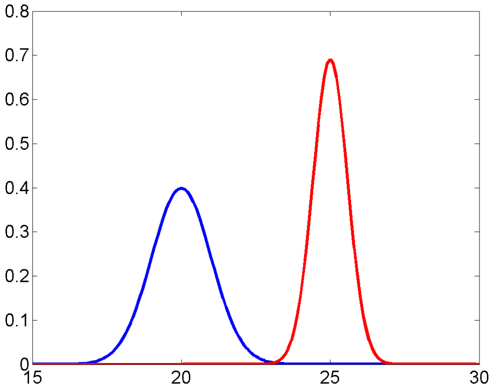
Bayesian Models of Brain and Behaviour

Bayesian Course
Wellcome Trust Centre for Neuroimaging at UCL
Feb 2013.

Optimal Data Fusion

For the prior (blue) we have $m_0 = 20$, $\lambda_0 = 1$ and for the likelihood (red) $m_D = 25$ and $\lambda_D = 3$.



Precision, λ, is inverse variance.

Bayes rule for Gaussians

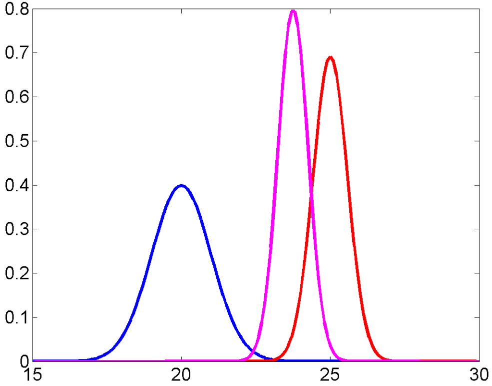
For a Gaussian prior with mean m_0 and precision λ_0, and a Gaussian likelihood with mean m_D and precision λ_D the posterior is Gaussian with

$$\begin{aligned} \lambda &= \lambda_0 + \lambda_D \\ m &= \frac{\lambda_0}{\lambda} m_0 + \frac{\lambda_D}{\lambda} m_D \end{aligned}$$

So, (1) precisions add and (2) the posterior mean is the sum of the prior and data means, but each weighted by their relative precision.

Bayes rule for Gaussians

For the prior (blue) $m_0 = 20$, $\lambda_0 = 1$ and the likelihood (red) $m_D = 25$ and $\lambda_D = 3$, the posterior (magenta) shows the posterior distribution with $m = 23.75$ and $\lambda = 4$.

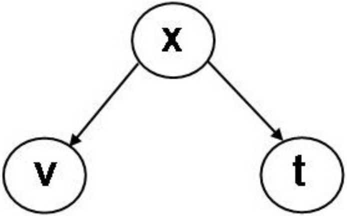


The posterior is closer to the likelihood because the likelihood has higher precision.

Sensory Integration

Ernst and Banks (2002) asked subjects which of two sequentially presented blocks was the taller. Subjects used either vision alone, touch alone or a combination of the two.

If vision v and touch t information are independent given



an object x then we have

$$p(v,t,x) = p(v|x)p(t|x)p(x)$$

Bayesian fusion of sensory information then produces a posterior density

$$p(x|v,t) = \frac{p(v|x)p(t|x)p(x)}{p(v,t)}$$

Sensory Integration

In the absence of prior information about block size (ie $p(x)$ is uniform), for Gaussian likelihoods, the posterior will also be a Gaussian with precision λ_{vt}. From Bayes rule for Gaussians we know that precisions add

$$\lambda_{vt} = \lambda_v + \lambda_t$$

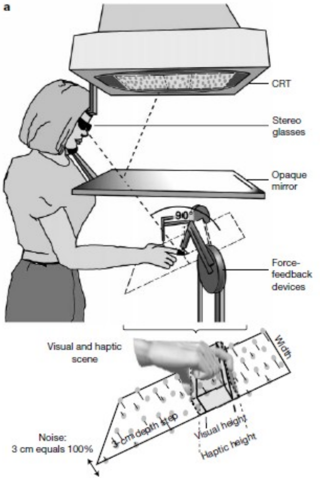
and the posterior mean is a relative-precision weighted combination

$$\begin{aligned} m_{vt} &= \frac{\lambda_v}{\lambda_{vt}} m_v + \frac{\lambda_t}{\lambda_{vt}} m_t \\ m_{vt} &= w_v m_v + w_t m_t \end{aligned}$$

with weights w_v and w_t.

