

# New Paper title

Gaurav Malhotra, Jennifer Coull & Emmanuel Daucé

August 11, 2011

## Abstract

Here the abstract...

## 1 Introduction

The word “temporal expectation” reflects the capacity of man and animal to orient the cognitive resources at a particular moment in time according to endogenous and exogenous cues. Endogenous cues, thought to activate top-down cognitive processes, associate arbitrary symbols to time intervals and are distinguished from exogenous cues where temporal information is a property of the stimulus (a moving car) and assumed to activate bottom-up processes, not requiring conscious control (Hauer & MacLeod, 2006). The use of endogenous cues thus involves complex cognitive mechanisms that need to be learnt and executed several times to make a correct prediction.

Specifying a mechanism for how subjects generate temporal expectations faces a number of challenges. These challenges lie in being able to understand and explain the results of behavioural and neurophysiological studies on the evolution of temporal expectations. The behavioural studies, largely based on reaction-time paradigm, investigate the change in response time of the subject to a change in the time at which a stimulus is presented. The neurophysiological studies, on the other hand, map the generation of temporal expectations to brain structures and processes. Therefore, a mechanism for generation of temporal expectations must explain the role of these brain structures and corresponding neural signals.

The processing of temporal expectation can be investigated through behavioural experiments that observe how long the subject takes to respond to a given stimulus. This reaction time (RT) forms a window through which the psychologist can investigate the information processing performed by the subject when it reacts to a stimulus.

In a typical temporal interval estimation task, the subject is given a warning stimulus (WS) about an upcoming response stimulus (RS), and the RT measures how long it takes the subject to respond to the RS. The temporal interval between the WS and the RS is referred as the *foreperiod* (i.e. the period before the RS). One of the main factor influencing RT is the variability of the time at which the RS is presented – i.e. the *time uncertainty* of the foreperiod (duration uncertainty of the foreperiod). Several studies have shown that increasing the variability of the foreperiod increased the reaction time of the subject (Klemmer, 1957; Karlin, 1959; Drazin, 1961).

We propose in the following a phenomenological approach that models changes in RT in terms of probabilities, stating that (1) “temporal expectation” can be mod-

elled as a probability of a certain event of interest to take place in the immediate next moment and (2) this probability is expected to represent the level of readiness of the subject, which we assume to be inversely related to its reaction time. Our model thus describes some of the processing steps involved in the calculation of discrete or continuous probabilities.

Following [REFS] and others, we assume that those probabilities are manipulated at a the level of one or several neuronal population(s) (the mapping of those different processing steps with existing brain structures if left to the final discussion).

If the RS is the event the subject is waiting for, we note  $p(\text{RS})$  the probability of the RS to take place at any time within a trial, in the absence of any internal or external cue. If we suppose now that a certain variable  $x$  represents the current *state* of the neural network, then  $p(\text{RS}|x)$  is a conditional probability representing the likelihood of the event to take place *given* the current state. The knowledge of  $x$  is thus expected to refine the initial probability, so that different values of  $x$  may correspond to different probabilities, making the final event more or less probable : if, for a certain  $x$ ,  $p(\text{RS}|x)$  is close to one, the subject is “almost sure” that the event will take place now, anticipating its motor response and shortening the RT. On the contrary, if for another  $x' \neq x$ ,  $p(\text{RS}|x')$  is close to zero, the subject finds very unlikely the event to take place now and the RT will be longer.

Of course, in the case of temporal tasks, we need to consider how the evolution of the internal state between the warning stimulus and the response stimulus modifies the expectations of the subject. We thus note  $x_{1:n} = (x_1, \dots, x_n)$  a sample of the different values taken by  $x$  for different instants of a particular trial, where  $x_1$  is the “initial” state (at the time of the WS), and  $x_n$  is the “final” state (at the time of the RS). The subscript  $i$  represents the time index, and  $x_i$  is the state at time step  $i$ . Then the series of  $p(\text{RS}|x_i)$ ’s indicates the evolution of the temporal expectations during the trial.

Between the WS and the RS, no feedback is given to the subject, and a complex ensemble of internal processes is needed in order to form correct expectations. The way this probability is formed in the case of temporal interval estimation tasks in the absence of sensory feedback is the subject of this paper. The problem is to define how this  $x$  and associated conditional probability evolves with time in the case of classical time interval experiments implying (or not) endogeneous cues.

We describe in the following the necessary processing stages implied in the calculation of appropriate temporal expectations.

## 2 First processing stage : Stochastic and probabilistic models of time uncertainty

The absence of feedback means that an internal state (or variable) needs to be generated through a certain neuronal process in order to represent the time passing during the trial.

The existence of internal variables explicitely modelling the passage of time is a subject of controversy in the timing litterature (Simen, Balci, deSouza, Cohen & Holmes, 2011) and is not central to the question we consider here.

For simplicity reasons, we assume that an internal “temporal variable”  $y$  may represent the passage of time (since the WS) in a manner that is compatible with the experimental data observed so far. It needs to respect some constraints, in

particular to be monotonically increasing for the duration of a trial.

## 2.1 Internal variability in time estimation

The estimation of time intervals is characterized by strong inter-trial variability. When a subject is asked to reproduce a particular temporal interval, typical responses display a coefficient of variation (CV) of the order of 0.1 in humans whatever the length of the interval. This linear scaling of the variability to the length of the interval is referred as the *scalar property* (Gibbon, 1977).

The strong variability in the subject's responses suggests a corresponding internal uncertainty about effective external time passing in the absence of sensory feedback. Simply speaking, some noise must be introduced in the process that generates temporal expectations, where the noise is a component of the process that generates the internal estimation. This means considering a *stochastic process*.

In modelling studies, it is generally considered that an internal temporal variable is updated continuously during the trial, thanks to the two following key component : A first population of neurons, called the “pacemaker”, displays a phasic activity at the onset of the trial at a frequency which remains constant during the whole trial. This activity may be modelled by a ticker [REF], a noisy periodic oscillator [REF] or simply a Poisson process [REF]. A second population of neurons, called the “accumulator”, collects and integrates the ticks of the first population, providing an activity that regularly increases during the trial.

This accumulator is expected to display a noisy ramping activity as observed, as for instance, in LIP neurons (Gold & Shadlen, 2001), in the case of a decision task where evidence about the final direction needs to be accumulated over time, until a threshold is reached. In models of temporal estimation, the frequency of the pacemaker (in combination with the gain of the accumulator) determines the length of the estimated temporal interval, and the subject's variability is modelled by the level of noise in the pacemaker.

Here we consider for simplicity the classical “noisy ticker” accumulation scheme, modelled as a discretized drifting diffusion process as proposed, e.g., by (Simen m. fl., 2011). Starting from 0, the temporal variable is updated at each time step as

$$y_{i+1} = y_i + \alpha + \beta \xi_i$$

where  $\alpha$  is the gain of the accumulator,  $\beta$  is the diffusion, and  $\xi_i$  is the result of a random draw according to the Normal law  $\mathcal{N}(0,1)$ . This “build-up” activity generates the temporal signal  $y_{1:n}$  at the origin of the foreperiod estimation.

## 2.2 Learning readiness to RS from observation

Consider now the different values taken by  $y$  at the end of each trial. Due to the accumulation of noisy activity, the final value of  $y$  is indeed different at each trial even in the case of a constant foreperiod <sup>1</sup>.

---

<sup>1</sup>in the case of a drifting diffusion process, it obeys to a Gaussian distribution of average  $n\alpha$  and standard deviation  $\beta\sqrt{n}$ , where  $n$  represents the length of the temporal interval. With a CV equal to  $\frac{\beta}{\alpha\sqrt{n}}$ , the internal estimate does not display the linear scaling observed in most temporal reproduction tasks. In order to display the scalar property, one possibility is to modulate the value of the gain  $\alpha$  as a function of a current expected interval stored in memory, as proposed by (Gibbon, 1991). Another possibility is to use adaptive  $\beta_i$  during trial, i.e.  $\beta_i = \beta\sqrt{\frac{|y_{i-1}|}{\alpha}}$  in which case the final variance is of the

The uncertainty about the final duration is then expressed by the distribution of the values of  $y$  at the time of the RS. The accumulation of a Poisson process, for instance, would result in a distribution of intervals that obeys to a Gamma distribution. In the case of the drifting diffusion approximation, it is given by a Gaussian distribution.

We use the notation  $f(y|RS)$  to describe the distribution of the  $y$ 's obtained at the end of the trial, summarized<sup>2</sup> by its mean  $\mu$  and standard deviation  $\sigma$ . If we note  $\theta = (\mu, \sigma)$  the couple, we use  $f(y|\theta)$  as the estimate of  $f(y|RS)$  given observations.

The RT of the subject is expected to reflect the value of  $p(RS|y_i)$  that represents its readiness to the upcoming response stimulus. If we consider the relation given by the Bayes theorem:

$$p(RS|y_i) = \frac{p(RS)}{f(y)} f(y|RS)$$

where  $p(RS) = \frac{1}{n}$  and  $f(y)$  is the distribution of  $y$ 's whatever the RS is present or not, and if we accept to approximate  $f(y)$  with a constant  $f(y) \simeq \frac{1}{y_{max}}$ , (which is approximately true in the case of a linearly increasing build-up activity), we can state that :

$$p(RS|y_i) \propto f(y|RS) \simeq f(y|\theta) \quad (1)$$

and consider the quantity  $f(y_i|\theta)$  as representative of the readiness of the subject.

The key point here is to notice that the distribution  $f$  is obtained by experience, i.e is progressively shaped through a parameter update at each new observation. It reflects the current knowledge of the subject regarding the distribution of the external temporal events.

The parameter update represents the learning process which is expected to take place at the synapses of the neuronal population receiving inputs from the accumulator. At the time of the RS, a synaptic plasticity mechanism is expected to occur, facilitating the response to the current subset of active neurons representing  $y$  in the accumulator, and depressing the other ones. The update mechanism is expected to display responses curves that faithfully reflect the distribution of values produced by the internal accumulator at the end of the trials.

In this model, we do not postulate any adaptation or change in firing rate at the level of the pacemaker or the accumulator. The adaptation processes we consider mostly take place at the output of the accumulator, that we refer as the “predictor”.

## 2.3 Variable foreperiod

An important factor influencing RTs is the variability of the time at which the stimulus is presented – i.e. the temporal predictability of the stimulus. Several studies have shown that increasing the temporal variability of the stimulus increased the reaction time of the subject (Klemmer, 1957; Karlin, 1959; Drazin, 1961) – i.e. the subject reacted faster to a stimulus when he knew the time at which the stimulus was going to occur with greater certainty.

