Sensory processing and probabalistic codes.

In this section we will begin to explore the idea that the brain performs Bayesian inference on data. As an example of how this might apply to perception, we consider the experiment done by Ernst and Banks **?** in which people were asked to judge the height of a block. The height of the block is x and the subjects were asked to judge its height under three conditions, visually v, by touch, that is haptically, h and by vision and touch together hv. In the visual trials noise is added using goggles. The set up is pictured in Fig. **??**

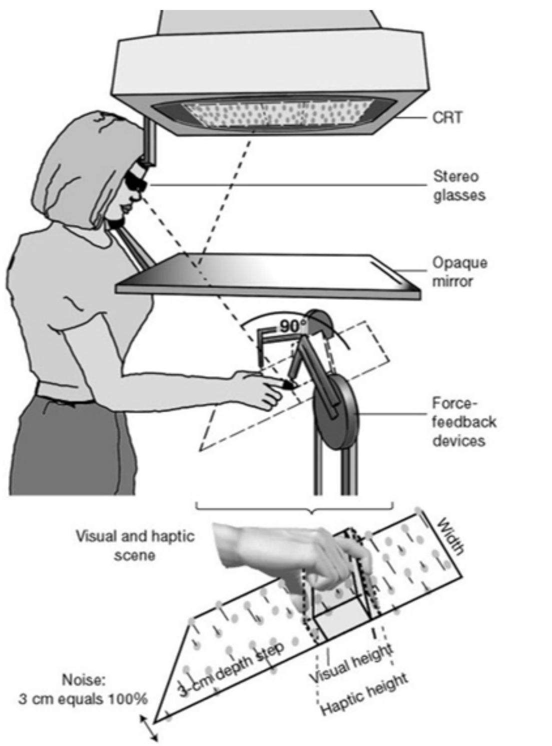
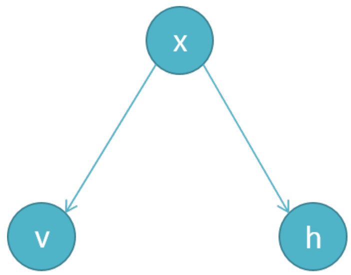


Figure 1: Illustration of the Ernst and Banks experiment. [Pictures from Rosalyn Moran who took it from the paper]

The joint probability is $p(v, h, x)$, it is assumed that V and H are conditionally independent:

$$p(v, h|x) = p(v|x)p(h|x) \tag{1}$$

In mathematics this would be modelled with a Markov chain $V \to X \to H$, in the world of Bayesian inference this is illustrated with a *directed acylcic graph* or *Bayesian network*:[1]



Either way, the idea is that although there is a relationship between V and H, if the block is big they will both be large, but that this relationship is only because they are both related to X. One consequence of this is

$$p(v, h, x) = p(v|x)p(h|x)p(x) \tag{2}$$

We are interested in the posterior judgement of the height of the block:

$$p(x|v, h) = \frac{p(v, h|x)p(x)}{p(v, h)} = \frac{p(v|x)p(h|x)p(x)}{p(v, h)} \tag{3}$$

It is assumed that the prior is uniform, so $p(x)$ is constant over some range of possible x, so if x is in that range

$$p(x|v, h) \propto \frac{p(v|x)p(h|x)}{p(v, h)} \tag{4}$$

The value of x that maximizes this is called the *maximum a posteriori*.

Now the purpose of the Ernst and Banks experiment is to study sensory fusion, the fusing of two noisy pieces of evidence. Bayesian analyses is used to show the optimal fusing of the data, this is what we will derive now. This is then compared to data to show that the brain approaches the problem in an optimal way.

