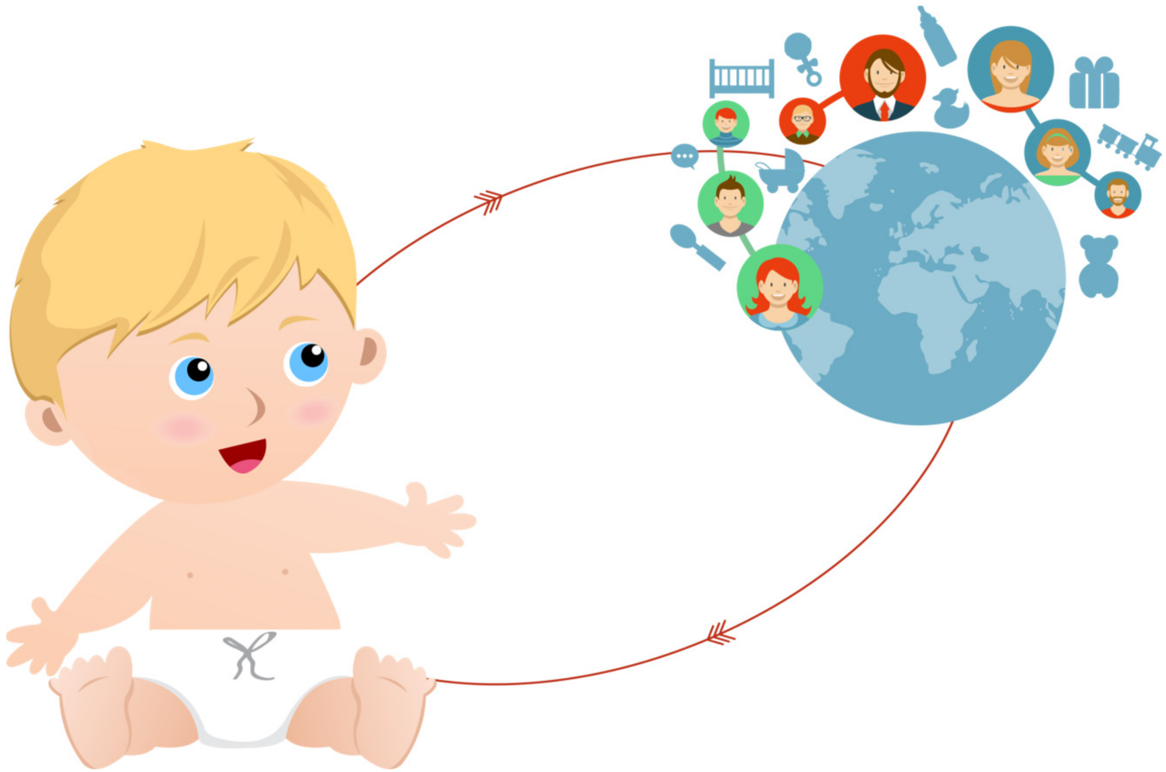


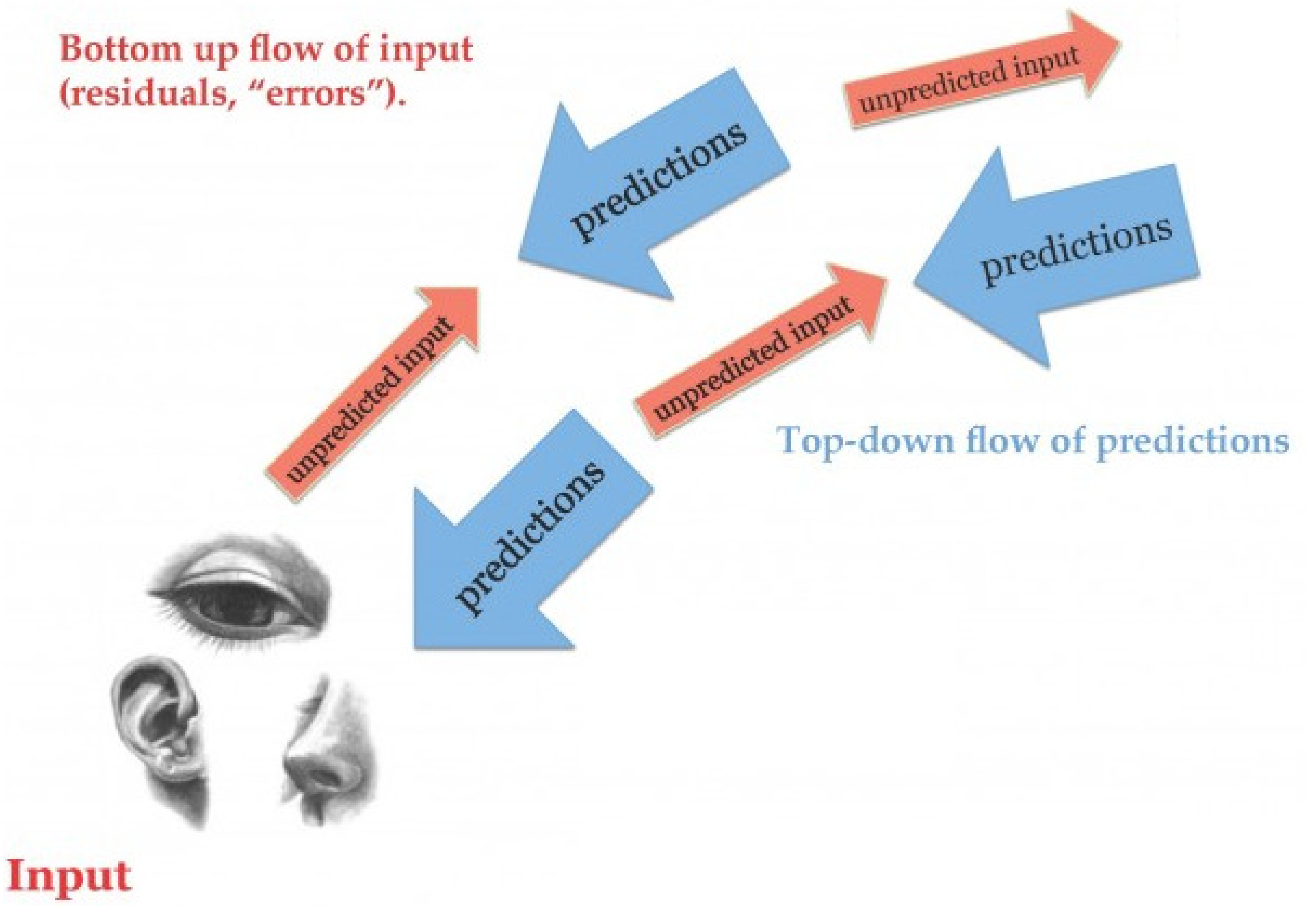
A Predictive Processing Account of the Developing Brain

Johan Kwisthout – ICDL / EpiRob 2016 workshop

Radboud University Radboudumc

A novel view of the brain