From a modelling point of view, this additional variability is expected to interfere with the self-generated internal variability; at the level of the predictor, the RS is

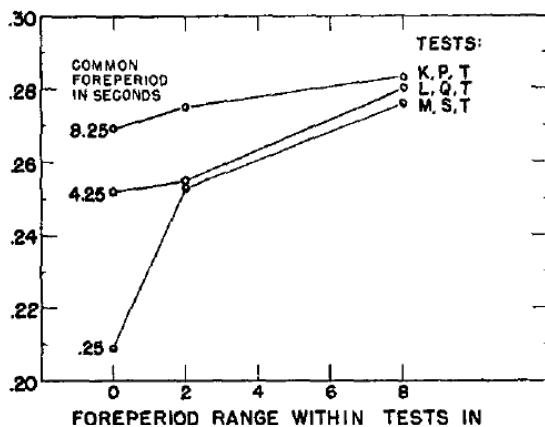
---

order of  $\beta\sqrt{\frac{n(n+1)}{2}}$  and the CV remains approximately constant, of the order of  $\sqrt{\frac{\beta}{2\alpha}}$  for large  $n$ .

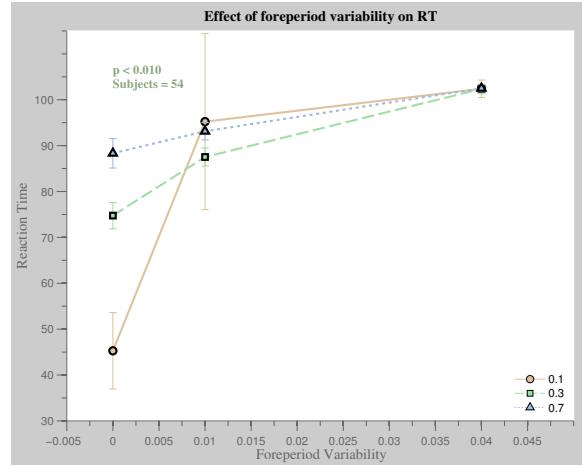
<sup>2</sup>In the case of a Gaussian estimate, the values of  $\mu$  and  $\sigma$  can be estimated after observing the final value of  $y$  in  $m$  different trials, and using non-biased estimators  $\mu = \frac{1}{m} \sum_{j=1}^m y_n^j$  and  $\sigma^2 = \frac{1}{m(m-1)} \sum_{j=1}^m (\tilde{\mu} - y_n^{(j)})^2$ .

expected to take place within a larger range of  $\gamma$  values (as internal and external variance are expected to add), leading to a higher uncertainty and a lower readiness to the RS.

Klemmer (1956) observed that the RT of participants in an experiment changed with a change in the variability of the foreperiod. These results are reproduced in Figure 1. Each line corresponds to a different mean foreperiod. It can be seen that as the variability in the foreperiod effect increases (along the x-axis), the reaction time increases (along the y-axis).



(a) Experimental results



(b) Simulation

Figure 1: (a) Effect of varying the foreperiod on reaction time for three different mean foreperiods. Reproduced from Klemmer (1956). (b) Reaction times obtained by training the model for three different foreperiods (0.1s, 0.3s and 0.7s) and three different stimulus variabilities.

This condition can be simulated by increasing the variability in the number of observations  $u_{1:n}$  before the response stimulus. In order to test the model, it was conditioned through a sequence of 160 trials on input stimuli that could be one of three different foreperiods – short (0.1s), medium (0.3s) or long (0.7s). A Gaussian noise was added to the input stimuli, which could acquire one of three variabilities: 0, 0.01s or 0.04s. Six subjects (instances of the simulation) were trained under each of these nine conditions, making a total of 54 subjects. After the training trials, a testing trial recorded the expectation of the subject at the mean foreperiod that it was trained on. Reaction times are then estimated simply as the inverse of this expectation, computed as  $C - e(y_i)$ , where  $C$  is a large constant. These results are graphically summarised in figure 1 where each point is the mean RT of a category and the vertical bars show the confidence interval based on the RTs of the six subjects that belong to that category.

The results from simulation reproduce two key aspects of the original observations by Klemmer: (a) reaction time increase with increase in the variability of the stimulus – as we move from left to right along the x-axis, the reaction time for all foreperiods tends to increase, (b) the increase in reaction time is much lesser for long foreperiod than for short foreperiods. Thus, the change in RT is much greater in case of the mean foreperiod of 0.1s than that for the mean foreperiod 0.7s.

The picture becomes however more complicated when the foreperiod variability is significantly higher than the internal variability. Most studies measure the reaction times to categorical foreperiods – where the length of the foreperiod is picked at random from a limited set of foreperiods ranging from one second to more than ten seconds. Then a key RT phenomenon, the *variable foreperiod effect*,

can be observed: the reaction time is *longest for the shortest foreperiod* (i.e. delay between WS and RS) and then decreases with saturation. This phenomenon was first recorded by Woodrow (Woodrow, 1914) and has been replicated a number of times since (see Niemi & Näätänen, 1981).

In order to replicate the foreperiod effect, we need first to take into consideration the effect of categorical foreperiod presentation.

### 3 Second processing stage : mixture models for categorical foreperiod experiments

The environment sometimes provides cues about the duration of a sensory signal. A jogger is likely to reach from A to B faster than a saunterer. The respective actions for jogging and sauntering are underlying *sources* of the sensory signal that are responsible for short and long durations. However, sources for sensory signal are not always available to observation. When waiting for a friend who is running late, the environment might provide limited clues about the duration of waiting. In order to predict the most likely time of arrival, the brain might need to speculate the underlying source of the delay. A detour to a petrol station would probably lead to a short delay while a minor accident will cause a relatively longer delay. While the source of delay might actually be hidden, subjects can better predict the likely duration of a sensory signal by attributing a *cause* to it.

Our main claim is thus that when faced to a variable foreperiod task, the subject will manipulate several concurrent hypotheses, the plausibility of which being possibly revised during the trial.

#### 3.1 Mixture model

If we suppose the knowledge about the sources of sensory data is internalized as a mixture model. This means that the predictor can model the sensory data using a linear superposition of probability densities. In the cases where the range of values associated with each source are well separated, the global density appear as composed of several density peaks, each peak corresponding to a distinct source.

Consider now the temporal variable  $y$  as observed at the end of a trial, a discrete random variable  $k \in \{1, \dots, K\}$ , and their joint probability distribution  $P(y, k)$ . The variable  $y$  is the observation and  $k$  is the source (the cause) generating this observation. A *mixture model* describes how the different observations originate from the different sources out of  $1, \dots, K$ . It is composed of a series of conditional densities  $f(\cdot|k)$ 's, giving the distribution of  $y$  when the source  $k$  is active, and a series of discrete probabilities  $\pi_k = p(k)$  giving how likely is the source  $k$  in the observed data (called the *mixing proportion*). Then, the readiness attained after observing  $y$  is given by:

$$f(y) = \sum_{k=1}^K f(y|k)\pi_k \quad (2)$$

The important quantities here are the mixing proportions  $\pi_k$ 's. They represent the credence attributed to the different sources. By construction, the sum  $\sum_{k=1}^K \pi_k$  is equal to 1. Provided the normalization constraint is respected, the mixing proportion may be modified, either within trial or between trial as new information is considered.

### 3.2 Guessing the sources from a series of observations

The duration of sensory signals is a property that could be predicted by recording a series of observations, attributing a source to each observation, and creating an probabilistic model for how this series of observations was generated *given* the source.

– Reference here about learning foreperiod which has recently been suggested to involve conditioning (Los & van den Heuvel, 2001).

In order to learn the mixture model, we need to consider a larger set of parameters than in the single foreperiod case :

- We suppose the densities  $f(y|k)$ 's are described by a set of parameters noted  $\theta_k$ . We note  $f(y|\theta_k) = f(y|k)$ .
- For simplicity, we moreover suppose that the densities are Gaussian distributions, i.e.  $\theta_k = (\mu_k, \sigma_k)$  where  $\mu$  represents the mean and  $\sigma$  represents the standard deviation.

We note  $\theta_k$  the set of parameters associated with source  $k$  and  $\theta = (\theta_1, \dots, \theta_k)$  the global set of parameters.

Since the sources associated to the different sensory signals are *unobserved*, the subject needs to model the generative process based on *incomplete* or partially observed data. The sources need to be guessed from the available observations. The parameters of this model need to be *learned* by performing something equivalent to a maximum likelihood estimate (MLE), on the basis of an incomplete data set composed of a series of  $m$  observations :  $\{y^1, \dots, y^m\}$ . This problem is complex and could be computationally intractable (McLachlan & Krishnan, 1997).

Modelling incomplete data is fortunately a well-studied problem and a number of solutions have been suggested to get around the problem of unknown values for hidden variables. One such solution is through the well-known *expectation-maximization* (EM) algorithm (Dempster, Laird & Rubin, 1977; McLachlan & Krishnan, 1997). This solution breaks down the problem of estimating sensory signals into two steps: (a) use the Baye's rule to *infer* the likely causes underlying the set of observations, based on existing knowledge of the generative model, and (b) use the inferred causes to *learn* the parameters of the predictor so that it can better estimate the likely values for sensory signals. These two steps are then repeated on the data set to improve the inference of causes and the likelihood of the observed sensory signals<sup>3</sup>.

<sup>3</sup>

The goal of learning is to obtain an optimal set of parameters which are able to make a “good” prediction. This means optimizing the likelihood of the data set given  $\theta$ . The EM algorithm optimizes the likelihood iteratively. Each iteration is called an epoch, and composed of two steps : the Estimation (E) step and the Maximization (M) step.

- Estimation : For a given set of parameters  $\theta$ , first identify the responsibilities  $t_k(y^j)$ 's as:

$$\forall j \in \{1, \dots, m\}, \forall k \in \{1, \dots, K\}, t_k(y^j) = \frac{f(y^j|\theta_k)\pi_k}{\sum_{\ell=1}^K f(y^j|\theta_\ell)\pi_\ell} \quad (3)$$

where  $\pi_k = p(k)$  is the mixing proportion.

- Maximization : update the parameters as:  $\forall k \in \{1, \dots, K\}$ :

$$m_k = \sum_{j=1}^m t_k(y^j) \quad (4)$$

$$\pi_k = \frac{m_k}{m} \quad (5)$$

With the Bayes' rule comes a quantity corresponding to the likelihood of source  $k$  generating observation  $y$ , called the *responsibility*, i.e.

$$t_k(y) = p(k|y) = \frac{f(y|k)\pi_k}{f(y)} \quad (8)$$

The responsibility  $t_k(y)$  is a discrete probability over  $k$  i.e.  $\sum_{k=1}^K t_k(y) = 1$ . It is used in general to identify the most likely source of the current signal.

If we suppose that the parameters are updated after each observation, the quantity  $t_k(y)$  indicates how likely is the source  $k$  to generate the current observation. It is representative of the competitive process between the different sources, and the update of the parameters is conditioned by its value. From a functional point of view, this competitive process (E-step) is expected to take place at the level of the predictor where different neuronal sub-populations are expected to compete for representing the current temporal feature, in a way similar that it has been observed, e.g., in the visual system(?, ?). Then, the update of the parameters (M-step) should be implemented in a way that is very similar to the synaptic mechanism we described earlier, facilitating *in the selected subpopulation of the predictor* the response to the current subset of active neurons representing  $y$  in the accumulator.

