ANC sysmposium 19 May 2014

Leaving Andy Clark's safe shores: Scaling Predictive Processing to higher cognition

Johan Kwisthout, Maria Otworowska, Harold Bekkering, Iris van Rooij

Donders Institute for Brain, Cognition and Behaviour

Radboud University Nijmegen

Lorentz workshop on HPI

Motivation of this work

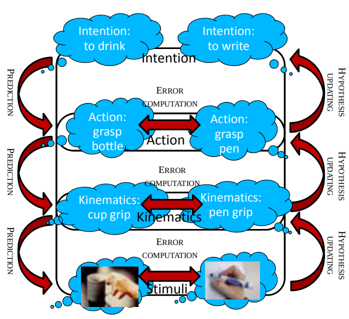
(Clark's BBS paper, p. 201)

"Questions also remain concerning the proper scope of the basic predictive processing account itself. Can that account really illuminate reason, imagination, and action selection in all its diversity? What do the local approximations to Bayesian reasoning look like as we **depart further and further from the safe shores of basic perception and motor control**? What **new forms of representation** are then required, and **how do they behave** in the context of the hierarchical predictive coding regime?"

Predictive Processing

- Brain as prediction machine
  - *The brain continuously makes predictions about future sensory evidence based its current best model of the causes of such evidence*
- Not **Friston**-ese, but **Clark**-ian / **Hohwy**-esque
- Focus on conceptual principle rather than Friston's "vanilla-flavor predictive coding" which is closely tied to the cortical hierarchy – we wish to abstract away from that and stay at Marr's computational level (for now)
- Computational translation of conceptual principles

PP flow of predictions and prediction errors



- Bottom-up inferential processing is augmented with top-down generative processing
- Predictions are made and compared to actual (or inferred) observations
- Prediction errors are used to update the hypotheses

Key (computational) aspects of PP

- Prediction ↔ Observation
  - (Both with particular *precision*)
- Prediction error
  - Difference between prediction and observation
- Lowering prediction error
  - Using original hypothesis and prediction error
  - Whether observation or prediction error is the "current" that drives bottom-up processing
- All within a Bayesian (= probabilistic) framework and assuming a hierarchy of predictions and hypotheses

Challenges for PP in higher cognition

- In "vanilla-flavoured predictive coding", Gaussian probability densities are assumed to model prior knowledge and likelihood of observations given the hypothesised causes of these observations
- With these Gaussian distributions, many aspects of predictive processing follow quite naturally:
  - Prediction = Mean + Variance (noise)
  - Precision of prediction = Inverse variance
  - Prediction error = Kullback-Leibler divergence between prediction and observation
  - Hypothesis updating = Variational Bayes approximation
- Does this scale to, e.g., action understanding?

Object-related actions on "cookies"


Selecting


Taking


Refusing


Sharing

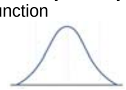

Crushing


Throwing

Challenges for HPC in higher cognition

Simple Gaussian representations are quite natural when we stay on the "safe shores of basic perception and motor control". However, these representations do not fit easily with higher cognition such as object-related actions:

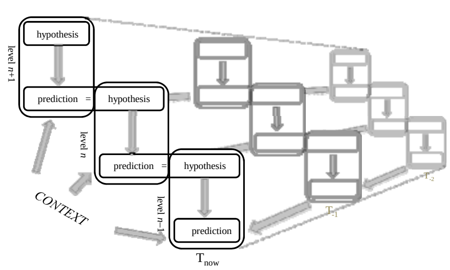
	Basic Motor Act	Object-related Action
Nature?	Simple / Continuous	Structured / Discrete
Monotone?	YES	NO
Ordered?	YES	NO
Model?	Probability Density Function	Probability Mass Function