"...essentially a sophisticated hypothesis-testing mechanism..."

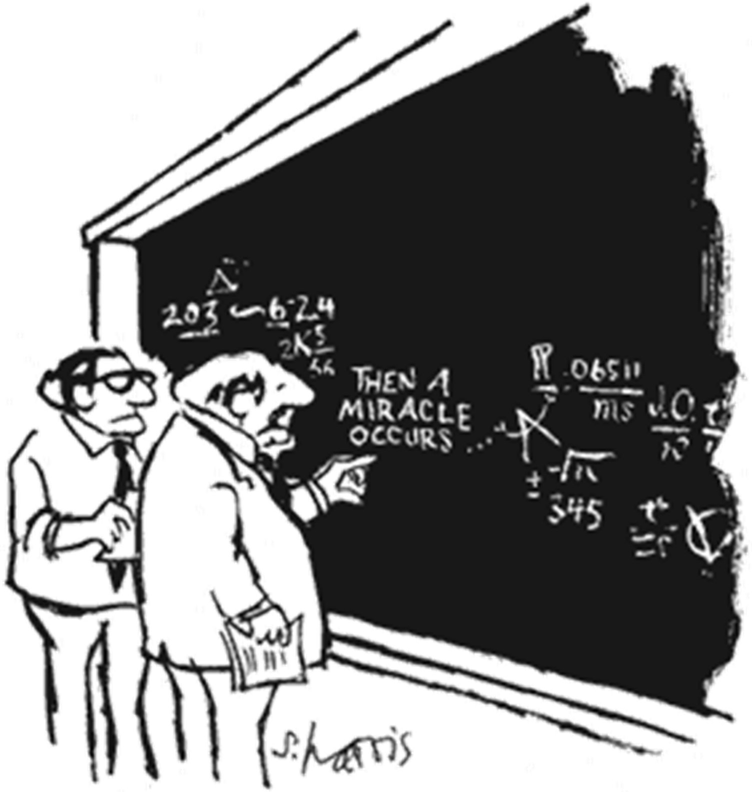


Clark, A. (2015). Surfing Uncertainty.

A novel view of the brain

- ✓ Explanatory power
- ✓ Neurophysiological evidence
- ✓ Rich formal machinery
- ✗ ... but lacking one crucial ingredient!