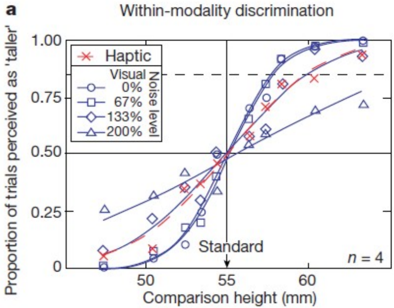
Vision and Touch

Ernst and Banks (2002) asked subjects which of two sequentially presented blocks was the taller. Subjects used either vision alone, touch alone or a combination of the two.



Vision and Touch Separately

They recorded the accuracy with which discrimination could be made and plotted this as a function of difference in block height. This was first done for each condition alone. One can then estimate precisions, λ_v and λ_t by fitting a cumulative Gaussian density function.

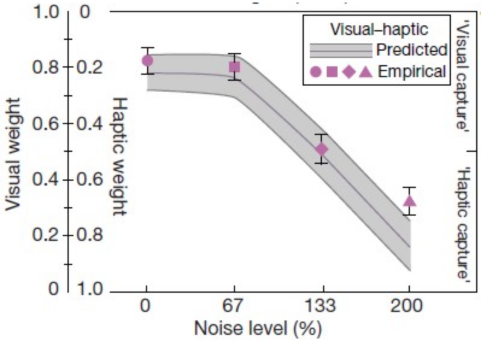


They manipulated the accuracy of the visual discrimination by adding noise onto one of the stereo images.

Vision and Touch Together

Optimal fusion predicts weights from Bayes rule

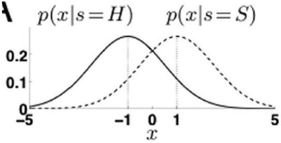
$$\lambda_{vt} = \lambda_v + \lambda_t$$
$$m_{vt} = \frac{\lambda_v}{\lambda_{vt}} m_v + \frac{\lambda_t}{\lambda_{vt}} m_t$$
$$m_{vt} = w_v m_v + w_t m_t$$



They observed visual capture at low levels of visual noise and haptic capture at high levels.

Likelihood Ratio Test

Given a sample x, from what density was it drawn ?



$$
\begin{aligned}
p(x|s=H) &= N(x; -1, \sigma^2) \\
p(x|s=S) &= N(x; 1, \sigma^2)
\end{aligned}
$$

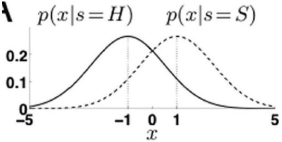
The Likelihood Ratio Test (LRT) is optimal for making this decision.

$$R = \frac{p(x|s=S)}{p(x|s=H)}$$

R is an odds ratio.

Bayesian Test

Given a sample x, from what density was it drawn ?



$$
\begin{aligned}
p(x|s=H) &= \mathrm{N}(x;-1,\sigma^2) \\
p(x|s=S) &= \mathrm{N}(x;1,\sigma^2)
\end{aligned}
$$

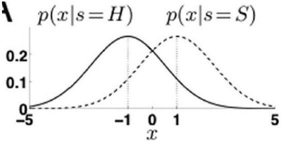
Given priors, we can compute the posterior odds

$$\frac{p(s=S|x)}{p(s=H|x)} = \frac{p(x|s=S)}{p(x|s=H)} \frac{p(s=S)}{p(s=H)}$$

This generalises LRT.

Sequential Bayes

Given a series of samples x_n, from what density are they drawn ?



For first sample

$$p(s|x_1) = \frac{p(x_1|s)p(s)}{\sum_{s'} p(x_1|s')p(s')}$$

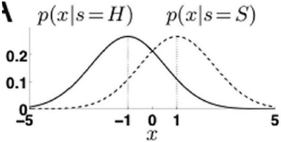
For second sample

$$p(s|x_1, x_2) = \frac{p(x_2|s)p(s|x_1)}{\sum_{s'} p(x_2|s')p(s')}$$

Posterior from first sample is prior for second sample.

Sequential Bayes

Given a series of samples x_n, from what density are they drawn ?



