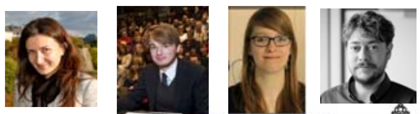
## Predictive Processing and Autism Spectrum Conditions

*Mini Symposium*
Thursday June 25th 2015



Donders Institute for Brain, Cognition and Behaviour
Radboud University Nijmegen

---

### Symposium Schedule

| | |
|---|---|
| 13.00-13.45 | Opening and "predictive processing primer" (Johan Kwisthout) |
| 13.45-14.30 | Sander van der Cruys (KU Leuven) *Can predictive coding provide a viable single-deficit account of ASD?* |
| 14.30-15.15 | Edita Poljac (DCC) *Adaptive Cognition in Autism* |
| Break plus poster session | |
| 16.15-17.15 | Panel discussion with Jan Buitelaar (UMC) and Floris de Lange (DCCN) |

Donders Institute for Brain, Cognition and Behaviour
Radboud University Nijmegen

---

"To be *precise*, the *details* don't matter"

## A primer on Predictive Processing

**Johan Kwisthout**

Symposium on Predictive Processing and Autism Spectrum Conditions

Donders Institute for Brain, Cognition and Behaviour
Radboud University Nijmegen

---

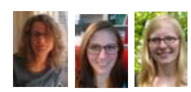
### Acknowledgements & Contributions

Lorentz workshop on HPI

Theoretical & computational investigations

Leuven workshop on PC/ASD

Behavioral & imaging studies (action understanding)



PP as explanatory mechanism in development of a "sense of agency"

Donders Institute for Brain, Cognition and Behaviour
Radboud University Nijmegen

---

### Outline of this presentation

- Predictive Processing as a concept
- Computational-level characterization of the processes involved in Predictive Processing
- Prediction errors & how to deal with them
- Precision of prediction and of prediction error
- Level of detail of predictions
- Open problems

Donders Institute for Brain, Cognition and Behaviour
Radboud University Nijmegen

---

### Motivation of much of our 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?"

Donders Institute for Brain, Cognition and Behaviour
Radboud University Nijmegen

## Predictive Processing vs. Predictive Coding

- Not **Friston**-ese, but **Clark**-onian / **Hohwy**-esque
- Focus on conceptual principle rather than Friston's "free-energy driven predictive coding" which is closely tied to the cortical hierarchy – we wish to abstract away from that and stay at Marr's computational level
- Computational translation of conceptual principles



## Predictive Processing

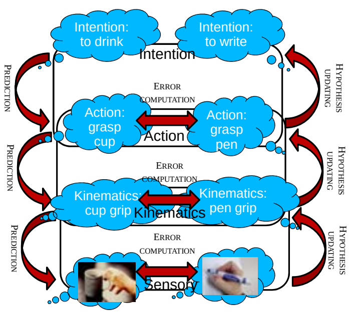
Brain as **prediction machine**
- *The brain continuously makes predictions about future sensory evidence based on its current best model of the causes of such evidence*

**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*



## 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 sub-processes

- Making predictions of expected input
- Comparing predicted inputs with actual inputs and computing the divergence between them
- Dealing with prediction errors
- Keywords
  - How to lower prediction error?
  - Precision of prediction and of prediction error
  - Level of detail of prediction



## 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 dependences

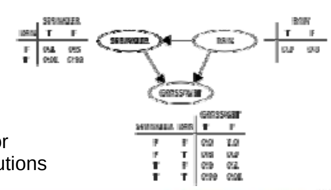
→ We argue (e.g., Otworowska et al., 2014) that simple Gaussian models are **not sufficiently rich** models

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

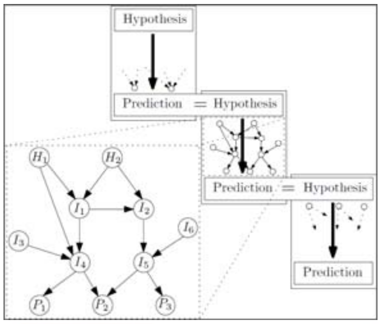


## Bayesian networks

- Bayesian networks describe stochastic variables and their dependence relations
- May describe both causal relations or stochastic co-occurrences (without causal interpretation)
- Variables may be
  - discrete
  - continuous
  - Gaussian
  - Otherwise
- Compute posterior probability distributions



## Computational model – prediction generation




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

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

**Intermediate**
variables
*Int* = {$I_1$, … $I_6$}



## 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
- Estimating the [size of the] error is defined as computing a KL- divergence or **relative entropy** between predicted distribution 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
- Four possible ways of "doing something" (Kwisthout, 2014):
  - Belief revision (revise hyp probability distribution)
  - Model revision (revise parameters in the CPTs)
  - Passive intervention (evidence gathering)
  - Active intervention (acting, i.e., setting variables)
- Each of them with the goal of lowering relative entropy