Where Predictive Processing is silent...

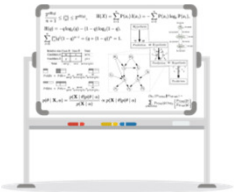


"I THINK YOU SHOULD BE MORE EXPLICIT HERE IN STEP TWO."

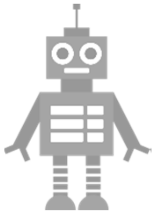
Overview of this presentation

Predictive Processing as **unifying** account

Concrete **computational** framework



How to **develop** generative models?



Robo-havioral research **methodology**

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

Predictive coding and predictive processing

Computational level Conceptual description	**Predictive Processing** [e.g., Clark] *Keywords*: predictions at various levels of detail, precision-weighted prediction errors, hypothesis updating, model revision	
Algorithmic level Process-level description	**Predictive Coding** [e.g., Friston] *Keywords*: low-level cognition, continuous Gaussian models, variational Bayes approximations	**Belief Propagation** [e.g., Sanborn] *Keywords*: high-level cognition, structured discrete models, sampling approximations, particle filtering
Implementational level Neuronal level description	**Cortical Microcircuits** [e.g., Bastos] *Keywords*: pyramid cells, feedforward-feedback connections	**Networks of Spiking Neurons** [e.g., Maass] *Keywords*: Boltzmann machines, switching rate, noisy spikes

Key sub-processes

- Making **predictions** of expected input based on the generative models that relate causes and effects

- Comparing predicted inputs with actual inputs and **computing precision-weighted prediction error**

- **Explaining away** prediction errors (minimizing overall prediction error)

- **Learning** and updating generative models based on the precision of the prediction errors

From conceptual idea to formal model

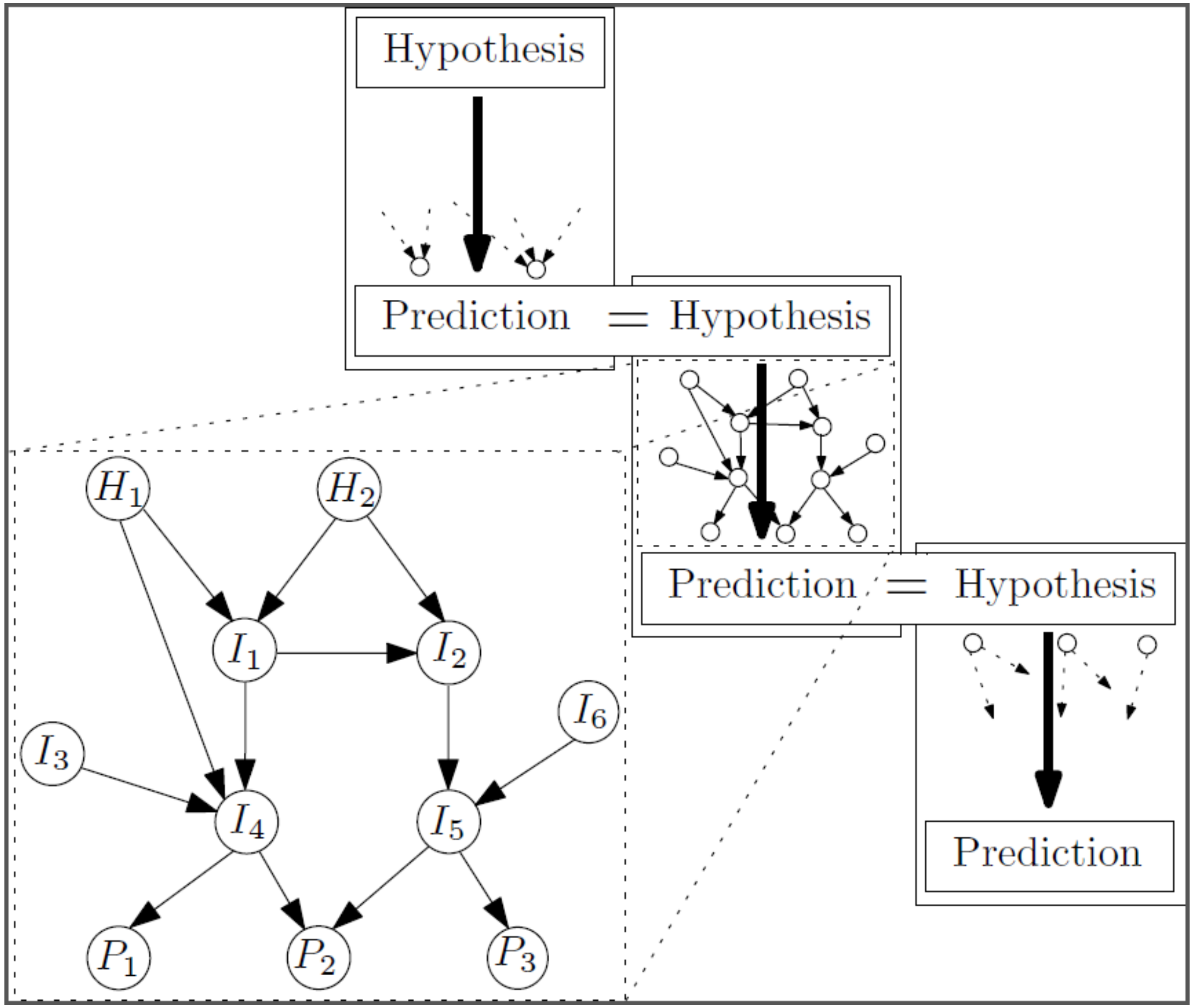
- Predictive processing is assumed to **explain and unify all of cognition**, including higher cognition