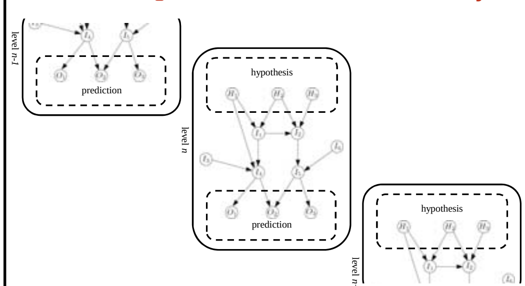
Challenges for PP in higher cognition

- How to translate the basic concepts of PP to structured domains? What representations do we need?
  - General principle = inference in structured graphical models (e.g., Bayesian networks)
  - Prediction = computation of posterior
  - Prediction error = Kullback Leibler divergence between prediction and observation
  - Precision of prediction = **Research Question 1**
  - Hypothesis updating = **Research Question 2**
  - Learning = **Research Question 3**

Top-down process: dynamic hierarchical BN



Zooming in on one level of the hierarchy



Research Question 1: precision

- How can we define precision in arbitrary probability distributions such as Bayesian networks, where there is no sense of "variance" as there is in simple Gaussians?
- One step back: what is precision actually?
  - In vanilla-flavored predictive coding?
  - In more conceptual predictive processing?

PRECISION → UNCERTAINTY

PRECISION → LEVEL OF DETAIL

Uncertainty

- In vanilla-flavored predictive coding, precision = inverse of model uncertainty = inverse of variance
- More in general, there is a well-known general notion of uncertainty in probability distributions and densities: Shannon entropy (in discrete distributions) and its extension to continuous distributions: differential entropy

$$h(X) = -\int_{\mathbb{X}} f(x)\log f(x)\,dx \;=\; \frac{1}{2}\ln(2\pi e\sigma^2)$$

- For a simple Gaussian, entropy is log-related to variance

Uncertainty

- So, we may generalize "precision" in simple Gaussians to "(differential) entropy" in arbitrary distributions without fundamentally changing the notion of uncertainty

$$H(X) = \sum_i P(x_i)\,I(x_i) = -\sum_i P(x_i)\log_b P(x_i)$$

$$h(X) = -\int_{\mathbb{X}} f(x)\log f(x)\,dx \;=\; \frac{1}{2}\ln(2\pi e\sigma^2) \quad \text{(simple Gaussians)}$$

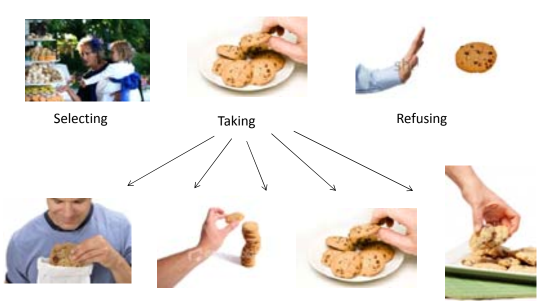
$$= \frac{k}{2}(1+\ln(2\pi)) + \frac{1}{2}\ln|\Sigma| \quad \text{(multivariate simple Gaussians)}$$

- No general analytical form for mixture distributions

Level of detail

- Conceptually speaking, "increasing precision" in predictive processing is more about "level of detail of predictions" than about "uncertainty about predictions"
- Our proposal: modulate the **aggregation level** of Bayesian random variables, their values and inter-dependences
- Quite literally: zooming in and out on probability mass functions!
- No real counterpart in Gaussian models