It must be noticed however that in order to reach the stationary distribution, the EM process needs to be repeated several times over the same data set. This is unrealistic from the neuronal point of view and the EM approach is here considered as a phenomenological model of the combination of a competitive process with an adaptive process, as it is expected to take place when the brain is learning categorical foreperiods.

One can also look at [REF:Wolpert] for instance for an implementation of EM in a model of motor decision in the case of ambiguous perception. A possible implementation of an EM process in a network using an STDP synaptic plasticity rule has been recently proposed by [REF:Maas].

### 3.3 Evidence of a decrease of uncertainty with time passing

We now come back in more detail to the question of the so-called “foreperiod effect”. Recall here it corresponds to a *decrease* of the reaction time to longer foreperiods in contradiction with the initial guess of longer RTs for longer foreperiods.

Why does the foreperiod effect arise? During the foreperiod the subject prepares for the RS. A short RT indicates a larger level of readiness or expectation compared to a longer RT. Therefore the foreperiod effect indicates that the expectation of the subject increases with the length of the foreperiod<sup>4</sup>.

Thus, one could explain the foreperiod effect if the uncertainty associated with the foreperiod decreases with the increase in the length of the foreperiod. Elithorn och Lawrence (1955) argued that this would indeed be the case if the subject calculated the expectation of a stimulus by taking into account how much time had already passed. They argued that rather than simply calculating the probability of

---


$$\mu_k = \frac{1}{m_k} \sum_{j=1}^m t_k(y^j) y^j \quad (6)$$

$$\sigma_k^2 = \frac{1}{m_k} \sum_{j=1}^m t_k(y^j) (y^j - \mu_k)^2 \quad (7)$$

<sup>4</sup>However, it is not the absolute length that is important as the foreperiod effect is robust to several different time ranges, as long as the ratio of the range of the foreperiods to the mean foreperiod is larger than a certain threshold (Drazin, 1961; Elliott, 1973).

the occurrence of a stimulus based on its temporal frequency, a subject calculates at each point of time the conditional probability of the occurrence of a stimulus (based on its temporal frequency) *given that* the stimulus has not occurred till the current point in time. Thus, this account argues that subjects compute expectation for a stimulus based not only the statistical properties of the environment (the temporal frequency), but also on a *strategic* process, using the information about the flow of time and integrating it with their knowledge of environmental statistics.

### 3.4 The “hazard function” approach

The *hazard function*, commonly known in System Reliability Theory, is a measure of the rate at which a system, functioning at time  $t$ , is likely to fail soon after this time (?, ?).

The likelihood function,  $f(y)$ , gives a probability density function of  $y$  being the final duration of the trial on the (subjective) time axis, the  $y$ -axis. Computed in this manner, the expectation does not use the information about how much time has passed so far. The probability of the RS to take place now, given that the final time is going to be greater than the current observation  $y_i$ , can be expressed as the conditional probability:  $p(\text{RS}|y, y \geq y_i) \propto f(y_i|y \geq y_i)$ , which corresponds to a recasting of the initial distribution on smaller interval. Using the Baye’s rule, we can indeed state that :

$$f(y_i|y \geq y_i) = \frac{\mathbf{1}_{y \geq y_i} f(y_i)}{p(y \geq y_i)} = \frac{\mathbf{1}_{y \geq y_i} f(y_i)}{1 - \int_0^{y_i} f(u) du} \quad (9)$$

Computed in this manner, the predictor likens the appearance of the RS to the failure of a system and continuously uses the information given by the passage of time to calculate the expectation of this failure. Note that even if the likelihood  $f(y_i)$  remains constant, the conditional probability  $f(y_i|y \geq y_i)$  will change because the integral in the denominator of the function will change.

This approach is appealing at first sight as it is supported by some data. In particular, support for such a hazard-based computation of expectation comes from studies that show that changing the distribution for different foreperiods changes the relation between foreperiod and reaction time (hereafter, FP-RT curve). Traditional foreperiod experiments (e.g. Woodrow, 1914) used a rectangular distribution of foreperiods, i.e. all foreperiods were equally likely and the conditional probability of observing a stimuli at a given point of time, steadily increased. If, instead, an experiment was to use a different distribution of foreperiods, it would change this conditional probability. Consequently, if the FP-RT curve relies on computation of the conditional probability, then changing the distribution of foreperiod should change the FP-RT curve. Baumeister och Joubert (1969) observed that this was indeed the case. They used four different foreperiod distributions and observed that the FP-RT curve changed in agreement with the strategic hypothesis of temporal expectation. Näätänen (1971) took this experiment to the extreme and devised a ‘non-aging’ distribution of foreperiods so that the shortest foreperiods were most likely and longest were least likely. He observed that for such non-aging distributions the foreperiod effect vanishes, i.e. the FP-RT curve is flat, which would be the case if the subject was using the conditional probability to calculate the expectation. Recently, Trillenberg, Verleger, Wascher, Wauschkuhn och Wessel (2000) replicated the result of Näätänen (1971) for a non-aging population, using a set of categorical foreperiods (compared to the continuous foreperiods used in the earlier

study).

It must be noticed that a brute calculation of the Hazard function, as given by (9), is diverging for large values of  $y$  as the denominator tends toward zero when  $y$  tends toward  $\infty$  (except in the case of the very specific “non aging” distribution where the probability of the hazard event decreases exponentially fast as the time increases). In order to avoid an unrealistic drift of the value of the expectation, we introduce a parameter  $\beta \in ]0, 1[$  that represents a general level of uncertainty regarding the prediction given by  $f$ , so that there is a certain level of chance, namely  $1 - \beta$ , that the RS may arrive at any time. It gives :

$$f(y_i|y \geq y_i) = \frac{\mathbf{1}_{y \geq y_i} f(y_i)}{1 - \beta \int_0^{y_i} f(u) du}$$

Baumeister och Joubert (1969) tested subjects for four different stimulus distributions and found that reaction times do indeed reflect the conditional probability of the stimulus at a point of time. Figure ?? shows the reaction times obtained in their study for four different distributions – rectangular (equal probability for each RS), symmetrical (low probability of RS as 0.1s and 0.8s, while high probability at 0.2s and 0.4s), skewed-left (probability of RS increases from left to right) and skewed-right (probability of RS decreases from left to right).

In the next set of simulations, we computed the expectation that would result from a hazard function. We did this for each stimulus distribution simulated above (rectangular, symmetrical, skewed left and skewed right) and the results are shown in figure 2.

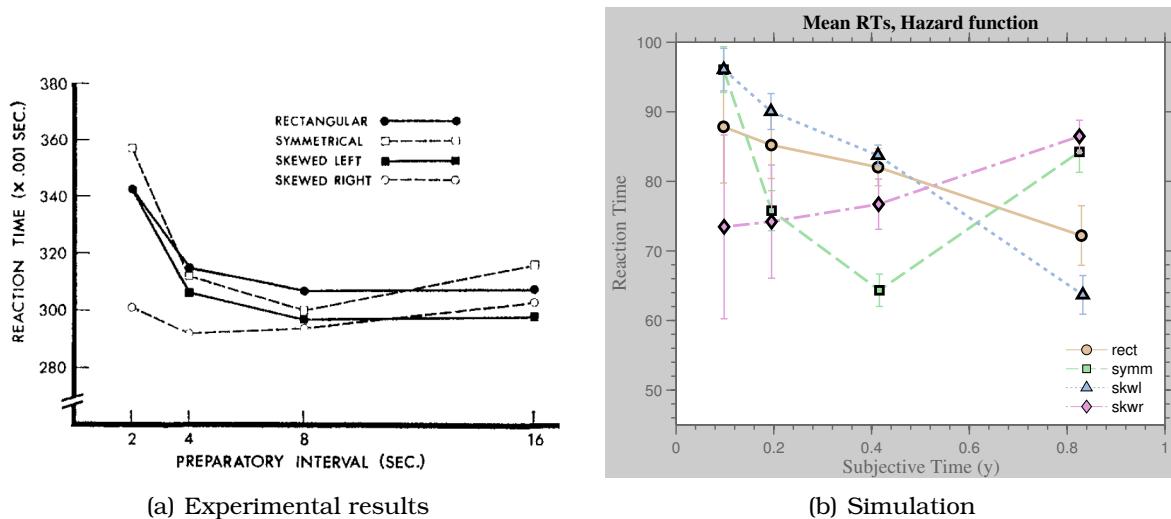


Figure 2: (a) Reaction times for stimuli with different distributions. Reproduced from Baumeister och Joubert (1969). (b) Reaction times for different distributions using the principle of hazard function, with parameter  $\beta = 0.8$  (see text).

Some of the expected effects of an internal calculation of a Hazard function are reproduced in the simulation. For instance, in the “skewed-right”, which approximately corresponds to the non-aging case, we observe a slow increase of the reaction time that is expected as an effect of uncertainty due to the internal variability. One can also remark that in the rectangular case, our mixture model displays a monotonical decrease of the reaction time, which is the main effect expected from the Foreperiod effect. This effect logically is enhanced in the skewed-left case, and the final increase of RT in the case of the symmetrical distribution is also reproduced.

So, while the results from simulating the hazard function show an overall foreperiod effect (except, of course, for the skewed-right case), there are two key differences that can be observed between the changing inference mechanism proposed in this study and a hazard function. Firstly, the sharp drop and saturation that is usually observed in foreperiod effect studies is not reproduced in this case. Secondly, these results do not reproduce the roughly equal RTs for all stimulus distributions at the largest interval that was observed by Baumeister och Joubert (1969).

The reason for this difference can be understood by looking at the characteristics of the hazard function... (TO BE ANALYSED)

TODO : NEURONAL IMPLEMENTATION?

## 4 Third processing stage : adapting the mixing proportions

### 4.1 Adaptive mixing proportions

The initial motivation of a mixture model of temporal expectation was the possibility to differentially (and independently) manipulate the expectations attached to the different underlying hypotheses.

It is worth noting that the mixing proportions  $\pi_k$ 's are both expected to adapt both between and *during* trials.

In order to allow this dynamic update, they must be considered as a separate signal arriving to the predictor (see figure 3). The level of readiness is then measured at every instant within the trial as the sum over every source of the product between a density function, parametrized by  $\theta_k$  and a mixing proportion noted  $\pi_i^k$  possibly changing as time passes, i.e :

$$f(y_i) = \sum_{k=1}^K f(y_i|k)\pi_i^k \quad (10)$$

This modelling choice rises the question of the neurological implementation of a bayesian choice, i.e. calculating a *posterior* probability on the basis of a product between a likelihood function ( $f$ ) and a prior ( $\pi$ ). This question is currently under debate in the modelling community and we refer to the work of [REF:Pouget, Deneve, Fiser] do decipher how this product may be implemented when the probabilities are encoded at the level of large neuronal populations.

