Sensory processing: the Eriksen flanker task

In this section we will talk about another experiment providing evidence for Bayesian inference in the brain. In this case we will start off examine our ability to pick out an individual in a crowd[1]

though in the end we appear to observe something else. In any case we begin by thinking about the selective attention and monitoring of contrast and similarity in visual input. In the task the participants are shown one of two stimuli, here a '**S**' or a '**H**' and they press a button in reponse, for example left for **S** and right for **H**. The target letter is flanked by distractors on each side the participants are told to ignore. There are two cases, one where the distractors are compatible: **HHH** and **SSS** and one where they are incompatible: **HSH** and **SHS**.

Participants are slower and less accurate when responding to incompatible flankers. This is studied from a Bayesian point of view in **?**; they use data from older experiments in which the participants were instructed to perform the task as quickly as possible. One surprising result, see Fig. 1, is that for very short reaction times the response with incompatible flankers can be worse than chance.

There are two hypothesis as to what is going on, the first is that there is a compatibility bias: the brain assumes that the world is smooth, so having **S** flankers biases it towards the expectation that middle letter is also a **S**. This makes sense, visual information often has a high regularity.

[1] A still from Hitchcock's film *Strangers on a Train*.

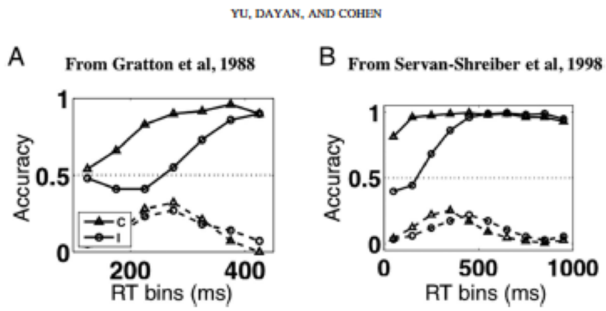

Figure 1: The solid lines show the accuracy against reaction time; the dashed lines the distribution of reaction times themselves. There are two traces, one marked C with compatible flankers, this is graphed with triangles and one marked I with incompatible flankers. The figure was taken (by Rosalyn) from **?** but they themselves have adopted it from other studies.

The second hypothesis argues that the areas of the brain that perform high level recognition assess visual stimuli over a large receptive field and so the recognition neurons are receiving cross talk from the distractors.

To see if these hypotheses account for the phenomenon we will construct a *generative model*, that is a model that could produce fictive data with a presumed similar distribution to the real data. Here, the generative model is a model of a neuronal response to the stimuli. We will couple the generative model to a *recognition model*, a model of how the brain might identify the target, based on the information encoded in the neurons. The recognition model describes the decision and ultimately determines which button will be pressed; this recognition model will perform Bayesian infernce.

Let the stimuli be labels s_1, s_2 and s_3, s_2 is the target, s_1 are the letter to the left and s_3 to the right. In the experiment $s_1 = s_3$, in the compatible condition $s_1 = s_2$, in the incompatible $s_1 \neq s_2$. The three neurons corresponding to three stimuli are labelled x_1, x_2 and x_3; these will have a response which is different, on average, depending on which letter is in each part of the stimulus, but they will respond in a variable way. In fact, for definiteness, lets say the ith stimulus $s_i = -1$ when the letter is **S** and $s_i = 1$ if it is **H**; the x_I are then Gaussian distibuted around these values. Of course, in implementing hypothesis two, each neuron will get an input from more than one s_i. The random variable M describes the trial condition, with $M = I$ for incompatible and $M = C$ for compatible. If we write $p(M = C) = \beta$ then for hypothesis one $\beta > 0.5$.

If $\mu_i = \mu(s_i)$ represented the -1 or one at the i subinterval letter position corresponding to **S** and **H** we can assume a distribution for the response, for hypothesis one this could be

$$p(\mathbf{x}|\mathbf{s}) = \mathcal{N}(\mu_1, \sigma)\mathcal{N}(\mu_2, \sigma)\mathcal{N}(\mu_3, \sigma) \tag{1}$$

where the σs represent the noise in the neuronal response and the responses are assumed to be conditionally independent. For hypothesis two there is a wider receptive field so we might have

$$p(\mathbf{x}|\mathbf{s}) = \mathcal{N}(t\mu_1 + d\mu_2, \sigma)\mathcal{N}(d\mu_1 + t\mu_2 + d\mu_3, \sigma)\mathcal{N}(d\mu_2 + t\mu_3, \sigma) \tag{2}$$

where here the d and t weight the influence of the distractor and the target on the neuronal response.