## Intuitive examples of lowering prediction error

- **Belief revision** – in cases with 'expected uncertainty' where the world model is well understood but inherently stochastic
- Prediction errors can be dealt with by **changing** the *Hyp* **distribution** to a (maybe a priori less probable) distribution that explains the observations
- E.g., your prior expectations about the weather in Scotland ("rainy") may predict a different picture, but this observation can be easily explained



## Intuitive examples of lowering prediction error

- **Model revision** – in cases with 'unexpected uncertainty' where the world model is still being learned or needs to be updated
- Prediction errors can be dealt with by changing some of the **model parameters** (tuning) such that the model can better predict the observations
- E.g., your model about what constitutes a friendly greeting may need updating (for the $30^+$-people amongst us)



## Intuitive examples of lowering prediction error

- **Passive intervention** – reduce prediction error by reducing uncertainty in the world: **add** additional observations
- This is what we intuitively do when confronted with the "train effect": when you're sitting in a train that is standing still at the station and you are looking at an opposite train – who is moving?
- You'd probably look at a stationary point to reduce uncertainty (e.g., the railway station buildings)

## Intuitive examples of lowering prediction error

- **Active intervention** – reduce prediction error by actively intervening in the world (active inference): bring prediction and observation closer together by **changing** the observation

- This has been proposed as a means of coupling action and perception in a single framework, where motor acts are the result of a mismatch between a "predicted" (expected) state and the actual, perceived state of the world



## Precision of prediction & prediction error

- **Precision** is a property of both a **prediction** and of a **prediction error**

- **Precision of a prediction** describes how much uncertainty there is in a prediction (and consequently, how informative the actual observation of what was predicted will be)

- **Precision of a prediction error** describes what proportion of this uncertainty can be attributed to inherent stochastic nature of the process that caused the outcome of the prediction



## Tossing coins

- What's the predicted outcome of tossing a **fair coin?**

**P(**  **) = 0.5**   **P(**  **) = 0.5**

- What's the predicted outcome of tossing a **coin that you know is biased but not to what side?**

**P(**  **) = 0.5**   **P(**  **) = 0.5**



## Why was that?

- The **subjective interpretation** of probabilities demands that either distribution is uniform as you have no information about *which* side is more likely

- However, where in the first case all the uncertainty is inherent in the **stochastic nature** of throwing coins, the uncertainty in the second case is due to ignorance about the "true" direction of the bias

- This is where the precision of the prediction error jumps in: it describes your **model confidence** and how much weight must be given to prediction errors



## Tossing coins again

- What's the predicted outcome of tossing a **fair coin?**
  - ***Uniform*** probability distribution of prediction
  - ***One*** bit of information in actual outcome
  - ***Prediction error*** (KL-divergence) of 1 bit
  - ***Low*** weight of p.e. as we have high confidence in the model

- What's the predicted outcome of tossing a **coin that you know is biased but not to what side?**
  - ***Uniform*** probability distribution of prediction
  - ***One*** bit of information in actual outcome
  - ***Prediction error*** (KL-divergence) of 1 bit
  - ***High*** weight of p.e. as we have low confidence in the model – we can use the actual outcome to improve our model!



## Precision

- Predictions are made with a particular **precision**

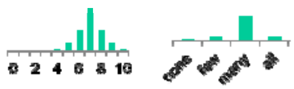
- "Precision" is to be understood in the *statistical* sense

- Predict the outcome of the throw

- What are *precise* and *imprecise* predictions here?
(in the context of predictive processing as defined earlier)

## NOT precision, but level of detail!

- Predictions are made with a particular **precision**
- “Precision” is to be understood in the *statistical* sense




a) Detail High    b) Detail Low

a) is **NOT** ‘more precise’ than b)!



## Precision

- Predictions are made with a particular **precision**
- “Precision” is to be understood in the *statistical* sense

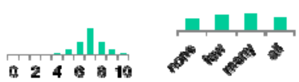
- Precision is a measure of uncertainty about the prediction
- Precision is defined at every possible level of detail
- Natural measure: **entropy**
  **E(*Pred*) = $-\sum_{p \in Pred}$ Pr(*p*)·log$_2$ Pr(*p*)**



## NOT precision, but level of detail!

- Predictions are made with a particular **precision**
- “Precision” is to be understood in the *statistical* sense


**Entropy 1.70**    **Entropy 1.19**




**Entropy 3.38**    **Entropy 1.98**



## Level of detail

- “Precision” can be formalized as “entropy” (or its continuous counterpart) as a measure of uncertainty
- Level of detail can be formalized as modulation of the **aggregation level** of Bayesian random variables, their values and inter-dependences
- Quite literally: zooming in and out on probability distributions!