### 4.2 Intra-trial variability : modeling the foreperiod effect with adaptive mixing proportions

#### 4.2.1 Model

Consider now that the priors attached to the different sources may vary with  $i$ , i.e the *prior* attached to the source  $k$  is *dependent* on the passage of time. We note  $f(y|k) = f(y|\theta_k)$  the distribution of the final  $y$ 's, i.e. the  $y$ 's measured at the end of the sequence, previously obtained through the EM process. Then the likelihood of the source  $k$  during the observation of  $y_i$  is the chance of the final  $y$  to be greater than  $y_i$ , i.e.

$$p(y \geq y_i|k) = 1 - \int_0^{y_i} f(y|k) dy \quad (11)$$

so that the prior attached to the source  $k$  given that  $y \geq y_i$  is:

$$\pi_k(y_i) = p(k|y \geq y_i) = \frac{p(y \geq y_i|k)\pi_k}{\sum_{\ell=1}^K p(y \geq y_i|\ell)\pi_\ell} \quad (12)$$

The expectation attached to the observation  $y_i$ , as defined by eq.10 now becomes:

$$f(y_i) = \sum_{k=1}^K f(y_i|\theta_k)\pi_k(y_i) \quad (13)$$

where  $\pi_k(y_i)$  governs the on-line revision of the gaussian mixture.

More in line with a psychological point of view, we consider in the following that the update of the current prediction should rely on discrete steps that reflect the change in the *belief* about the possible sources causing the observation given the current time estimate. There is evidence from behavioural studies (Coull, 2009) that subjects use the informative value of the unidirectional nature of time in order to predict the duration of a sensory signal. As the trial proceeds and time goes forward, the subject will be in a position to rule out certain sources of the sensory signal, knowing that the duration of the signal is beyond a particular value.

The unidirectional nature of time means that the temporal sources underlying sensory signals form an ordered sequence for when a source can be ruled out. This knowledge of the unidirectional nature of time is what is called the “strategic” account of time in the calculation of the likelihood, as proposed by XXX and XXX.

**– Here you can add detailed reference to data or to lit review section –** Sources that are associated with smaller durations can be ruled out early, so occur early in the sequence. In contrast, sources that are associated with larger durations cannot be ruled out for a longer period of time and therefore occur later in this sequence.

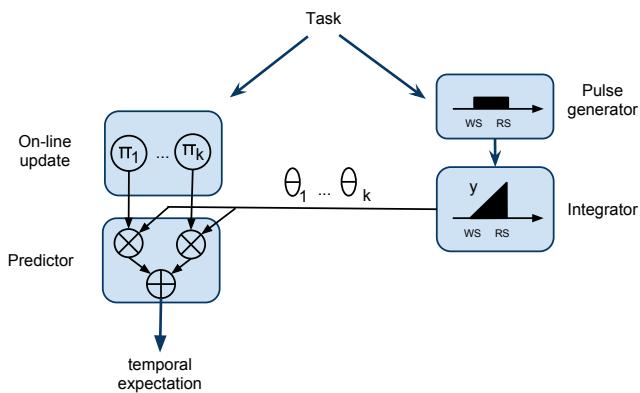


Figure 3: Schematic representation of the double information path leading to the formation of an adaptive mixture, with, on one side, the adaptive calculation of the mixing proportion, and on the other side, the calculation of the likelihood of the different sources at generating  $y$  (where  $y$  is an internal signal representing the time since the WS) – See text for more detail –.

Given such an ordered set, the predictor can choose which inferences it can rule out. Let us consider this mechanism of ruling out sources more formally.

So far we have assumed that the discrete variable  $k \in \{1, \dots, K\}$  represents one of the sources that underly the observed signal  $y_i$  at instant  $i$ . Now, we assume that these  $K$  sources form an ordered sequence on the basis of the mean of the temporal variable related to each source, i.e.

$$\mu_y^1 \leq \mu_y^2 \leq \dots \leq \mu_y^K \quad (14)$$

where  $\mu_y^k$  is the mean of the Gaussian source  $k$  along the dimension  $y$ .

The model can keep track of the possible sources at a particular instant of time by maintaining an index  $s$  into the sequence. At a particular instant, all sources beyond this index form a set of *valid sources*. All sources below the index are invalid and are presumed to not underlie the observed signal. During the trial, the index  $s$  is updated, changing the list of valid sources. But how decide to move from one index to the next?

How much time needs to pass before the subject decides to rule out a source as invalid? We can answer this question by considering each subset of valid sources as the subject's *belief state* and changing the index as a *transition* from one belief state to another. At the beginning of a trial, the subject might believe that the observed stimuli can arise out of any of the hidden sources. Additionally, the subject uses the observed stimuli to *infer* whether it should remain in the given belief state, or make a transition to the next one, believing that the observed signal does not arise from the first (shortest) temporal source. This process of dynamically revising inferences can be seen as a state-space model shown in Figure ??.

Let us assume that the belief states of the subject are represented by a variable  $s$ , so that  $s_1, \dots, s_K$  are  $K$  belief states with the belief state  $s_k$  corresponding to the set of valid sources  $\{k, \dots, K\}$  (see Figure ??). As each belief state assumes a different mixture of sources underlying the observed signal, the likelihood of observing a signal  $y_i$  at instant  $i$ , given in Equation ??, now becomes a function of the belief state at that instant. Let us assume that a variable  $q_i$  records the belief state at instant  $i$ , then

$$f(y_i|q_i) = \sum_{k=1}^K f(y_i|k) \pi_k(q_i) \quad (15)$$

gives the likelihood of the observed signal  $y_i$  as a function of belief state given by  $q_i$ .

The mixing proportions,  $\pi_k(q_i) = P(k|q_i)$ , now become a function of this belief state. Equation 15, then gives us a way to compute the likelihood dynamically, updating it as the belief states update during the course of a trial. However, in order to use it, the subject must be able to (a) infer the belief states with the passage of time, i.e. assign a state from the set  $\{s_1, \dots, s_K\}$  to  $q_i$ , and (b) compute the mixing proportions  $\pi_k$ 's as a function of this belief state  $q$ . We look at how the subject can solve each of these problems next.

The first problem, that of defining the valid sources for every  $q$ , is the problem of computing the mixing proportions,  $\pi_k$  as a function of this belief state. Recall that, from definition,  $\pi_k(q) = P(k|q)$ . In a state  $s_k$ , all the sources  $\ell$ , with  $\ell < k$  are considered invalid. Therefore, we have the following two constraints on calculating

the mixing proportions for the belief state  $q_i = s_k$ :

$$\begin{aligned}\pi_\ell(s_k) &= 0 \quad \text{if} \quad \ell < k \\ \sum_{\ell=k}^K \pi_\ell(s_k) &= 1\end{aligned}\tag{16}$$

This means that when the predictor sets the mixing proportion for one source to be zero, it must redistribute the prior probabilities for picking other sources. There are three choices here: (a) *uniform redistribution*: the predictor distributes the mixing proportions equally over the remaining sources, (b) *fully optimistic*: the predictor assumes that the signal arises from the next source in the sequence, or (c) *optimistically biased*: the predictor passes the mixing proportion of the invalidated source to the next one:

- Uniform redistribution:

$$\pi_\ell(s_k) = \begin{cases} 0 & \text{if } \ell < k \\ \frac{\pi_\ell}{\sum_{j=k}^K \pi_j} & \text{if } \ell \geq k \end{cases}\tag{17}$$

- Fully optimistic:

$$\pi_\ell(s_k) = \begin{cases} 0 & \text{if } \ell < k \\ 1 & \text{if } \ell = k \\ 0 & \text{if } \ell > k \end{cases}\tag{18}$$

- Optimistically biased:

$$\pi_\ell(s_k) = \begin{cases} 0 & \text{if } \ell < k \\ \sum_{j=1}^k \pi_j & \text{if } \ell = k \\ \pi_\ell & \text{if } \ell > k \end{cases}\tag{19}$$

where the  $\pi_k$ 's are the values of the mixing proportions at the start of the trial. Which rule an subject employs for recalculating the priors for each source will depend on the strategy that the subject is employing. We will come back to this question in sections ?? and ??.

We now come to the second problem, that is of inferring the hidden state of the HMM. Finding the belief state at each instant, is the problem of statistical inference of the states for a hidden Markov chain. At an instant  $i$ , the subject has made the observations  $y_1, \dots, y_i$  and, based on these observations, wants to infer the belief state  $q_i$  from the set  $\{s_1, \dots, s_K\}$ . In other words, given the series of observations,  $y_{1:i} = (y_1, \dots, y_i)$  the subject needs to estimate the probabilities  $p(q|y_{1:i})$ .

We assume that the subject has learnt the set of parameters for the HMM  $\Phi = (\boldsymbol{\theta}, \mathbf{A})$ , where  $\boldsymbol{\theta}$  are the model parameters, and  $\mathbf{A} = (a_{lk})_{l \in 1 \dots K, k \in 1 \dots K}$  is the transition matrix for the HMM and gives the *transition probabilities*,  $a_{lk}$ , of switching from state  $s_l$  to state  $s_k$ . This problem, of propagating the conditional probability of a system's states based on a sequence of observations, is the *filtering problem* from control theory. The problem can be solved through the discrete analog of the Kalman filter (Kalman, 1960) which solves the problem by estimating the desired conditional probabilities recursively at each instant of time from the estimate at

the previous instant. Here, we assume that the subject keeps track of only the most probable (MAP) belief state, instead of retaining the conditional probability of each belief state; i.e. the subject only remembers the most likely belief state at the previous instant,  $\hat{q}_{i-1}$ , rather than the probability  $P(q_{i-1} = s_l | y_{1:i-1})$ :

$$\begin{aligned}\hat{q}_i &= \arg \max_{k \in \{1, \dots, K\}} P(s_k | y_i, \hat{q}_{i-1}) \\ &= \arg \max_{k \in \{1, \dots, K\}} f(y_i | s_k) P(s_k | \hat{q}_{i-1})\end{aligned}\tag{20}$$

This recursive calculation of probabilities of belief states can be simplified by making another plausible assumption. We assume, as illustrated in Figure ??, that the stochastic process is a *left-to-right* Markov chain of order one, i.e. all elements of  $\mathbf{A}$  except the upper bidiagonal are zero. The subject invalidates only one signal source at a time and never makes a previously invalidated source valid, thus modelling the unidirectional nature of time. This assumption means that the transition matrix  $\mathbf{A}$  is upper bidiagonal:

$$\hat{q}_i = \arg \max_{k \in \{\hat{q}_{i-1}, \hat{q}_{i-1}+1\}} f(y_i | s_k) a_{\hat{q}_{i-1} k}\tag{21}$$

This last equation tells us that the switching of belief states is entirely dependent on the likelihood of the observation in the current and next belief states and the transition probability between these states. If the transition probabilities along the diagonal of  $\mathbf{A}$  (i.e.,  $a_{kk}$ ) are smaller than the off-diagonal probabilities (i.e.,  $a_{kk+1}$ ), then the subject will show a ‘fickleness’ and switch to the next belief state even though the observation is better explained by the current belief state. On the other hand, if the off-diagonal probabilities are larger, the subject will show a reluctance to switch to the next belief state and the likelihood of observation from next belief state will have to cross a threshold before the subject invalidates a signal source.

