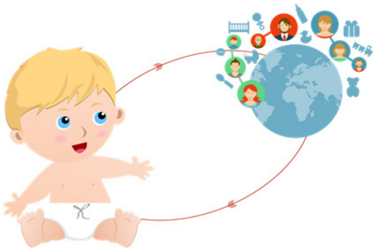
DONDERS INSTITUTE



Learning and Development

Predictive Processing Research Seminar

Radboud University Radboudumc

Organizational stuff

- 21st March: Andy Clark lecture in Amsterdam
- 23rd March: PP Seminar on the same topic! (enactivism, embodied and embedded PP)
- Birds of a feather report / announcements
- Call for new topics / discussions starting April 13nd

Recall: Predictive Processing

Brain as **prediction machine**
- *The brain continuously makes predictions about future sensory evidence based on its current best model of the causes of such evidence*

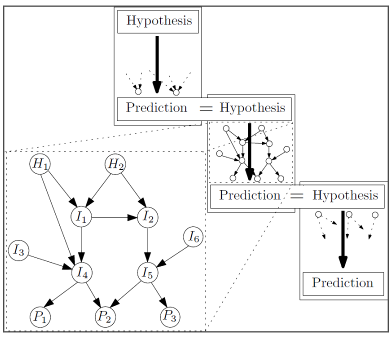
Bayesian Brain
- *The brain combines prior knowledge with sensory evidence (from various sources) in a Bayesian way*

Hierarchical Brain
- *The brain is organized in a hierarchical way, where "high level" information influences "low level" information and vice versa*

Recall: Key sub-processes

- Making **predictions** of expected input based on the generative models
- Comparing predicted inputs with actual inputs and **computing prediction error**
- **Explaining away** prediction errors (minimizing prediction error)
- **Learning** and adapting generative models

Recall: Computational model



Hypothesis variables $Hyp = \{H_1, H_2\}$

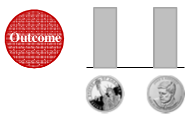
Prediction variables $Pred = \{P_1, P_2\}$

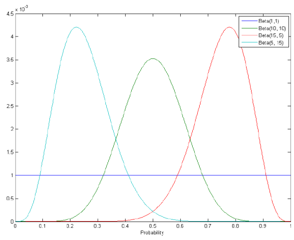
Intermediate variables $Int = \{I_1, \ldots I_6\}$

Hyperprior

- Hyperpriors are **priors over priors**
- Define a distribution P(Outcome) over Heads and Tails
- A hyperprior now describes a **distribution over x**, such that P(Outcome = Heads) = x [and P(Outcome = Tails) = 1 – x]
- What does it mean and what does it look like?

Beta distribution





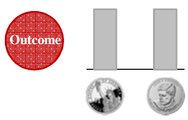
$$P(p;\alpha,\beta) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}$$

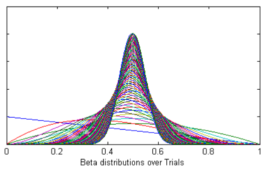
$$\mathrm{B}(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt = \frac{\Gamma(x)\,\Gamma(y)}{\Gamma(x+y)}.$$

Beta function as hyperprior

- Hyperpriors are **conjugate priors** over the corresponding likelihood function
- Conjugate here means: if the likelihood distribution is of the family X, choosing a conjugate prior ensures that the **posterior distribution** is also of the family X
- In particular, a beta distribution is a conjugate prior over a binomial distribution (in this case: outcomes of coin tosses)
- Dirichlet distributions are conjugate priors over categorical distributions, Gaussians are conjugate priors over themselves, etc.