(Kwisthout & Van Rooij, in press)



## Modulating level of detail


High detail $H$, high detail $P$

$\mathrm{P}(P|H) =$

| | $p_1$ | $p_2$ | $p_3$ | $p_4$ |
|---|---|---|---|---|
| $h_1$ | 0.3 | 0.2 | 0.4 | 0.1 |
| $h_2$ | 0.6 | 0.1 | 0.2 | 0.1 |
| $h_3$ | 0.2 | 0.4 | 0.3 | 0.1 |
| $h_4$ | 0.1 | 0.1 | 0.4 | 0.4 |

Low detail $P$, high detail $H$

$\mathrm{P}(P|H) =$

<table>
<tr><th></th><th>$p_1$</th><th>$p_2$</th><th>$p_3$</th><th>$p_4$</th></tr>
<tr><td>$h_1$</td><td colspan="2">0.5</td><td colspan="2">0.5</td></tr>
<tr><td>$h_2$</td><td colspan="2">0.7</td><td colspan="2">0.3</td></tr>
<tr><td>$h_3$</td><td colspan="2">0.6</td><td colspan="2">0.4</td></tr>
<tr><td>$h_4$</td><td colspan="2">0.2</td><td colspan="2">0.8</td></tr>
</table>

$\mathrm{P}(p_1 \vee p_2|H) = \mathrm{P}(p_1|H) + \mathrm{P}(p_2|H)$

High detail $P$, low detail $H$

$\mathrm{P}(P|H) =$

<table>
<tr><th></th><th>$p_1$</th><th>$p_2$</th><th>$p_3$</th><th>$p_4$</th></tr>
<tr><td>$h_1$</td><td rowspan="2">0.45</td><td rowspan="2">0.15</td><td rowspan="2">0.3</td><td rowspan="2">0.1</td></tr>
<tr><td>$h_2$</td></tr>
<tr><td>$h_3$</td><td rowspan="2">0.15</td><td rowspan="2">0.25</td><td rowspan="2">0.35</td><td rowspan="2">0.25</td></tr>
<tr><td>$h_4$</td></tr>
</table>

$\mathrm{P}(P|h_1 \vee h_2) = \frac{1}{2}(\mathrm{P}(P|h_1) + \mathrm{P}(P|h_2))$



## Modulating level of detail


High detail $H$, high detail $P$

$\mathrm{P}(P|H) =$

| | $p_1$ | $p_2$ | $p_3$ | $p_4$ |
|---|---|---|---|---|
| $h_1$ | 0.3 | 0.2 | 0.4 | 0.1 |
| $h_2$ | 0.6 | 0.1 | 0.2 | 0.1 |
| $h_3$ | 0.2 | 0.4 | 0.3 | 0.1 |
| $h_4$ | 0.1 | 0.1 | 0.4 | 0.4 |

Low detail $P$, low detail $H$

$\mathrm{P}(P|H) =$

<table>
<tr><th></th><th>$p_1$</th><th>$p_2$</th><th>$p_3$</th><th>$p_4$</th></tr>
<tr><td>$h_1$</td><td colspan="2" rowspan="2">0.6</td><td colspan="2" rowspan="2">0.4</td></tr>
<tr><td>$h_2$</td></tr>
<tr><td>$h_3$</td><td colspan="2" rowspan="2">0.4</td><td colspan="2" rowspan="2">0.6</td></tr>
<tr><td>$h_4$</td></tr>
</table>

## How to lower prediction error

- Apart from:
  - Model revision
  - Belief revision
  - Active intervention (active inference)
  - Passive intervention (obtaining evidence)

  we can also **increase precision** by **decreasing the level of detail** of the prediction or by **increasing** the level of detail of the hypothesis (the model)

- E.g., when we predicted "7 pins" – which is somewhat too detailed given the information we have – we may **lower prediction error** when 8 pins fall by revising "7 pins" to "many pins" → this decreases uncertainty and thus **increases precision**…



## How to lower prediction error

- Apart from:
  - Model revision
  - Belief revision
  - Active intervention (active inference)
  - Passive intervention (obtaining evidence)

  we can also **increase precision** by **decreasing the level of detail** of the prediction or by **increasing** the level of detail of the hypothesis (the model)

- E.g., when we have a very general model of a bowler, throwing on average 5 pins plus or minus 4 we may **lower prediction error** by selecting a more detailed model of the bowler (e.g., "a quite proficient bowler throwing on average 7 plus or minus 2")



## The optimal prediction

- Naturally, there is a trade-off between "level-of-detail" and "precision"

➢ At the lowest level, we can make a very uninformative prediction ("stuff will happen") that is by definition **correct** but not very **useful**

➢ In contrast, we can make a very **detailed prediction** that is almost certainly not correct, leading to prediction error and thus **cognitive load** to reduce this error

- The optimal **expected prediction error / information gain** will occur when the two are balanced



## Open research problems

- What mechanism for error minimization does the brain employ under which circumstances?
- What is the optimal balance between information and possible prediction error and how is it achieved?
- Is attention related to level-of-detail, rather than to precision? How could we investigate this?
- Are AS conditions related to 'problems' with the precision of the prediction error (as Sander will talk about) or to problems with the level of detail of the predictions, or both (may it be the same 'problem' that appears in different guises)?