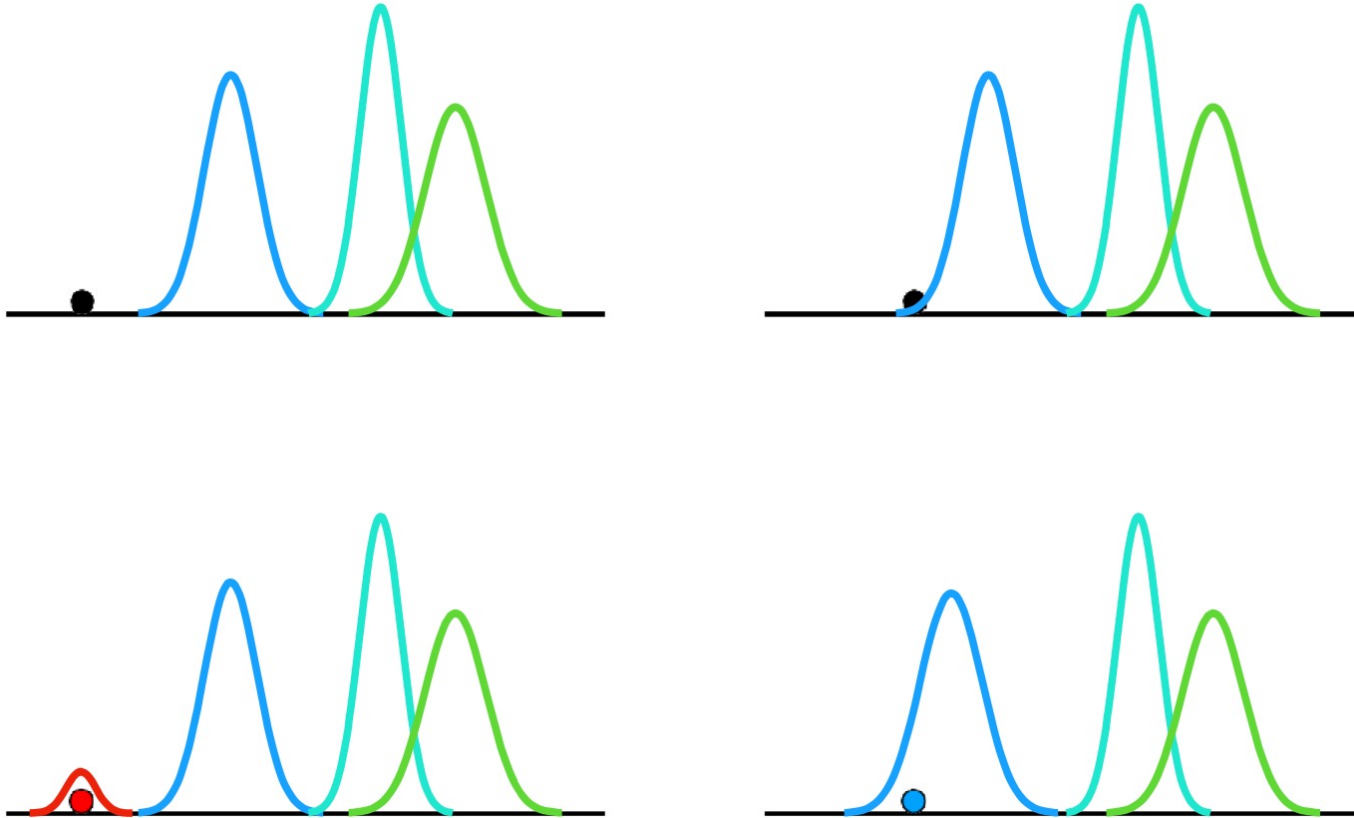


Fig. 1. Categorization and explanation in our framework (schematic). In the top panels, the same initial belief is depicted on the left and right, with separate explanations (causes) represented by Gaussians. On the left, the new observation (black dot) has a larger deviation from the existing causes than on the right. Here, the model infers a new cause and fits a corresponding cause to explain the observation (red Gaussian, bottom left). On the right, a less extreme observation is integrated into an existing cause (blue Gaussian, bottom right), which leads to a change in the structure of the corresponding explanation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