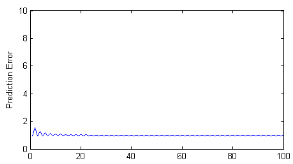
Bayesian Updating





s = heads, n-s = tails

$$= \frac{x^{s+\alpha\mathrm{Prior}-1}(1-x)^{n-s+\beta\mathrm{Prior}-1}}{\int_0^1 \left(x^{s+\alpha\mathrm{Prior}-1}(1-x)^{n-s+\beta\mathrm{Prior}-1}\right)dx}$$



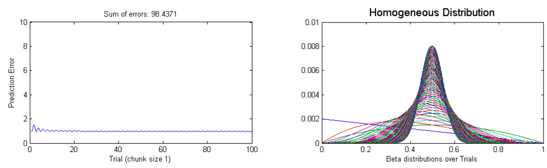
Precision-weighted prediction error

- Precision-weighted prediction error describes the **size of the effect** of a prediction error on the updating of the model
- The higher this *pwpe*, the **bigger the effect** on the generative model a prediction error is
- The higher this *pwpe*, the **more reducible uncertainty** there is in the environment
- We define this *pwpe* as the **KL divergence** between the hyperprior 'before' and 'after' updating with the new data
- Note that this is a **descriptive** measure, not a normative!

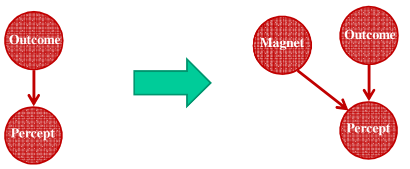
Model updating cannot be the whole story

- There is evidence that 6mo infants process prediction errors, but the generative models undergo substantial post-natal development (Emberson et al., 2015)
- Immaturity of these models leads to sub-optimal processing of sensory stimuli in infants (Lee et al., 2012)
- How can models grow, incorporate new experiences, become more fine-grained etc.?
- There are computational arguments that Model Updating isn't the whole story, as well as empirical evidence!

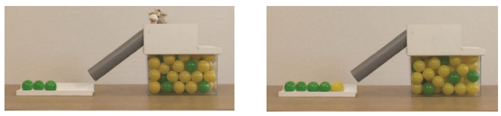
Model Updating?

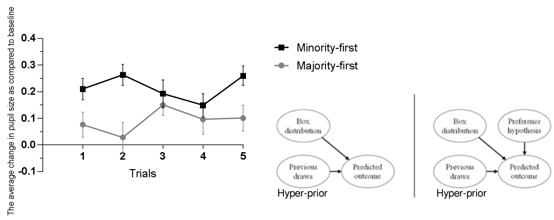


Model Revision



Model Revision

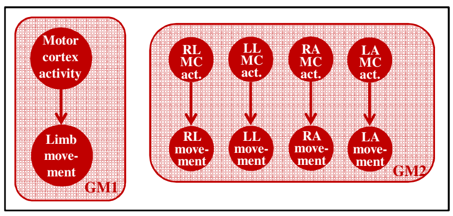




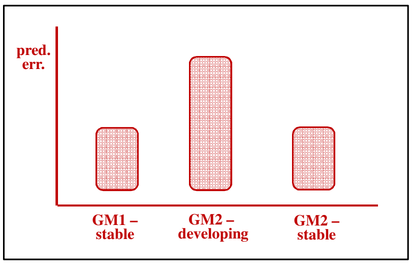
Model Refinement

- Apart from adding context dependencies, adding hypotheses, changing structure of the models etc., another process that kicks in is Model Refinement
- In new & uncertain situations, very detailed predictions will carry lots of information but are likely to be false
- When you become more confident it becomes attractive to increase the *level of detail* of your predictions to make more informative predictions
- Think motor babbling in infants

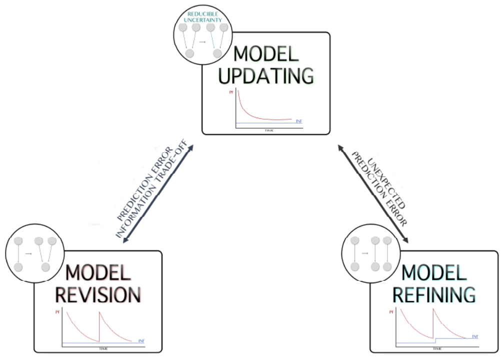
Model Refinement



Model Refinement



Model Updating, Revision, and Refinement



Model development open questions

- Where do the new hypotheses **come from**?
- How are they **integrated** with existing models?
- How are models **'split'** into more detailed models?
- How is the **trade-off** between informativeness and prediction error resolved? Is this context dependent? Governed by **hyperpriors**?
- How can we test this all?