- To model, e.g., complex social interactions, Theory of Mind, intention recognition, and problem solving, we need **rich enough** knowledge structures to model complex, non-monotone, non-linear dependences

→ We argue (Otworowska et al., 2014) that simple Gaussian models are **not sufficiently rich** models for higher cognition

→ We propose to use **causal Bayesian networks** as knowledge structures instead to describe predictive processing

Computational model – prediction generation



Hypothesis variables

$Hyp = \{H_1, H_2\}$

Prediction variables

$Pred = \{P_1, P_2\}$

Intermediate variables

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

Computational model – error estimation

- Prediction and Observation are **probability distributions** over the prediction variables *Pred*

- Prediction is defined as computing the **posterior distribution** $\Pr_{(Pred)}$ given the parameters in the network

- Prediction error is set difference $\Pr_{(obs)} - \Pr_{(Pred)}$

- Estimating the *size* of this error is defined as computing the KL-divergence or **relative entropy** between predicted and observed distribution

$$D_{\mathrm{KL}}(\Pr_{(\mathrm{Pred})} \| \Pr_{(\mathrm{Obs})}) = \sum_{\mathbf{p} \in \Omega(\mathrm{Pred})} \Pr_{\mathrm{Pred}}(\mathbf{p}) \log \left(\frac{\Pr_{\mathrm{Pred}}(\mathbf{p})}{\Pr_{\mathrm{Obs}}(\mathbf{p})} \right)$$

Computational model – error minimization

- Prediction error minimization: "doing something" such that $D_{KL}(\ Pred\ ||\ Obs\ )$ is minimized
- Several possible ways of "**doing something**" (Kwisthout et al, 2016):
  - Revising beliefs about current state of the world
  - Gathering information (e.g., look around)
  - Active inference (move your arm)
  - Modulate model by contextual influences (oh yeah, I'm on the moon – less gravity!)
- In this talk: long-term development / change of model

Building generative models

- **Thought experiment**: let's assume I give you a coin and tell you that it may or may not be biased to either side, what would your predicted outcome be?

P(

) = 0.5

P(

) = 0.5

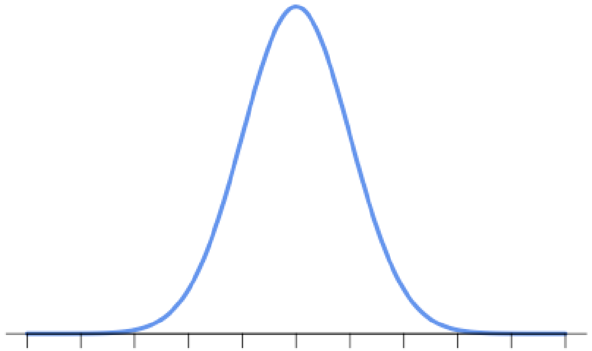
- But why? According to the Jaynesian principle of maximum entropy your prediction will be such that it carries no information that is not actually there

- This happens to be the uniform distribution

Building generative models

- **Second order** probability distribution

P("P() = X")



- The precision of this distribution is the inverse variance
- It indicates the confidence you have in this distribution
- This will change, using Bayesian updating, to a more narrow distribution given more evidence