Let $X_n = \{x_1, x_2, ..., x_n\}$

$$p(s|X_n) = \frac{p(x_n|s)p(s|X_{n-1})}{\sum_{s'} p(x_n|s')p(s')}$$

Today's prior is yesterdays posterior.

$$\frac{p(s=S|X_n)}{p(s=H|X_n)} = \frac{p(x_n|s=S)}{p(x_n|s=H)} \frac{p(s=S|X_{n-1})}{p(s=H|X_{n-1})}$$

Without prior at n=1, this is sequential LRT.

Flanker Task

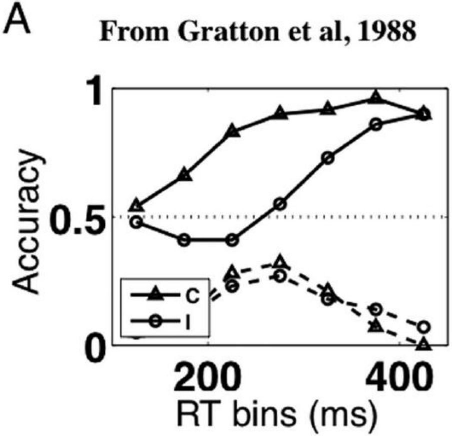
In the Eriksen Flanker task subjects have to implement the following stimulus-response mappings

Stimulus	*Response*
1.*HHH*	*Right*
2.*SHS*	*Right*
3.*SSS*	*Left*
4.*HSH*	*Left*

Put simply, the subject should press the right button if the central cue is H and left if it is S. On trial type one and three the flankers are compatible ($M = C$) and on two and four they are incompatible ($M = I$).

Decision Making Dynamics

If subjects are too slow an auditory beep is emitted. This is the *deadlined* Flanker task.

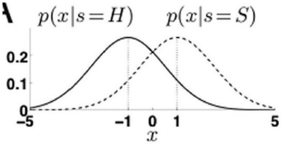


On incompatible trials initial average accuracy dips below the chance level.

Likelihood

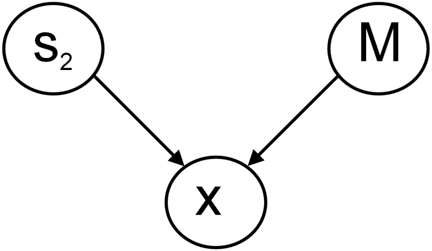
Yu et al. (2009) assume three populations of neurons, x, that are driven by the three stimuli, s

$$p(x|s) = \prod_{i=1}^{3} \mathrm{N}(x_i; \mu_i, \sigma^2)$$



$$\begin{aligned} p(x|s = SHS) &= p(x|s_2 = H, M = I) \\ &= \mathrm{N}(x_1; 1, \sigma^2)\mathrm{N}(x_2; -1, \sigma^2)\mathrm{N}(x_3; 1, \sigma^2) \end{aligned}$$

Generative Model



Joint probability

$$p(x, s_2, M) = p(x|s_2, M)p(s_2)p(M)$$

Likelihood

$$p(x|s_2, M) = \prod_{i=1}^{3} p(x_i|s_2, M)$$

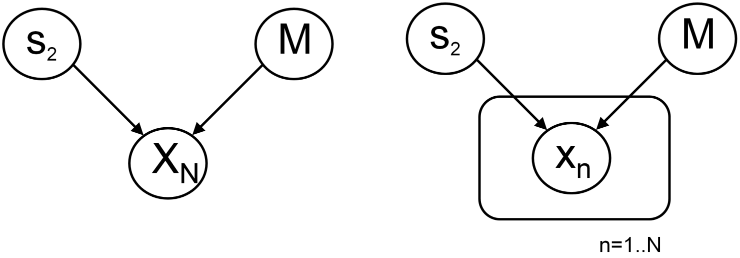
Dynamics

Consider a discrete set of time points $t(n)$ within the trial with $n = 1, 2, ..N$.

Denote x_n as population vector observed at time $t(n)$ and $X_n = [x_0, x_1, ..., x_n]$ as all vectors observed up until time point $t(n)$.

Yu et al. (2009) formulate a discrete time inferential model.