A generative framework for the study of delusions

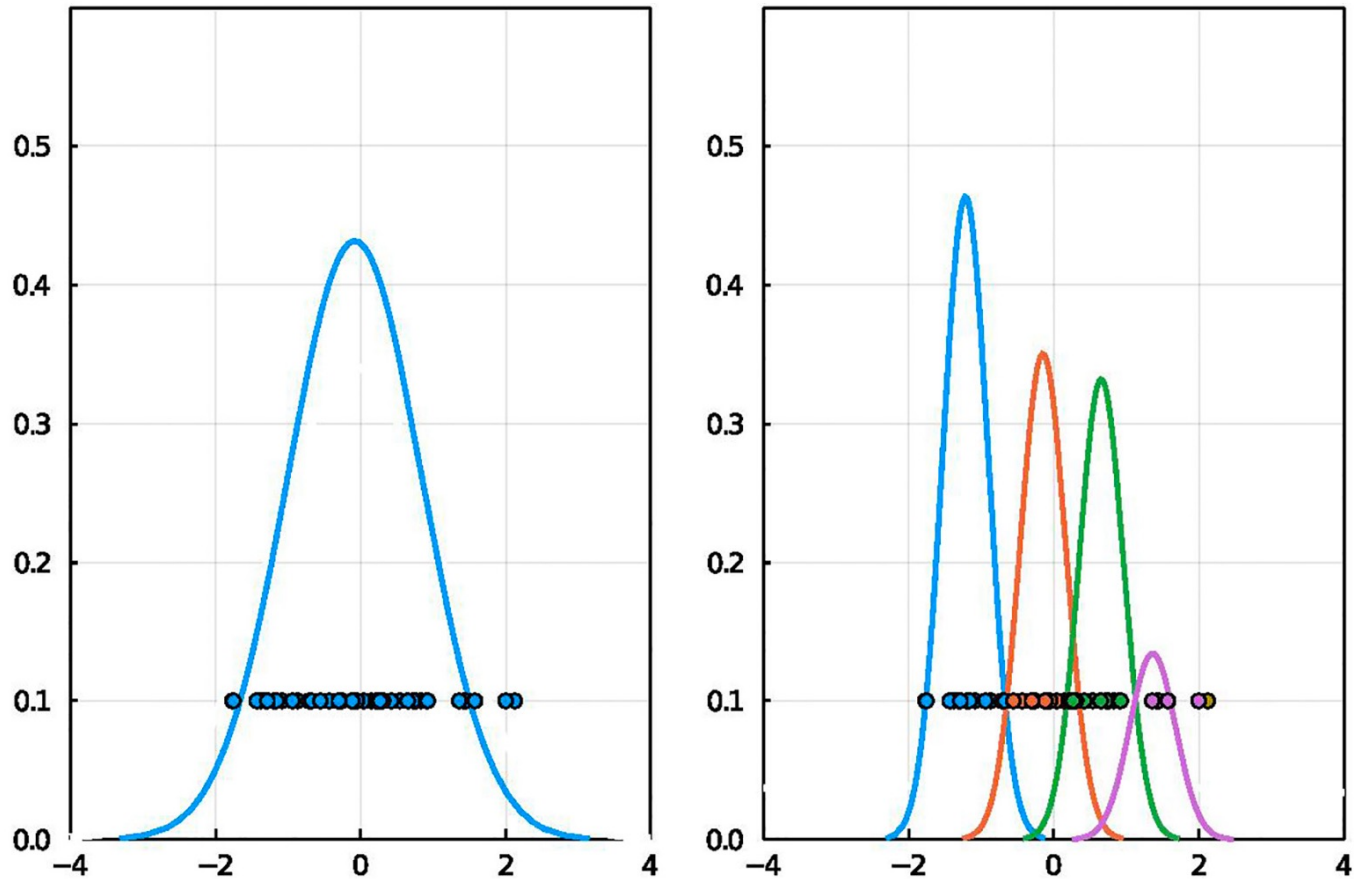


Fig. 2. A simulation of delusional mis-identification. In a case study, [Hirstein and Ramachandran \(1997\)](#) presented Capgras' patient DS with a sequence of photographs of the same model's face looking in different directions (here, we represent the photographs as points on a line; observations that are perceptually similar fall close on this abstract dimension). The left panel shows a simulation of inference in healthy observers: a single underlying cause ("the same person, photographed multiple times"; represented as a single Gaussian) is inferred. On the right, the inference observed in patient DS simulated ("different women who looked just like each other"; represented by multiple Gaussians). The two simulations from our model differed only in the expected precision (left: $\mu_r = \frac{1}{100}$, right: $\mu_r = 100$). Inputs and all other parameters were equal.