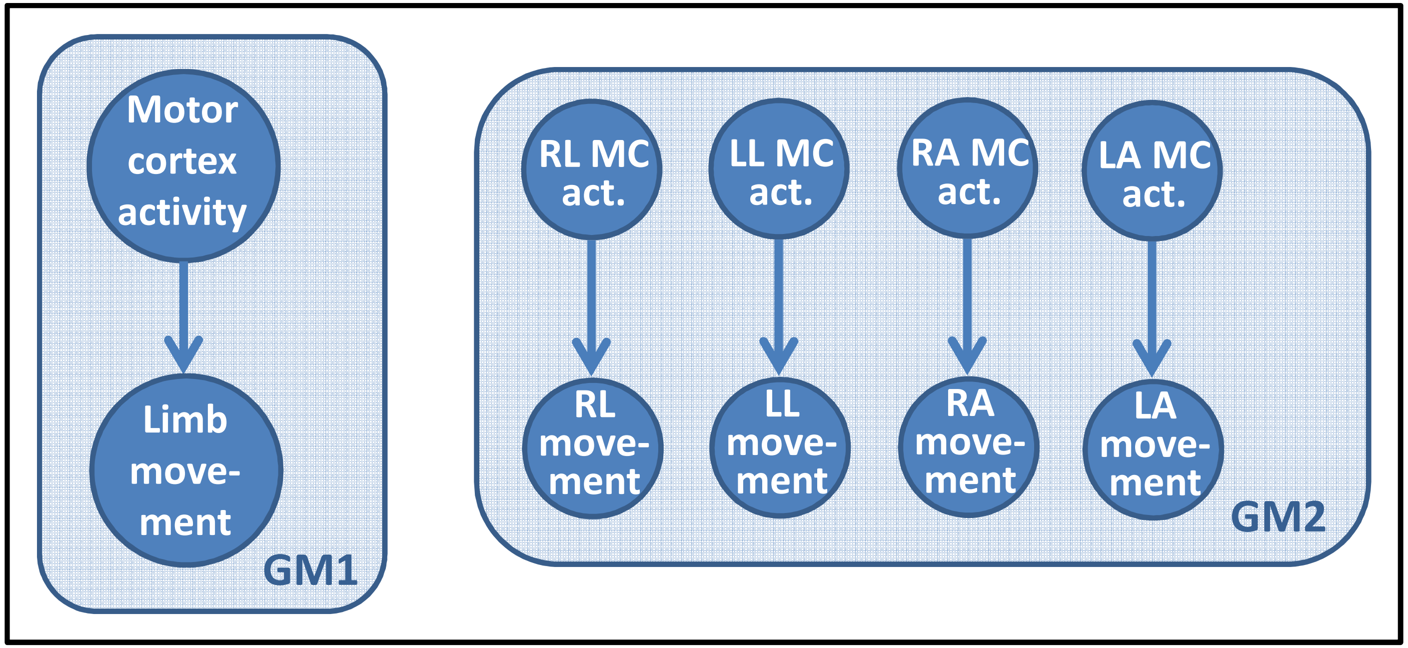
Updating or developing generative models?

- Generative models are updated using **precision-weighted prediction errors** – if there is lots of reducible uncertainty, precision of the prediction error is higher than when all the uncertainty is irreducible

- But this already assumes that there is a generative model **in the first place**! We need to know that “outcome of coin toss” is a binary variable with Heads and Tails as possible values