Equations 17 / 18 / 19 and 21 together describe a mechanism through which the predictor can maintain an *active buffer* of valid inferences – a buffer which will evolve with time. The temporally evolving mixing proportions for each source, calculated through Equation 17 (or Equations 18/19), can then be plugged into Equation ?? to get the posterior probability,  $t_k(y_i) = f(k|y_i)$ , for  $k$  at instance  $i$ . Using this posterior probability, the predictor infers the likely source (or likely set of sources) that are responsible for the sensory signal. As the mixing proportions change with time, the predictor’s inference about the probability of sources underlying the sensory signal will also vary with time.

Finding the belief state,  $\hat{q}_i$ , for each instant of time solves the inference problem that arose out of Equation 15. As the trial proceeds, the predictor will recalculate the  $\pi_k$ ’s whenever its belief state changes, using one of Equations 17, 18 or 19, updating the set of valid inferences.

#### 4.2.2 simulation results

In order to test for the foreperiod effect, the simulation is again divided into conditioning and testing phases. Each subject (model instance) was first trained during a conditioning phase of 160 trials, each of which could be one of four different

durations – 0.1s, 0.2s, 0.4s and 0.8s<sup>5</sup> and each duration occurred with the *same* probability. The model then uses the EM algorithm (Equation ??) to learn the generative model that would lead to this observed data<sup>6</sup>. This generative model can then be used to predict a RS in the future. In order to test the expectation that the model will generate for each RS, we simulated the model on a trial of duration 1.0s and recorded the expectation  $e(y_i)$  generated at each point of time.

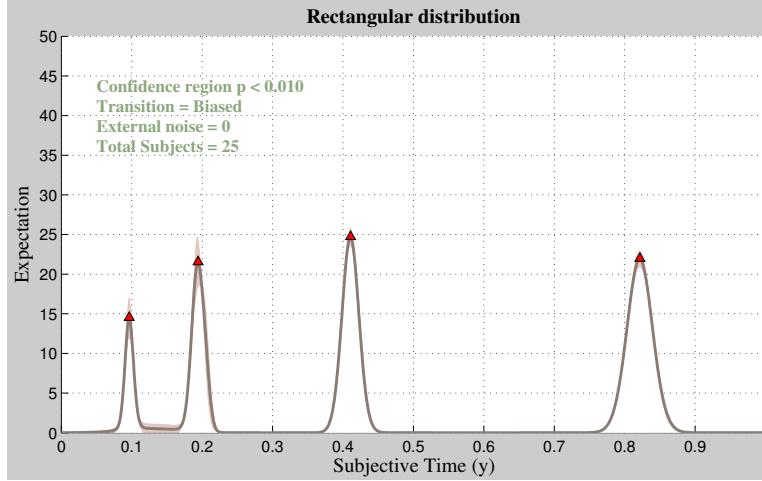


Figure 4: Expectations generated by the model at different points of time when it is trained for a sequence of foreperiods. The bold line shows mean expectation for 20 subjects while the shaded region shows the confidence region for this expectation. Peak expectations are marked by a triangle and are used to determine the reaction times.

Figure 4 shows the mean expectation curve (black line) during the testing trials for 20 subjects along with a confidence region ( $p < 0.001$ ) for this curve (shaded region). As in the previous results, the reaction time is found by calculating the inverse of peak expectation around each interval. Figure 5(a) shows typical foreperiod effect obtained in an experimental study – as the length of foreperiod increases, the reaction time decreases. As the foreperiod continues to increase, the decrease in RT shows saturation. Figure 5(b) plots this reaction time obtained from the simulation. In this simulation, the switching mechanism for the HMM has been set to *optimistically biased* transitions (Equation ??). We also tested the model for the *uniform redistribution* switching mechanism (Equation ??) and the results are shown in Figure 5(c).

The key observations from these simulations are: (a) reaction times, in general, decrease for increasing length of foreperiod – i.e. the model is able to make use of the informative value of time in order to generate expectations of the response stimulus, (b) the pattern of change in RT is closer to the observed pattern for the optimistically biased mechanism than the uniform redistribution mechanism – while the optimistically biased mechanism leads to a sharp change in RT followed by saturation, the uniform redistribution mechanism leads to a small change at short foreperiods and a sudden decrease for the last foreperiod, (c) there is large variation in RTs at shorter FPs compared to longer FPs.

<sup>5</sup>We have chosen this time scale to compare our results from behavioural studies investigating interval timing. However, our mathematical model is independent of scale and can be applied to durations both in the millisecond and tens-of-seconds range.

<sup>6</sup>In our simulations, we fixed the number of hidden sources according to the experiment (here four). But this can be learnt using a variational scheme as in ? (?).

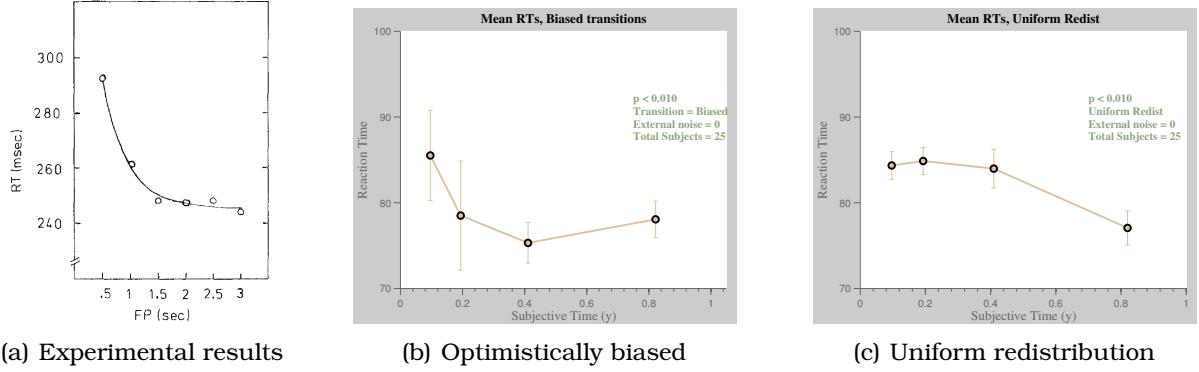


Figure 5: Reaction times as a function of foreperiod. As the length of foreperiod increases, the reaction time decreases. For the simulation, all foreperiods were equally likely.

Figure 6(a) shows results of simulating the model under each of these conditions. The simulation is performed in a similar manner to the simulation for the rectangular distribution described above – 160 conditioning trials followed by a testing trial that records the expectation. The results again accumulate data from 20 different subjects and show the mean and confidence interval for the peak RT around each RS. The switching mechanism used to simulate these results is *optimistically biased*. Again, we experimented and replaced this switching mechanism with *uniform redistribution*, the results of which are shown in Figure 6(b).

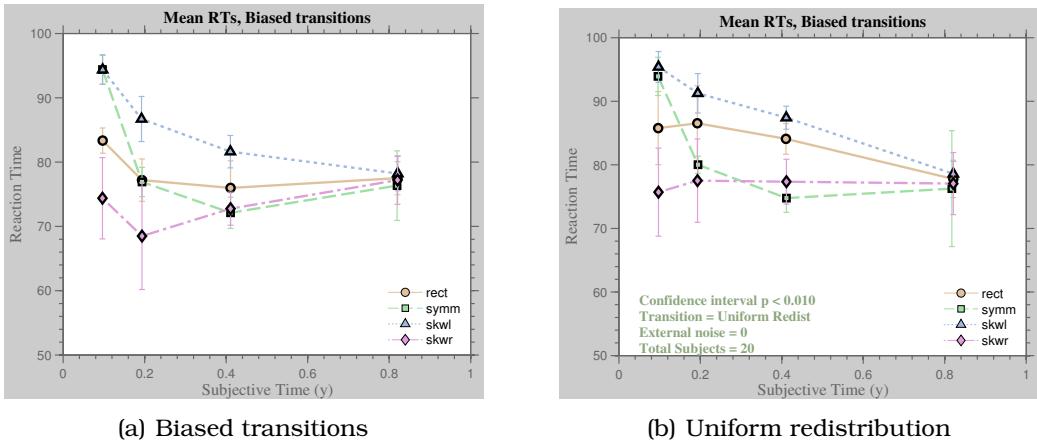


Figure 6: Reaction times for different distributions of foreperiod durations.

The results from the simulation replicate the following key observations of Baumeister och Joubert (1969): (a) the reaction times for different distributions differ at short foreperiods but not at the longest foreperiod, (b) the reaction times for rectangular, symmetrical and skewed-left distributions decrease monotonically with increasing foreperiod but the skewed-right distributions don't. The lack of foreperiod effect for skewed-right distribution is particularly significant as the informative value of time (increasing conditional probability of stimulus) is counterbalanced by decreasing probability of stimulus, an effect that has been replicated in several other experiments (Näätänen, 1971; Trillenberg m. fl., 2000).

## 4.3 Cue-based variability : taking into account endogeneous cues

### 4.3.1 Model

Consider now the observed value is the pair  $(\mathbf{x}, y)$ , consisting of the sensory signal  $\mathbf{x}$  and the time-varying internal signal  $y$ .

Let us call *prediction* the density calculated in two steps:

- (a) inverse step : for a sensory feature  $\mathbf{x}$ , identify the most probable source generating it. The calculation of the responsibilities allows to give a “score” to every source.
- (b) forward step : generate a probability density which is a mixture of source-related emission density functions.

An observation is now couple of measures  $(\mathbf{x}, y) \in \mathbb{R}^{d-1} \times \mathbb{R}$  (where the vector  $\mathbf{x}$  represents the sensory feature and the scalar  $y$  represents the *duration*), a discrete random variable  $k \in \{1, \dots, K\}$ , and their joint probability distribution  $P(\mathbf{x}, y, k)$ .

The objective is to predict  $y$  given  $\mathbf{x}$ , under the hypothesis that  $\mathbf{x}$  and  $y$  depend on the source  $k$  only, i.e.:

$$f(\mathbf{x}, y|k) = f(\mathbf{x}|k)f(y|k)$$

Where :

- $f(\mathbf{x}|k)$  is the probability density of the feature  $\mathbf{x}$  given hidden source  $k$ .
- $f(y|k)$  is the probability density of the final duration  $y$  given hidden source  $k$ .