Generative Model



Joint probability

$$p(X_N, s_2, M) = p(X_N|s_2, M)p(s_2)p(M)$$

Likelihood

$$p(X_N|s_2, M) = \prod_{n=1}^{N} p(x_n|s_2, M)$$

Inference

The following joint probability is updated recursively

$$p(s_2, M|X_n) = \frac{p(x_n|s_2, M)p(s_2, M|X_{n-1})}{\sum_{s_2', M'} p(x_n|s_2', M')p(s_2', M'|X_{n-1})}$$

Then marginalise over M to get decision probability

$$p(s_2 = H|X_n) = p(s_2 = H, M = C|X_n) + p(s_2 = H, M = I|X_n)$$

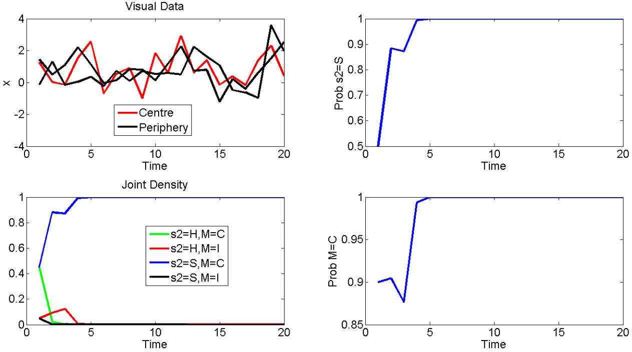
Initialise with

$$\begin{aligned} p(s_2 = H, M = C|X_0) &= p(s_2 = H)p(M = C) \\ p(s_2 = H, M = C|X_0) &= 0.5\beta \\ p(s_2 = H, M = I|X_0) &= 0.5(1 - \beta) \end{aligned}$$

where $p(M = C) = \beta$.

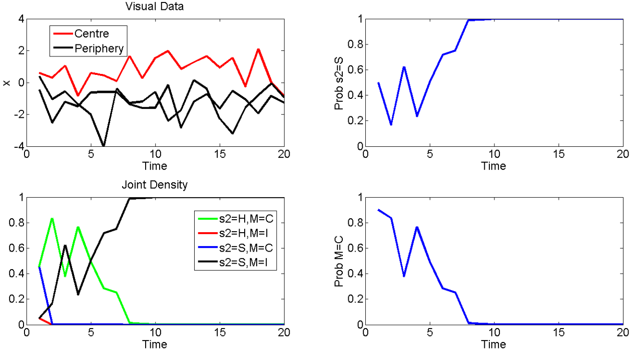
Compatible Trial

Stimulus set=SSS.



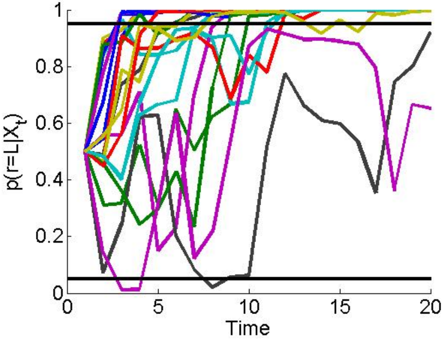
Incompatible Trial

Stimulus set=HSH.



Inference

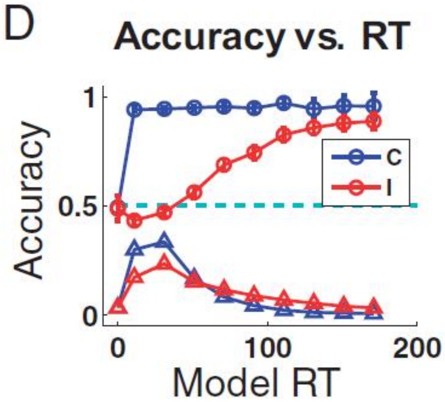
On most trials (18 out of 20) evidence slowly accumulates in favour of the central stimulus being $s_2 = S$. This is reflected in the posterior probability $p(s_2 = S|X_n)$.



This corresponds to evidence for a left button press.

Compatibility Bias Model

For compatibility bias $\beta > 0.5$



The model also shows the initial dip for incompatible flankers.

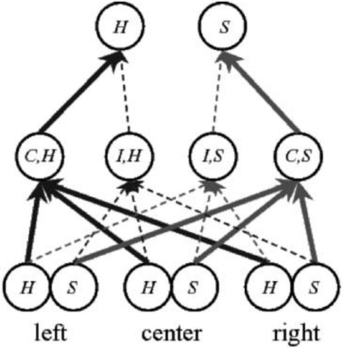
Neural Implementation

The Bayesian inference equations

$$p(s_2, M|X_n) = \frac{p(x_n|s_2, M)p(s_2, M|X_{n-1})}{\sum_{s_2', M'} p(x_n|s_2', M')p(s_2', M'|X_{n-1})}$$

$$p(s_2 = H|X_n) = p(s_2 = H, M = C|X_n) + p(s_2 = H, M = I|X_n)$$

can be implemented as a network model.