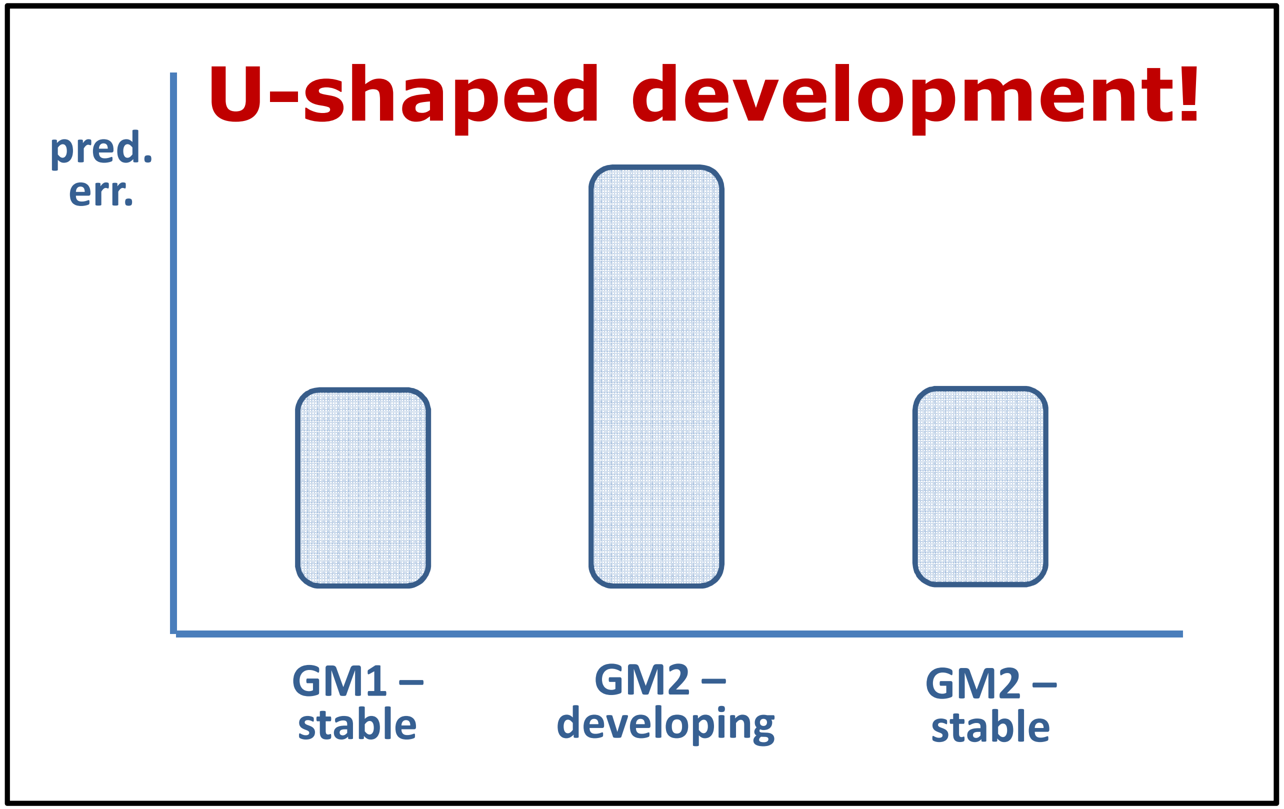
- But how do we **develop** such generative models in the first place? Where do the hypotheses come from, how are new hypotheses **integrated** in existing models?

From lower to higher detailed models



- Proposed rough idea: We start with very coarse and broad models (like GM1) and **refine** them to more detailed models over time

From lower to higher detailed models



- This leads to this particular **pattern** of prediction errors

Many open questions

- Currently my students are working on many (theoretical, computational, and developmental) **models and theories** based on this principle

- E.g., how are initial “least detailed” models generated based on isolated experiences that are generalized? What triggers a model revision? If we (e.g.)`split’ a variable, how is this computationally realized

- Challenge: we cannot just take a computational model of the shelf to base our theory on! They **don’t exist**!

- Probability theory etc. **assumes** a given state space

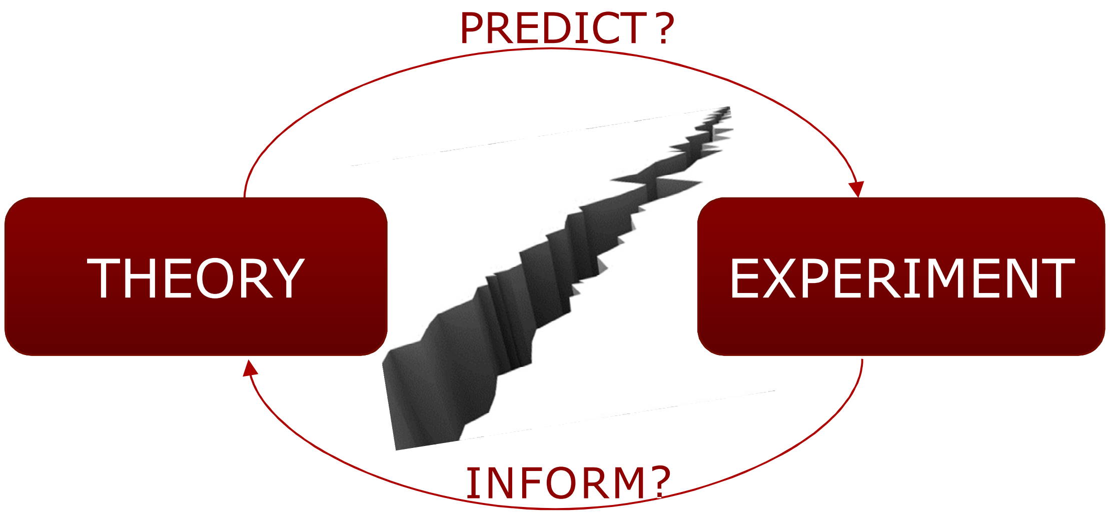
Many challenges

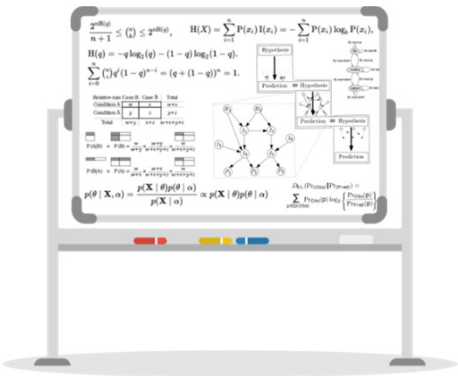
- We use modelling using a framework for which **part of the math** is yet to be developed