Then it can be shown that :

$$f(y|\mathbf{x}) = \sum_{k=1}^K f(y|k)t_k(\mathbf{x}) \quad (22)$$

where  $t_k(\mathbf{x})$  is the responsibility given by

$$t_k(\mathbf{x}) = p(k|\mathbf{x}) = \frac{f(\mathbf{x}|k)\pi_k}{f(\mathbf{x})}$$

The role of  $t_k(\mathbf{x})$  is here to bias the mixing proportions  $\pi_k$  according to initial sensory information.

The parameters associated to the distribution of couples  $(\mathbf{x}, y)$  are learned using the same EM algorithm, adapted to the case of multivariate Gaussian densities (with an additional constraint of independence between  $\mathbf{x}$  and  $y$ ).

The effect of time passing on the mixture composed of  $t_1(\mathbf{x}), \dots, t_K(\mathbf{x})$  is the same as previously stated, i.e.:

$$f(y_i|\mathbf{x}, q_i) = \sum_{k=1}^K f(y_i|k)t_k(\mathbf{x}, q_i) \quad (23)$$

with

$$t_k(\mathbf{x}, q_i) = p(k|\mathbf{x}, q_i) = \frac{f(\mathbf{x}|k)\pi_k(q_i)}{f(\mathbf{x})}$$

### 4.3.2 Simulation results

Coull och Nobre (1998) showed that visual cues can be used by subjects as an exogenous factor for computing temporal expectation. In the simulations so far, we have assumed that the only distinguishable property of stimuli is their duration

– i.e. all stimuli have the same  $x$ , but could differ in the associated  $y$ . In the next experiment, we generalise this constraint and allow two different values of sensory stimuli  $x$ . These two values of sensory stimuli could be interpreted as two different visual stimuli, e.g. colors or shapes. Additionally, we allow for the possibility that the sensory stimuli can be used to predict the duration of stimuli – i.e. serves as an endogenous cue. This is done by associating the two values of  $x$  with two different durations of presentation – short (0.15s) or long (0.75s).

A key result of the Coull och Nobre (1998) study was that *valid* cues showed a reduction in RT compared to *invalid* cues at short SOAs but not long SOAs. In order to code for these valid and invalid cues, the model was trained on a set of conditioning trials where a value of  $x$  is associated with one of the two duration for 80% of trials and with the other duration for 20% of trials. The other value of  $x$  reversed this association, being associated with the first duration for 20% of trials while with the second duration for 80% of trials. In this way, each value of  $x$  validly predicts a duration for 80% of trials. We set the number of signal sources (prior hypotheses) to two, so that the model learns to generate both the valid and invalid data. Figure 7 shows the Gaussians associated to each source at the end of a typical conditioning phase. Each point on the graph shows the vector  $u = (x, y)$  at the end of a trial. The two ellipses show contours of the two Gaussian sources that are used to model the data. The colour of each dot shows the responsibility (the posterior) of each source for the given data point. One can observe that the model was able to classify the data according to its duration – both valid and invalid cues that appear for short durations are classified as one signal source (Gaussian) while longer durations are classified as the other source.

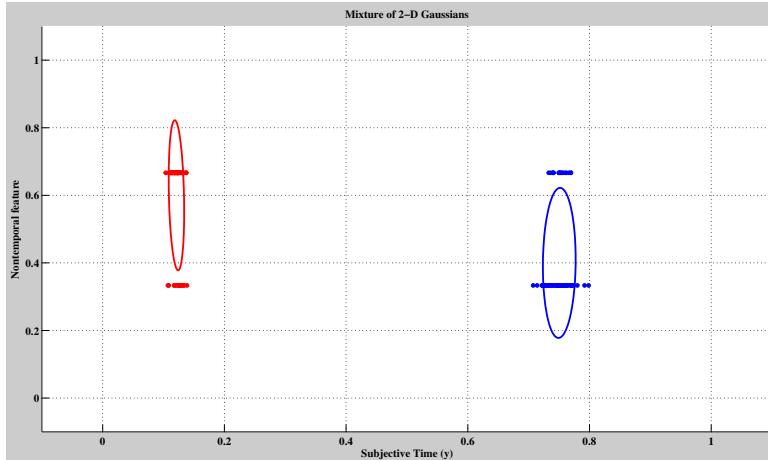


Figure 7: Classification through EM in case of non-temporal cue. The x-axis shows the duration of each stimuli while the y-axis shows a non-temporal feature (vector  $x$ ). The two ellipses show contours of the Gaussian probability densities for the two sources. The Gaussian on the left is the pdf for signal sources which appear for short duration while the pdf on the right is for signal sources that appear for a longer duration.

Once the model has learnt these parameters, we can find the RT that will be displayed by the model when it is given a stimulus with a particular value of the sensory dimension  $x$  and a duration. In order to do find these RTs, we generate the expectation  $e(y_i)$  for the two values of  $x$  for all values  $0 \dots 1$  – i.e. both when the cue  $x$  is valid or invalid. We then calculate the peak expectation around the short and long durations, which is reversed (as above) to find the RT. Figure 8(a) shows the expectations generated by the model as a result of learning for different durations

during the trial. Figure 8(b) shows these RTs for each cue, classified according to whether the cue was valid or invalid.

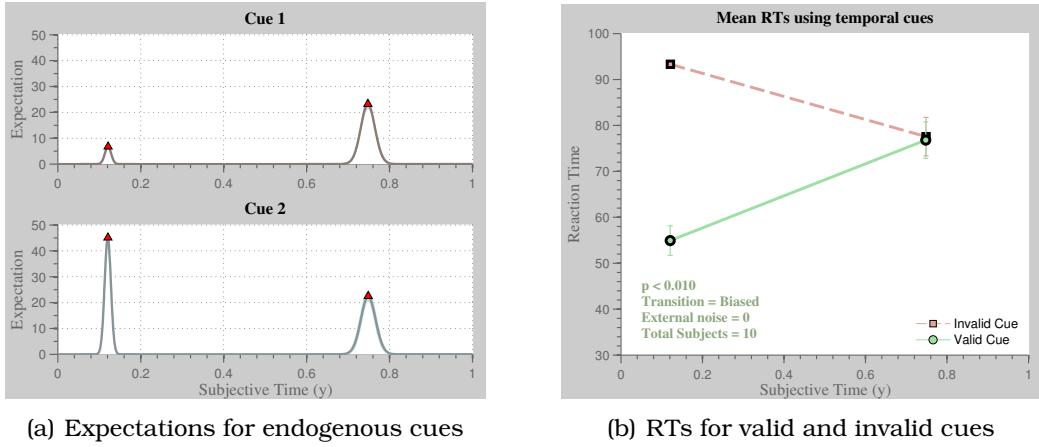


Figure 8: Expectations and reaction times for two endogenous cues. Figure (a) plots the expectations for each cue for all durations while Figure (b) plots the reaction times for valid and invalid cues at short and long durations.

The reader can observe that the simulation reproduces the key result from Coull och Nobre (1998): valid cues significantly reduce RTs at the shorter interval but have no significant effect on RT at the longer interval. The reason for this result is clear: the informative value of time helps to predict the RS in case of an invalidly cued long interval (i.e. when the cue predicted that the interval will be short but turned out to be long). Thus the informative value of time *compensates* for the lack of information present in the invalid cue. One aspect of the simulation results that are different from those observed by Coull och Nobre (1998) are that the reaction time in case of valid trials is longer at long foreperiods compared to short foreperiods. The reason for this is the inbuilt Webber's law in the model – longer intervals have larger variance and lower peak expectations while shorter intervals have short variance but a sharp (and large) peak expectation (compare the two expectations for Cue 2 in Figure 8(a)).

#### 4.4 Inter-trial variability : the sequential effect

It has been observed that that RTs show a *sequential effect*: the RT of a subject is longer if the foreperiod on the immediately previous trial (hereafter, PFP) is longer than the current foreperiod (hereafter, FP) as compared to when the PFP is the same length or shorter than FP (Woodrow, 1914; Niemi & Näätänen, 1981). Thus, when a subject has just observed a longer foreperiod, it seems less prepared for a stimulus than when it has seen a comparatively shorter foreperiod. Klemmer (1956) replicated this sequential effect and also observed that it is a short-term effect so that the foreperiod two trials back (PPFP, if you may) has no effect on RT. Furthermore, the effect of PFP on RT is asymmetrical: the RT of a subject does not significantly change when PFP is shorter than FP (Karlin, 1959; Granjon & Reynard, 1977). These asymmetrical effects from PFP on RT mean that a short FP is more likely to ‘suffer’ from the effects of PFP than a comparatively longer FP: both short and long PFPs will have sequential effect on a short FP, while only long PFPs will affect a long FP. Hence the effect from PFP will, at least in part, be responsible for the longer RT at shorter foreperiods.

TODO

As we reviewed at the beginning of this report, researchers have attributed the foreperiod effect to two different mechanisms – a strategic mechanism where subjects update the conditional probability of occurrence of RS with the passage of time (Elithorn & Lawrence, 1955) and a conditioning mechanism where learning during one trial changes the expectation of RS during subsequent trials (Los & van den Heuvel, 2001). These two mechanisms are generally considered to act separately on the generation of expectation (Vallesi & Shallice, 2007) and it is not understood how the two types of learning can be combined together. Our next simulation is aimed at understanding the mechanism for the combination of these two types of information.

In the probabilistic framework presented above, the strategic mechanism of generating expectations consists of two operations: learning the probabilities of occurrence of different response stimuli and updating these probabilities with the passage of time. As we showed above, the probabilities of different response stimuli is learnt through the EM algorithm. This provides a *long-term* memory in the form of probability densities for different signal sources and their mixing proportions. If we now suppose that, in addition to this long-term memory, subjects also have a *short-term* memory of the duration of a response stimulus. Specifically, subjects could increase or decrease the mixing proportion of a signal source in response to whether that signal source lead to an observed stimulus. This gives a mechanism for encoding sequential effects from one trial to the next akin to the ones proposed by (?) and Los, Knol och Boers (2001).

In a similar manner to previous simulations, the model was trained using a series of 160 trials for RS of either short (0.2s), medium (0.4s) or long (0.8s) durations. In this case, however, instead of recording the expectation during only one trial, the expectation was recorded for a series of two trials, where each trial could be of either short, medium or long duration. In between the trials, the mixing proportions were adjusted based on a short-term learning mechanism. For the purpose of this study, we tested four different kind of adjustments to mixing proportions between trials (a) short-term potentiation: small increase in the mixing proportion of the source that correctly predicted the RS, (b) short-term depression: small decrease in the mixing proportion of the sources that incorrectly predicted the RS, (c) time-dependent depression, positive decay: like short-term depression but with larger depression for sources farther away from RS than closer, and (d) time-dependent depression, negative decay: like short-term depression but with larger depression for sources closer to RS than farther away. Figure 9 shows the RTs for the second trial at each duration of RS and each length of foreperiod during the first trial. Each sub-figure shows the results for one type of learning mechanism.