A generative framework for the study of delusions

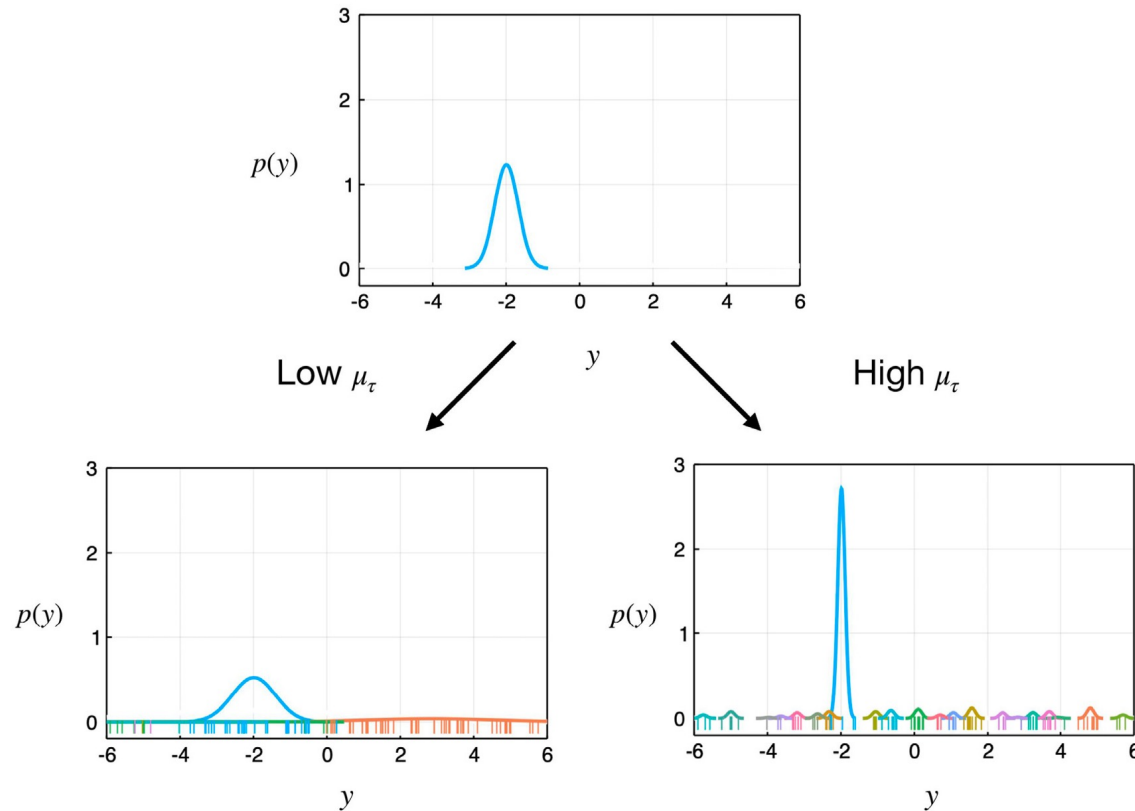


Fig. 3. Belief preserving evidence integration. Initial belief (upper row) and final belief (lower row) after inference given new observations. The difference in final beliefs is a function of the expected precision μ_τ alone. All other settings and inputs are the same. Bottom left: $\mu_\tau \sim \text{HN}(100, 10)$. The existing explanation (blue Gaussian) is elaborated (i.e., broadened) in response to new observations, which are to a considerable extent integrated into the already existing, but now elaborated, model. Bottom right: $\mu_\tau \sim \text{HN}(1/100, 10)$. The existing explanation is narrowed, but its dominance remains unaffected. New observations which do not fit it exactly are explained away (i.e., assigned to their own little ad hoc explanations). While both of these ways of processing the same information correspond to Bayesian inference (albeit under different values for μ_τ), the inference process on the right can be characterized as delusional. Further details and the code for reproducing this simulation can be found here: <https://tinyurl.com/y3m79qdw>. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

A generative framework for the study of delusions

Thanks

Remarks on parameter recovery

- *Parameter recovery* means the ability to estimate the ‘ground truth’ parameter values put into a simulation
- This is never possible to do exactly because simulation is stochastic
- The stochastic nature of the simulation means that the parameter value best able to explain the simulated data is not the same as that used in the simulation. By nature, the simulation only noisily represents the ‘ground truth’ parameter values.
- It is therefore misleading to call the value used in the simulation ‘true’. One could just as well argue that the estimated value is the ‘true’ one since it best explains the data.
- In order to get an idea of how exactly a parameter can be recovered, run many (on the order of tens to hundreds) simulations and look at the distribution of parameter estimates.
- **Estimating a parameter that isn’t well recovered from simulations can still make sense because adjusting its value can improve the model by leading to a better explanation of the data**

Remarks on model comparison / model selection

- There is a range of scores that help in choosing a well-performing model: AIC (Akaike information criterion), BIC (Bayesian information criterion), Bayes factors, LME (log-model evidence), free energy, etc.
- Each model gets a particular score (which is on its own uninterpretable!)
- The difference in score between models is what counts
- However, model selection is not straightforward. AIC and BIC penalize complexity based on simple heuristics, which may not reflect complexity accurately. LME is better on that count, but is very sensitive to the modeller's choice of priors.
- **Three important considerations:**
 1. **Does the model allow me to answer my question of interest?**
 2. **Does the *prior predictive* distribution of observations make sense?**
 3. **Does the *posterior predictive* distribution of observations make sense?**

When the answer to all three is yes, the model is fine.