- We aim to contribute computational models to a unifying theory of the brain that paints in rather **broad strokes**

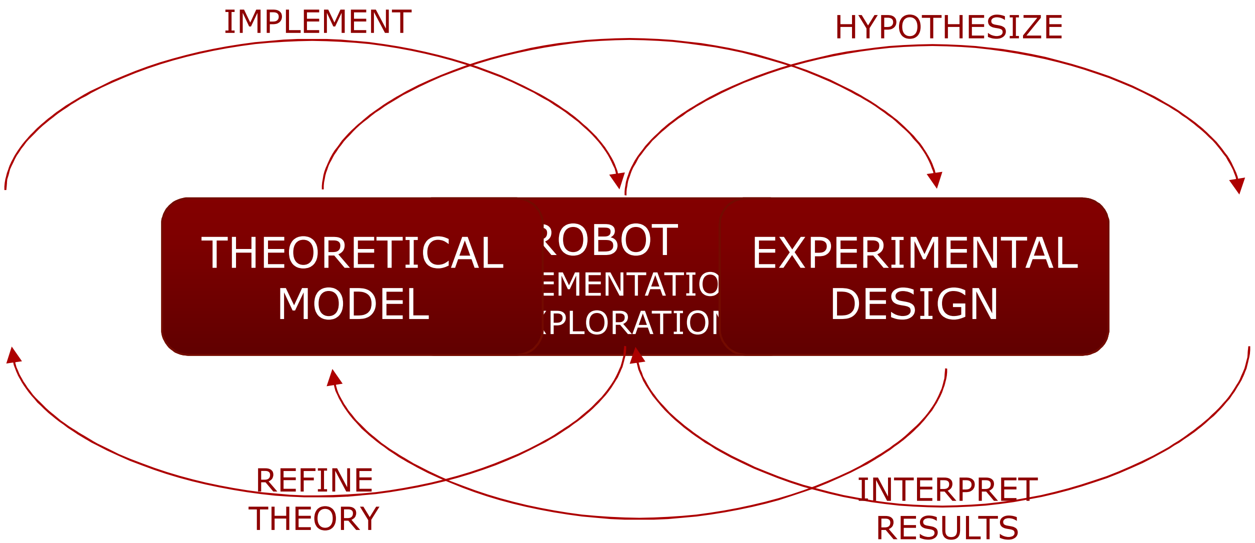
- We want our theorizing and models to **inform** and **be informed** by experimental infant studies