The final piece of the generator is time; the idea is that the neurons draw repeatedly from their distribution, independently each time, so, for example:

$$p(\mathbf{x}(t)|\mathbf{s}) = \mathcal{N}(\mu_1, \sigma)\mathcal{N}(\mu_2, \sigma)\mathcal{N}(\mu_3, \sigma) \tag{3}$$

for all t. We will also treat t as a discrete variable.

Now we consider the recognition model. This assumes an *ideal behaviour*, that is inference based on an optimal Bayesian strategy. Lets imagine that the target is **H** and we are interested in estimating s_2. Ultimately to calculated the probability that $s_2 = \mathbf{H}$ we will need to marginalize over the condition M; but for now lets calculate $p(s_2 = \mathbf{H}, M = C|\mathbf{x}(t))$. Using the Bayes rule we have

$$p(\mathbf{H}, C|\mathbf{x}(t), \text{previous } t) = \frac{p(\mathbf{x}|\mathbf{H}, C)P(\mathbf{H}, C|\text{previous } t)}{Z} \tag{4}$$

Here Z just stands for the normalization stuff on the bottom which we won't look at closely. Clearly the clever bit is that the recognition model is performing a Bayesian loop, updating the posterior for time t based on the previous prior and then using that as the prior when working out time $t - 1$. The distribution for $\mathbf{x}$ conditioned on the value of the target doesn't depend on previous results since we are assuming the generator model is time independent; we know this distribution in that we've assumed, as above, that it is Gaussian.

The overall idea is that the model will keep working until it reaches a threshold $p(s_2 = \mathbf{H}|\text{previous times}) > 0.9$ for example and at that point the recognition model will call its decision.

Now at the start the prior is not based on any evidence

$$P(\mathbf{H}, C|\text{at the start}) = 0.5\beta \tag{5}$$

since at this time the condition, consistent or inconsistent, is independent of the identity of the target, **H** or **S**. We also have values for these quantities, we assume our priors are $p(M = C) = \beta$, representing our expectation of consistency and $P(s_2 = \mathbf{H}) = 0.5$ since we have no reason to expect one over the other. Now

$$p(\mathbf{H}, C|\mathbf{x}(t = 1), \text{at the start}) = \frac{0.5\beta p(\mathbf{x}|\mathbf{H}, C)}{Z} \tag{6}$$

This gives us a posterior for time $t = 1$. We would need to do the same thing for $P(s_2 = \mathbf{H}, M = I|\mathbf{x}(t = 1), \text{at the start})$ to calculate $P(s_2 = \mathbf{H}|\mathbf{x}(t = 1), \text{at the start})$. If that is greater than 0.9, or less than 0.1, we stop and declare an answer. Otherwise we iterate:

$$p(\mathbf{H}, C|\mathbf{x}(t = 2), t = 0 \text{ and } t = 1) = \frac{p(\mathbf{x}|\mathbf{H}, C)p(\mathbf{H}, C|\mathbf{x}(t = 1), t = 0)}{Z} \tag{7}$$

Clearly, although this hasn't been fleshed out in detail, it shows how we could build a model to calculate the reaction time distribution and the accuracy.

In **?** this is simulated for the two hypotheses. The results can be seen in Fig. 3 and Fig. **??**.

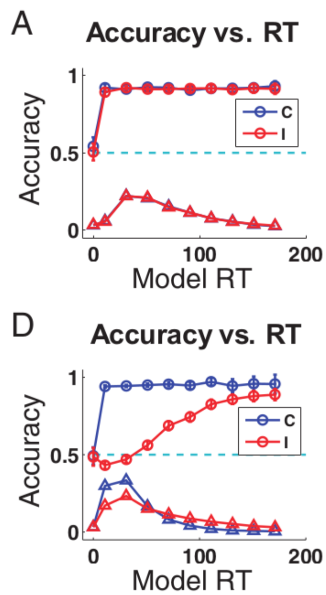

Figure 2: This shows the model results for the first hypothesis, with, in (**top**) $\beta = 1$ so there is no consistency bias and (**bottom**) with $\beta = 0.9$ so there is a bias. Taken from **?**.

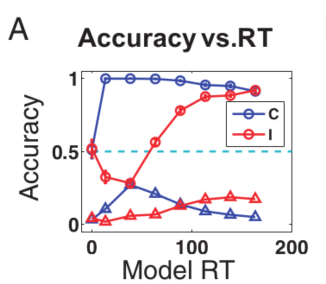

Figure 3: This shows the model results for the second hypothesis. Taken from **?**.