Comparing the results for different learning mechanisms, it can be observed that the time-dependent depression with larger depression for sources farther away from actual RS give results that are closest to the asymmetrical effect of previous foreperiod on RTs frequently observed in psychological experiments (Karlin, 1959; Vallesi & Shallice, 2007). For this conditioning rule, the longest previous foreperiod affects the RT for the short and medium foreperiod, but has no affect on the longest foreperiod. The medium previous foreperiod affects only the RT for short foreperiod.

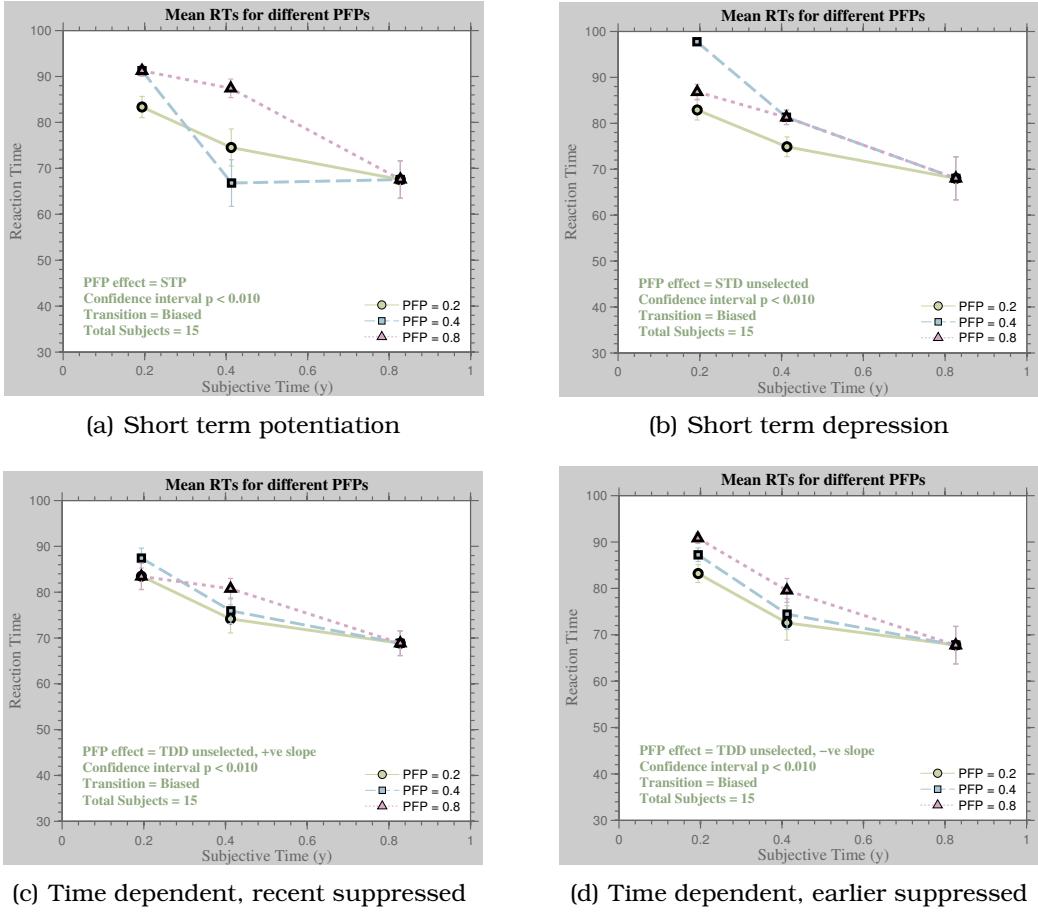


Figure 9: Effect of previous foreperiods (PFPs) on RTs as a result of different conditioning mechanisms. Three different durations of foreperiods are used: short (0.2s), medium (0.4s) and long (0.8s)

## 5 Discussion : Neural substrates of temporal prediction

NOT CHANGED FROM THE INITIAL DRAFT –*à* A RECAST ACCORDING TO OUR MODELLING PROPOSAL IS NEEDED

The behavioural studies discussed above have highlighted that computation of temporal expectation involves a number of functionally discrete processes. When an animal reacts to a stimulus, the information processing performed by the animal will rely on these discrete processes. The question that now arises is how are these processes realised in the animal's brain. Answering this question will help us answer another, more important question: do different processes rely on disjoint cognitive resources and how does computation of temporal expectation depend on more general cognitive mechanisms of memory, attention, conscious control (intentionality) and prediction.

A number of neuropsychological and neurological studies have investigated the mapping of temporal prediction in the brain. While several different brain regions are involved in timing (see Coull, Cheng & Meck, 2011; Buhusi & Meck, 2005), three broad regions are implicated by studies involving temporal predictions: the prefrontal cortex, the cerebellum and the parietal cortex. Within each of these regions, investigations have revealed anatomically localized zones that are involved in functionally dissociable processes involved in generating temporal predictions.

### 5.0.1 The prefrontal cortex

In the previous section we discussed how the RTs reveal that animals model temporal uncertainty of stimulus and the foreperiod effect reveals that they compute stimulus expectancy that changes with time. This time-varying expectancy seems to crucially depend on the prefrontal cortex. In experiments with 38 patients suffering from brain lesions to frontal lobes, Stuss m. fl. (2005) found that lesions to right dorsolateral prefrontal cortex (centered around Broadman area 9/46v) affected the variable foreperiod effect. While controls and patients with lesions to other frontal regions showed a decrease in RT with increase in FP, patients with lesions to this region showed a significant increase in RT. Stuss m. fl. (2005) concluded that the right dorsolateral prefrontal cortex (hereafter, rDLPFC) plays a crucial role in a monitoring process that allows the animal to use the longer FP to prepare more fully for a fast response.

This monitoring role of rDLPFC is confirmed by TMS studies. Vallesi, Shallice och Walsh (2007) replicated the results of Stuss m. fl. (2005) in healthy adults by applying transcranial magnetic stimulation to the rDLPFC. In their experiment, Vallesi m. fl. (2007) apply TMS to three brain regions: the right angular gyrus, the left DLPFC and rDLPFC. They observed that TMS applied to rDLPFC significantly reduced the foreperiod effect while it had no effect on FP when applied to the left contralateral site or the right angular gyrus. Additionally, Vallesi m. fl. (2007) observed that TMS to rDLPFC did not significantly affect the sequential effect, supporting the results from Vallesi och Shallice (2007) which showed a developmental dissociation between foreperiod and sequential effects (discussed above).

The involvement of the rDLPFC in temporal processing suggested by these lesion studies is corroborated by brain imaging using fMRI. Rao, Mayer och Harrington (2001) conducted event-related fMRI recordings during a time perception task and surveyed the time-course of activation of different brain regions. By looking at this time-course of activation, Rao m. fl. (2001) were able to infer whether a region was involved in encoding (early activation) or comparison and response (late activation). They found that the rDLPFC was activated late and uniquely for performing time discriminations during this task, confirming its role in a comparison of time intervals. This comparison of intervals is an *explicit timing* task and contrasts with the foreperiod effect usually observed in *implicit timing* tasks. However, both tasks involve a monitoring of time intervals and require a temporary storage and comparison of information. Rao m. fl. (2001) argued that their results suggest that the rDLPFC participates in a working memory of time interval and is involved in an 'executive circuit' responsible for comparison and manipulation of stored information. These conclusions are supported by studies that show rDLPFC is crucial to temporal reproduction (Basso, Nichelli, Wharton, Peterson & Grafman, 2003; Jones, Rosenkranz, Rothwell & Jahanshahi, 2004), a task that involves monitoring the passage of time and comparing it to an interval stored in memory.

### 5.0.2 The parietal cortex

A number of studies suggest that the parietal cortex plays a role in predicting the length of time intervals. We mentioned above the study performed by Coull och Nobre (1998) where the recorded RTs showed that participants learnt to predict SOAs based on endogenous cues. In addition to recording RTs, this study also measured brain activity of these participants using PET and fMRI. They found that

spatial orienting led to preferential activation of right parietal cortex while temporal orienting (using endogenous cues) led to preferential activation of the left parietal cortex. Specifically, both the PET and fMRI recordings showed that temporal orienting activates the left cerebellum (see below) and the left intraparietal sulcus (IPS), extending into the inferior parietal lobule (Broadman area 40). Coull och Nobre (1998) concluded that regions within the left parietal cortex perform an analogical role in temporal orientation to the role performed by the right parietal cortex in spatial orientation and form a part of the frontoparietal attentional network (Mesulam, 1981; Posner & Petersen, 1990).

In the above experiment participants learnt to predict temporal intervals based on endogenous cues, implying top-down processing. These results are complemented by a recent study in which participants learnt to predict time implicitly, based on exogenous cues. Coull, Vidal, Goulon, Nazarian och Craig (2008) designed an experiment in which the participants viewed an animation showing a car approaching a wall at a decelerating velocity. The animation stopped before the car came to a halt (or collided with the wall) and the participants were asked to judge whether the car would have collided with the wall. In order to make this judgement, the participants must predict the time-to-contact (TTC) and compare it with the time at which the car would come to a halt, given the velocity and (constant) deceleration. Thus, the velocity and the deceleration were exogenous cues that allowed the participants to predict time. The certainty with which a participant could predict TTC depended on the starting distance between the car and the wall, its velocity and deceleration, parameters that varied from trial to trial. Using event-based fMRI recording, Coull m. fl. (2008) observed that performing the TTC judgments activated the left pars opercularis of the inferior frontal lobe and the supramarginal gyrus of the inferior parietal cortex. Furthermore, Coull m. fl. (2008) also observed that as the uncertainty of a trial's outcome increased, the activity in left inferior parietal cortex increased<sup>7</sup>, confirming a temporally predictive role based on spatial cues.

An earlier study conducted by Assmus m. fl. (2003) also found neural activity in the left inferior parietal cortex when participants used implicit spatial cues in making temporal judgements. A follow-up study not only reproduced these results, but also confirmed a functional role of the left inferior parietal cortex in TTC judgements through a parametric design. Assmus, Marshall, Noth, Zilles och Fink (2005) varied the difficulty of making TTC judgements and observed that as the difficulty of predicting the TTC increased, there was a linear increase in the BOLD-signal recorded in the left inferior parietal cortex. The authors concluded that this region performed a functionally important role in spatio-temporal integration. These studies demonstrate that the left inferior parietal cortex is involved in bottom-up processing of stimuli when temporal predictions need to be implicitly computed. They complement studies of temporal orienting where a top-down processing of cues explicitly predicts a time interval and activates the same brain region.

