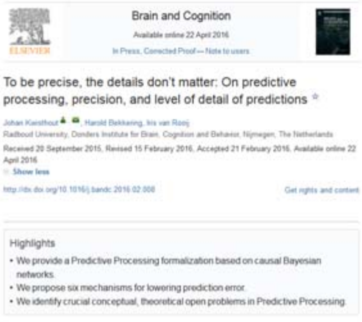
DONDERS INSTITUTE


Brain and Cognition
Available online 22 April 2016
In Press, Corrected Proof — Note to users

To be precise, the details don't matter: On predictive processing, precision, and level of detail of predictions

Johan Kwisthout, Harold Bekkering, Iris van Rooij
Radboud University, Donders Institute for Brain, Cognition and Behavior, Nijmegen, The Netherlands
Received 20 September 2015, Revised 15 February 2016, Accepted 21 February 2016, Available online 22 April 2016
Show less
http://dx.doi.org/10.1016/j.bandc.2016.02.008
Get rights and content

Highlights
- We provide a Predictive Processing formalization based on causal Bayesian networks.
- We propose six mechanisms for lowering prediction error.
- We identify crucial conceptual, theoretical open problems in Predictive Processing.

UvA Predictive Processing Workshop

Radboud University Radboudumc

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

→ We propose to use **causal Bayesian networks** as *structured* generative models to describe predictive processing

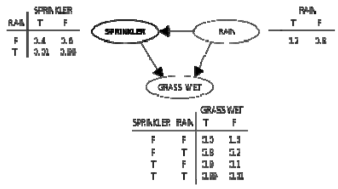
Predictive Processing

- Brain as **prediction machine**
- (approximate) **Bayesian** brain
- **Hierarchically** organized brain
- Prediction-error **minimizing** brain
- No **specific** algorithmic and implementational claims (contrasting predictive coding)

Causal Bayesian networks (Pearl, 2000)

- Bayesian networks efficiently describe stochastic variables and their conditional (in-)dependence relations
- In **causal Bayesian networks**, the arcs have a causal interpretation (not just stochastic dependency)
- In our modeling we assume discrete (categorical) probability distributions

(cfm. Friston et al., 2015)

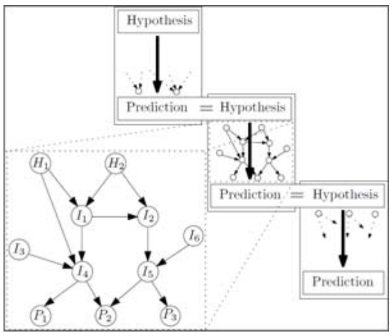


Friston et al. (2015). Active Inference and epistemic value. *Cognitive Neuroscience*

Key sub-processes

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

Computational model



Each layer in the hierarchy is a causal Bayesian network

Hypothesis variables
Hyp = {H_1, H_2}

Prediction variables
Pred = {P_1, P_2}

Intermediate variables
Int = {I_1, … I_6}

Making predictions and computing errors

- 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 a KL- divergence or **relative entropy** between predicted distribution and observed distribution

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

Computational model – error minimization

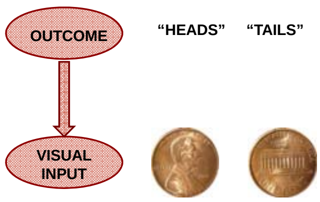
- Prediction error minimization: “doing something” such that D_{KL}(*Obs* || *Pred*) is minimized
- Six possible ways of “doing something” (Kwisthout, van Rooij, & Bekkering, 2017):
  - Belief revision (revise hyp probability distribution)
  - Model revision (revise parameters in the CPTs)
  - Passive intervention (evidence gathering)
  - Active intervention (acting, i.e., setting variables)
  - More fine-grained models / less detailed predictions
- Each of them with the goal of lowering relative entropy

Entropy & Precision-weighted prediction errors

- **Entropy 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)
- The more **details** (categories) in the prediction, the more **information** in the observation (and the **higher** the prediction error 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 → **precision-weighted prediction errors**

Intuitive examples of lowering prediction error

- **Belief revision** – in cases with ‘expected uncertainty’ where the world model is stable but there is information carried by the prediction error



Hyperprior ↔ Precision-weighted prediction error

- **Hyperprior** on distribution describing how **confident** we are in this generative model (here: Dirichlet distribution)
- Precision-weighted prediction error describes the **size of the effect** of a prediction error on the updating of the model
- Formally defined as the **KL divergence** between the hyperprior ‘before’ and ‘after’ updating with the new data
- The higher this weighted prediction error, the **bigger the effect** on the generative model a prediction error is and the **more reducible uncertainty** there is in the environment
- Note: **idealized** mathematical definition…

Intuitive examples of lowering prediction error

- **Model revision** – in cases with ‘unexpected uncertainty’ where the world model suddenly turns out to be misinformative
- 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)

Precision in discrete distribution

- Predictions are made with a particular **precision**
- Precision in Gaussian distributions = 1 / variance = 1 / σ
- Precision in discrete distributions is a function of how **detailed** the prediction is and how **uncertain**
- Measure of uncertainty in a distribution: **entropy**
  E(*Pred*) = $-\sum_{p \in Pred}$ Pr(*p*)·log$_2$ Pr(*p*)
- Measure of state space granularity: **cardinality**
  Ω(*Pred*) = number of values that variable *Pred* can take
- **Shannon redundancy** measures ‘pure’ (granularity-normalized) uncertainty

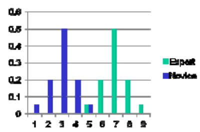
Intuitive examples of lowering prediction error

- **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

More / less detailed models

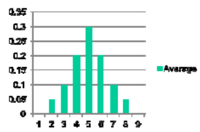
- **Predictions** have a state space, but so do **hypotheses**
- The level of detail of the hypothesis state space is a measure on how fine-grained or coarse or models are

Fine-grained model of bowlers



In-group?

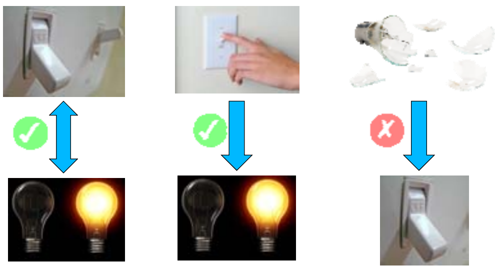
Coarse model of bowlers



Out-group?

Active and passive intervention

Causal Bayesian network – relationship along arcs is causal: turning the switch causes the light bulb to go on, not vice versa!



Modulating level of detail in models and predictions

High detail H, high detail P

$P(P|H) =$

	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

$P(P|H) =$

<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>

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

High detail P, low detail H

$P(P|H) =$

<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 h_2</td><td>0.45</td><td>0.15</td><td>0.3</td><td>0.1</td></tr>
<tr><td>h_3 h_4</td><td>0.15</td><td>0.25</td><td>0.35</td><td>0.25</td></tr>
</table>

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

Summary

- When formalizing Predictive Processing in terms of categorical (structured, discrete) generative models, the **state space granularity** becomes important
- More detailed predictions allow for more information processing, at the cost of **higher prediction errors**
- **Lowering detail of predictions** is one way of dealing with (uninformative) prediction errors
- **Making more refined models** is one way of increasing informative-ness of the predictions
- Information / prediction error **trade-off** in the brain