Aggregation of actions



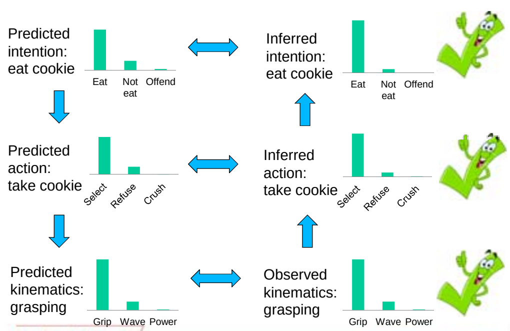
Probability mass function

Zooming in

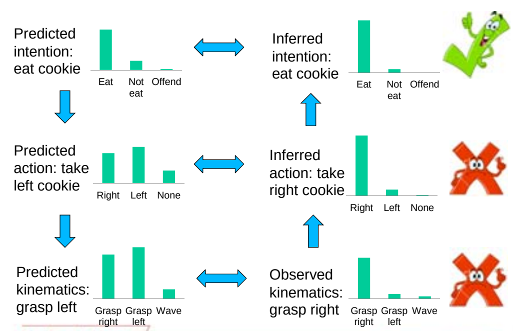
Precision: as precise as possible

- More precise predictions have lower probability, but are more specific (i.e., have more detail), than less precise predictions
- Objective: minimize prediction error (proxy for free energy) by making good predictions that are likely to be true and as precise as possible
- Too precise: likely to generate prediction error when comparing prediction with observation
- Too general: likely to generate prediction error **on higher level** due to prior of observation

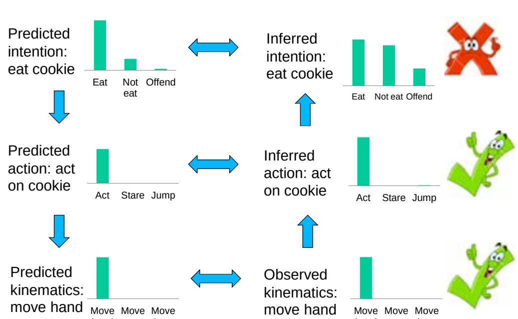
Example: right precision



Example: too precise



Example: too general



Hypothesis updating: modulating precision

- We can lower prediction error by updating beliefs (I understand your action to be refusing a cookie, rather than selecting one)
- We can lower prediction error by model updating (Some people consider it polite to first reject something being offered, only to accept when insisted)
- We can lower prediction error by intervention ("Please, take a cookie, they're delicious!")
- **We can lower prediction error by modulating precision (I understand that you want to pick *a* cookie, rather than *that specific* chocolate cookie)**

What is the right level of detail?

- In vanilla-flavored predictive coding, so-called hyper-priors determine which precision is warrented in which contexts
- A similar notion can be thought of for deciding which level of detail is appropriate in which context
- But note: “What level of detail is appropriate” comes hideously close to “What information is relevant here”
- Are we looking at the **Frame Problem** in disguise?

Frame Problem

- The Frame Problem, in the formulation such as given by Jerry Fodor or Daniel Dennett, is the problem of determining what is relevant in a particular context for a particular goal
- In principle, anything might be relevant to everything – and sometimes strange relations can be and are made. This is a huge intractable computational burden!
- We have as yet no clue whatsoever how cognition deals with this problem, or how to solve it in artificial intelligence…

Donders Institute for Brain, Cognition and Behaviour Radboud University Nijmegen

Issues to be resolved

- Hyper-priors and aggregation-priors are typically assumed to solve the problem of determining what the appropriate level of uncertainty is and what is relevant in a particular context
- However, they potentially hide/mask the Frame Problem
- Andy Clark (personal communication): “Predictive processing does not have anything to say about the Frame Problem” → this does not mean it is not relevant, of course…

Donders Institute for Brain, Cognition and Behaviour Radboud University Nijmegen

Challenges for PP in higher cognition

- How to translate the basic concepts of PP to structured domains? What representations do we need?
  - General principle = inference in (dynamical) graphical models (such as Bayesian networks)
  - Prediction = computation of posterior
  - Prediction error = Kullback Leibler divergence between prediction and observation
  - Uncertainty in prediction = **entropy**
  - Level of detail in prediction = **aggregation level**
  - Hypothesis updating = **belief/model revision, active inference (intervention), less/more detail**
  - Learning generative models = ***future research***

Donders Institute for Brain, Cognition and Behaviour Radboud University Nijmegen