It is assumed that the haptic and visual channels have independent Gaussian noise around the true value, that is $\mathcal{N}(x, \sigma_v^2)$ and $\mathcal{N}(x, \sigma_h^2)$ respectively, so

$$p(v|x) = \frac{1}{\sqrt{2\pi\sigma_v^2}} e^{-\frac{(v-x)^2}{2\sigma_v^2}} \tag{5}$$

[1]Picture from Rosalyn Moran

and

$$p(h|x) = \frac{1}{\sqrt{2\pi\sigma_v^2}} e^{-\frac{(h-x)^2}{2\sigma_h^2}} \tag{6}$$

and we ignore the normalizing factor $p(v, h)$ since it is independent of x, then

$$p(x|v, h) \propto p(v|x)p(h|x) \tag{7}$$

Now if we multiply out the two Gaussians the exponent is

$$-\frac{(h-x)^2}{2\sigma_h^2} - \frac{(v-x)^2}{2\sigma_v^2} = -\left(\frac{1}{2\sigma_h^2} + \frac{1}{2\sigma_v^2}\right)x^2 + \left(\frac{h}{\sigma_h^2} + \frac{v}{\sigma_v^2}\right)x + \text{other stuff} \tag{8}$$

so if we define σ by

$$\frac{1}{\sigma^2} = \frac{1}{\sigma_v^2} + \frac{1}{\sigma_h^2} \tag{9}$$

and $\bar{x}$ by

$$\bar{x} = \frac{\sigma^2}{\sigma_v^2}v + \frac{\sigma^2}{\sigma_h^2}h \tag{10}$$

this gives

$$-\frac{(h-x)^2}{2\sigma_h^2} - \frac{(v-x)^2}{2\sigma_v^2} = -\frac{(x-\bar{x})^2}{2\sigma^2} + \text{other stuff} \tag{11}$$

In other words, the optimal guess for x is the variance weighted measurements of v and h; the factors σ^2/σ_v^2 and σ^2/σ_h^2 quantify how much the variability of v and h contribute to the variability of $\bar{x}$, and the more they contribute, the lower the weighting is given to the measurement. This is summarized by writing $\lambda_v = \sigma^2/\sigma_v^2$ and $\lambda_h = \sigma^2/\sigma_h^2$, so

$$\lambda_v + \lambda_h = \frac{\sigma^2}{\sigma_v^2} + \frac{\sigma^2}{\sigma_h^2} = \frac{1}{1/\sigma_v^2 + 1/\sigma_h^2}\left(\frac{1}{\sigma_v^2} + \frac{1}{\sigma_h^2}\right) = 1 \tag{12}$$

so λ_v and λ_h are the visual and haptic weights.

In the actual experiment the visual noise is changed, allowing the experimenters to manipulate the two weights. They then do multiple experiments where they ask the subjects to decide which of two blocks is larger and manage through a regression analysis to decide how the subjects are weighting the two pieces of evidence. In other words it is assumed that the subjects are combining the haptic and visual estimates

$$\xi = \mu_v v + \mu_h h \tag{13}$$

where ξ is the subjects estimate of x and μ_v and μ_h are some weighting factors to be estimated from the behavioural data by regression; for details see the paper. They then compare the measured weightings, μ_v and μ_h, to those estimated by the optimal Bayesian model, that is λ_v and λ_h. The result is shown in Fig. **??** and demonstrate Bayesian optimal perception in this aspect of human sensory fusion.

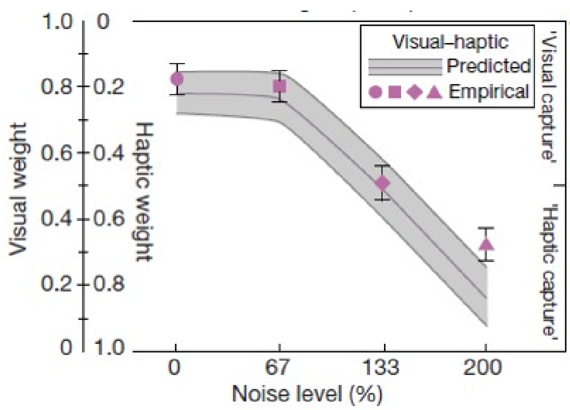


Figure 2: Results of the Ernst and Banks experiment. As the weights are changed by manipulating the visual noise by adjusting the goggles, the weighting of how the visual and haptic information are fused should change according to the grey line; by a clever set of tests it is possible to estimate the true weightings used by the subjects, these are the red features. The match is very good.[Pictures from Rosalyn Moran who took it from the paper]