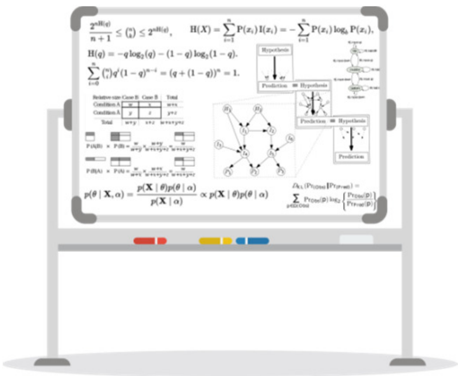
Research methodology



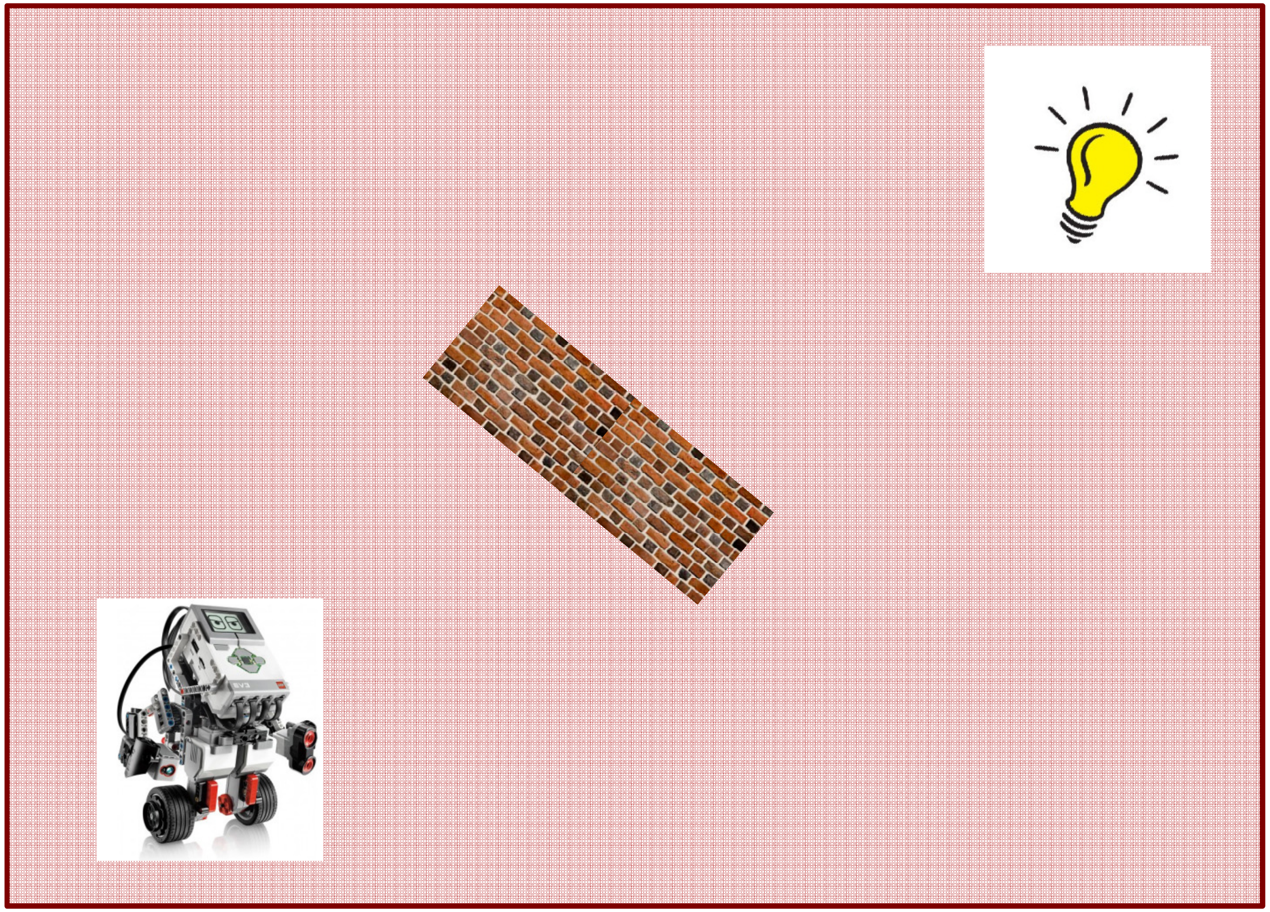


Research methodology





Implementing and exploring

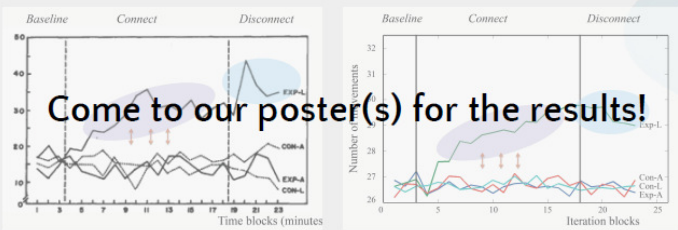


See our posters!

The mobile-paradigm as a measure of infants' sense of agency: Insights from babybot simulations

Lorijn Zaadnoordijk, Maria Otworowska, Johan Kwisthout, Sabine Hunnius, Iris van Rooij

- Developing sense of agency
- Mobile paradigm
- Simulated infant:
  - Operant conditioning mechanism
  - Incapable of causal learning
  - Proof of concept
- Are the observed behavioral patterns evidence for causal learning?

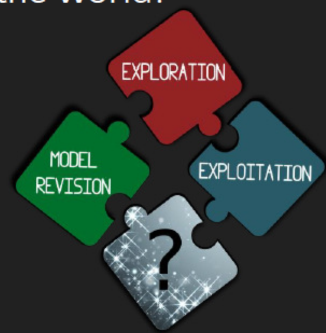


Come to our poster(s) for the results!

Causal learning in the crib: A predictive processing formalization and babybot simulation

Maria Otworowska, Lorijn Zaadnoordijk, Erwin de Wolff, Johan Kwisthout, Iris van Rooij

- Causal learning in Predictive Processing
- How do infants learn generative models of how their actions cause events in the world?



POSTER: Full formalization, simulations, results and **highly relevant** conclusions!