The hidden layer representations are *self-exciting* and require *divisive normalisation*. In the compatibility bias model the compatible pathway is initially excited.

Approximate Inference

As the number of stimuli grows exact inference becomes intractable. Instead, we can initially *assume* compatibility.

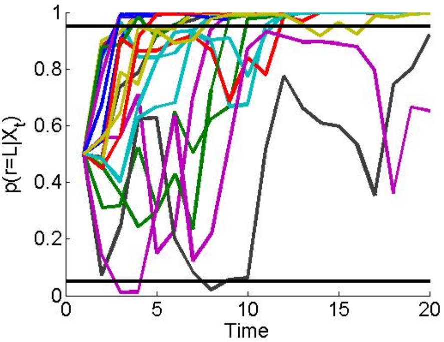
$$p(s_2 = H|X_t) = \frac{p(x_1(t)|s_1 = H)p(x_2(t)|s_2 = H)p(x_3(t)|s_3 = H)p(s_2 = H|X_{t-1})}{\sum_{s=H,S} p(x_1(t)|s_1 = s)p(x_2(t)|s_2 = s)p(x_3(t)|s_3 = s)p(s_2 = s|X_{t-1})}$$

If the flankers are detected to be incompatible we can then switch to an inferential scheme which ignores them

$$p(s_2 = H|X_t) = p(x_2(t)|s_2 = H)p(s_2 = H|X_{t-1})$$

Conflict detection

Compatibility can be inferred from a conflict detector

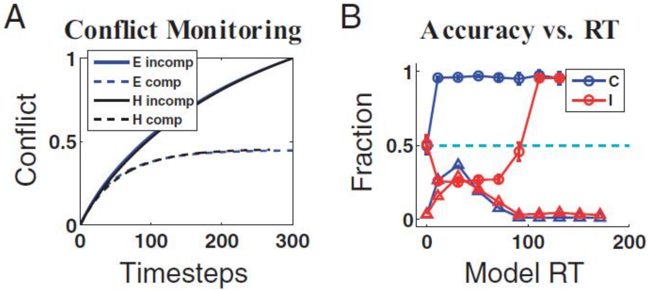


which measures the energy in the decision region (Botvinick et al. 2001)

$$E_t = E_{t-1} + p(s_2 = H|X_t)p(s_2 = S|X_t)$$

Approximate Inference

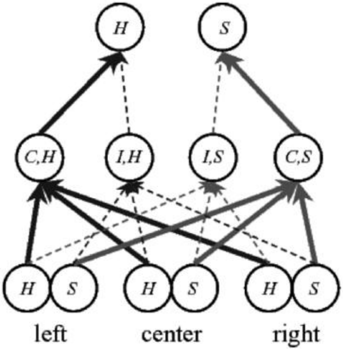
Detecting conflict using an energy measure gives similar results to using an entropy measure, H



Approximate inference yields behaviour similar to exact inference and empirical data.

Neural Implementation

Output of conflict monitoring enhances $M = C$ or $M = I$ pathway.



References

W. Penny (2012) Bayesian models of brain and behaviour. ISRN Biomathematics, Article ID 785791.

M. Botvinick et al. (2001) Psych Review 108, 624-652.

M. Ernst and M. Banks (2002) Nature 415, 429-433.

A. Yu, P. Dayan and J. Cohen (2009) J Exp Psych 35,700-717.

Bayes rule for Gaussians

For a Gaussian prior with mean m_0 and precision λ_0, and a Gaussian likelihood with mean m_D and precision λ_D the posterior is Gaussian with

$$
\begin{aligned}
m &= \frac{\lambda_0}{\lambda} m_0 + \frac{\lambda_D}{\lambda} m_D \\
&= m_0 - m_0 + \frac{\lambda_0}{\lambda} m_0 + \frac{\lambda_D}{\lambda} m_D \\
&= m_0 - \frac{\lambda_D}{\lambda} m_0 + \frac{\lambda_D}{\lambda} m_D \\
&= m_0 + \frac{\lambda_D}{\lambda}(m_D - m_0)
\end{aligned}
$$

Prediction m_0 is updated based on new data m_D.