The increase in activity in the parietal cortex with the increase in difficulty in making a temporal decision implies that the parietal cortex is involved in resolving temporal uncertainty. However, there is evidence showing that the role of parietal cortex is not limited to the resolution temporal uncertainty in stimulus but

---

<sup>7</sup>This increase was observed for the *egocentric* and not *allocentric* point of view. An increase in activity was also observed in the right temporal pole during trials with an allocentric point of view and bilaterally for mid/superior temporal cortex and dorsomedial visual cortex during trials with an egocentric point of view.

also extends to resolving the uncertainty in choosing a response. Sakai m. fl. (2000) conducted experiments in which participants carried out choice reaction time tasks that could contain temporal uncertainty, response uncertainty, or both. Simultaneously, Sakai m. fl. (2000) performed fMRI and observed that the neural activity in left intraparietal sulcus increased (relative to control) when participants performed the temporal uncertainty task. Interestingly, they observed that activity in this region also increased when the participants performed the response uncertainty task. In contrast, the presupplementary motor area / rostral cingulate motor area (preSMA/rCMA) showed increased activation only during the response uncertainty task while the posterior lobe of the cerebellum showed increased activation only during the temporal uncertainty task. These results demonstrate that while the preSMA/rCMA and cerebellum are selectively involved in response selection and timing adjustment respectively, the IPS is involved in both types of decisions.

The above studies indicate that the left parietal cortex is involved in computation of time intervals. In particular, the increasing activity in this area with rise in temporal uncertainty demonstrates a relation to generation of temporal expectation. However, one could rightly ask whether this relation is causal or epiphenomenal – i.e. is the neural activity in the left parietal cortex responsible for making the temporal prediction or merely a consequence of this prediction, made elsewhere? In order to examine a causal role, one needs to look at the time course of the neural activity and establish whether the activity precedes the behaviour. While the studies presented above use fMRI and provide low temporal resolution, evidence for an anticipatory role of the parietal cortex is available from single unit recordings in primates. Mackay och Crammond (1987) recorded from 2365 single units in the posterior parietal lobe of awake monkeys and found that 168 units, located either side of IPS, showed anticipatory activity. They found that many neurons in this region increased their discharge rate on somatic approach and also immediately prior to the expected time of reward.

In another study, Duhamel, Colby och Goldberg (1992) also found anticipatory activity in neurons in the lateral intraparietal (LIP) area<sup>8</sup> in preparation of a change in the neuron's receptive field before a saccade. Certain neurons in this area fired only when the monkey intended to make a saccade that would bring a stimulus into the new fixation's receptive field, leading them to argue that cells in this area perform a *predictive remapping* of the stimulus from the current coordinates to the predicted coordinates. These results were extended by Snyder, Batista och Andersen (1997), who showed that anticipatory activity in cells recorded from the LIP depended on the type of intended movement. For the same sensory stimulus, one set of neurons increased discharge anticipating saccadic eye movement while another set of neurons increased discharge anticipating reaching movement, showing that neurons encoded intended movement. Together, these results indicate that the anticipatory activity in posterior parietal cortex is general to performing prediction, be it the predictive mapping in the sensory pathway or the predicted movement in the motor pathway.

### 5.0.3 The cerebellum

We mentioned above that the experiments by Coull och Nobre (1998) and Sakai m. fl. (2000) showed neural activity in the left cerebellum during temporal orient-

---

<sup>8</sup>It has been argued that the mid-posterior IPS is the human homologue of the LIP of the macaque monkey brain (Culham & Kanwisher, 2001)

ing task. The participation of cerebellum in computing temporal expectation is corroborated by data available from patients with cerebellar lesions. Ivry och Keele (1989) showed that these patients showed an impairment of motor timing, a result that has been replicated by several other studies (see Coull m. fl., 2011). Temporary cerebellar lesions in healthy adults through TMS also reproduce this finding. Koch m. fl. (2007) found that applying TMS to cerebellum impaired timing during a temporal reproduction task. They also noticed that timing was impaired only in the millisecond range (when subjects were asked to reproduce 400–600ms intervals), but not in the seconds range (1,600–2,400ms). The selective impairment to timing in the millisecond but not seconds range seems to be a robust effect (Fierro m. fl., 2007; Lee m. fl., 2007).

The predictive role of cerebellum in computing temporal expectation is clear from studies on implicit timing. In a study similar to that of Assmus m. fl. (2005), O'Reilly, Mesulam och Nobre (2008) conducted an experiment in which participants predicted the position of a moving target that became occluded for a fixed period of time (600 ms). During this period of occlusion either the velocity or the position of the target was perturbed. Participants were asked to make either a 'temporal-spatial' judgement – where they compared the distance covered by the target during occlusion to their prediction – or a 'spatial' judgement – where participants compared their prediction to the position of the target. O'Reilly m. fl. (2008) recorded the participants' neural activity during the experiment and found that while areas in the frontal and parietal cortex were activated for both types of judgements, the posterior cerebellum (lobule VII, crus I) was activated specifically when the participants made a temporal-spatial prediction. In an independent study, Beudel, Renken, Leenders och de Jong (2009) also found that the prefrontal cortex, preSMA and left cerebellum were activated when participants used spatial cues (distance and velocity) to judge whether a moving target would reach a given point within or after 3 seconds. Beudel m. fl. (2009) controlled for neural activation as a result of spatial processing and also showed that these regions are activated for momentary temporal estimation (regions activated for speed judgement – regions activated for spatial processing).

The hypothesis that the cerebellum is involved in predictive processing is supported by the role accorded to it in motor control. An animal can learn to perform fast movements by either using feedback or by internally modelling the consequences of its movement. Two types of internal models have been proposed: *inverse models* compute the motor commands that must be issued in order to achieve a desired trajectory of the controlled object and *forward models* compute the consequences of outgoing motor commands on controlled objects (Kawato, 1999). The key advantage to forward modelling is that it reduces the delay associated with feedback from visual and proprioceptive systems and allows an animal to predict the consequences of motor commands before realising the movement. Wolpert and colleagues have proposed that the cerebellum realises such a forward model (Miall, Weir, Wolpert & Stein, 1993; Wolpert, Ghahramani & Jordan, 1995; Wolpert, Miall & Kawato, 1998), supporting this claim using neurophysiological data (Kawano m. fl., 1996), functional imaging data (Imamizu m. fl., 2000; Blakemore, Wolpert & Frith, 1998), data from lesion studies (Baizer, Kralj-Hans & Glickstein, 1999) and single-cell studies (Miall, Keating, Malkmus & Thach, 1998).

Given this role of cerebellum in motor control, O'Reilly m. fl. (2008) have argued that their results extend the notion of forward modelling to the perceptual domain – i.e. just like the animal feeds motor commands to the forward model to obtain

a likely trajectory, it is able to predict a temporal property from perceptual stimuli by feeding this stimuli (such as position and velocity) to a forward model, which outputs the likely time for the stimuli to acquire a given state (such as a different position). Different forward models in the cerebellum can give different likelihoods for movement trajectories / temporal properties. Thus the forward model can be given a general probabilistic interpretation according to which it gives the likelihood of motor actions or perceptual stimuli. Wolpert och Kawato (1998) have given the forward model such an interpretation for the problem of adapting motor control to multiple environments. In the current study, we will describe a probabilistic model that associates temporal properties with the likelihood of observing certain perceptual stimuli and can therefore be seen as a probabilistic interpretation of a forward model for predicting time. **I'm still fleshing out this argument.**

## Referenser

- Assmus, A., Marshall, J. C., Noth, J., Zilles, K. & Fink, G. R. (2005). Difficulty of perceptual spatiotemporal integration modulates the neural activity of left inferior parietal cortex. *Neuroscience*, 132, 923–927.
- Assmus, A., Marshall, J. C., Ritzl, A., Noth, J., Zilles, K. & Fink, G. R. (2003). Left inferior parietal cortex integrates time and space during collision judgments. *Neuroimage*, 20, S82–S88.
- Baizer, J. S., Kralj-Hans, I. & Glickstein, M. (1999). Cerebellar lesions and prism adaptation in macaque monkeys. *Journal of Neurophysiology*, 81, 1960–1965.
- Basso, G., Nichelli, P., Wharton, C. M., Peterson, M. & Grafman, J. (2003). Distributed neural systems for temporal production: a functional mri study. *Brain Res Bull*, 59, 405–411.
- Baumeister, A. A. & Joubert, C. E. (1969). Interactive effects on reaction time of preparatory interval length and preparatory interval frequency. *Journal of Experimental Psychology*, 82(2), 393–395.
- Beudel, M., Renken, R., Leenders, K. L. & de Jong, B. M. (2009). Cerebral representations of space and time. *Neuroimage*, 44, 1032–1040.
- Blakemore, S. J., Wopert, D. M. & Frith, C. D. (1998). Central cancellation of self produced tickle sensation. *Nature Neuroscience*, 1, 635–640.
- Buhusi, C. V. & Meck, W. H. (2005). What makes us tick? functional and neural mechanisms of interval timing. *Nature Reviews Neuroscience*, 6, 755–765.
- Coull, J. T. (2009). Neural substrates of mounting temporal expectation. *PLoS Biology*, 7(8), 1–4.
- Coull, J. T., Cheng, R.-K. & Meck, W. H. (2011). Neuroanatomical and neurochemical substrates of timing. *Neuropsychopharmacology Reviews*, 36, 3–25.
- Coull, J. T. & Nobre, A. C. (1998). Where and when to pay attention: The neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *The Journal of Neuroscience*, 18(18), 7426–7435.

- Coull, J. T., Vidal, F., Goulon, C., Nazarian, B. & Craig, C. (2008). Using time-to-contact information to assess potential collision modulates both visual and temporal prediction networks. *Frontiers in Human Neuroscience*, 2, 1–12.
- Culham, J. C. & Kanwisher, N. G. (2001). Neuroimaging of cognitive functions in human parietal cortex. *Current Opinion in Neurobiology*, 11, 157–163.
- Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1), 1–38.
- Drazin, D. (1961). Effects of foreperiod, foreperiod variability, and probability of stimulus occurrence on simple reaction time. *Journal of Experimental Psychology*, 62, 43–50.
- Duhamel, J. R., Colby, C. & Goldberg, M. E. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science*, 255, 90–92.
- Elithorn, A. & Lawrence, C. (1955). Central inhibition – some refractory observations. *Quarterly Journal of Experimental Psychology*, 11, 211–220.
- Elliott, R. (1973). Some confounding factors in the study of preparatory set in reaction time. *Memory & Cognition*, 7, 13–18.
- Fierro, B., Palermo, A., Puma, A., Francolini, M., Panetta, M. L. & al, O. D. et. (2007). Role of the cerebellum in time perception: a tms study in normal subjects. *Journal of Neurological Science*, 263, 107–112.
- Gibbon, J. (1977). Scalar expectancy theory and weber's law in animal timing. *Psychological Review*, 84(3), 279–325.
- Gibbon, J. (1991). Origins of scalar timing. *Learning and Motivation*, 22, 3–38.
- Gold, J. I. & Shadlen, M. N. (2001, January). Neural computations that underlie decisions about sensory stimuli. *TRENDS in Cognitive Science*, 5(1), 10–16.
- Granjon, M. & Reynard, G. (1977). Effect of the length of the runs of repetitions on the simple rt-isi relationship. *Quarterly Journal of Experimental Psychology*, 29, 283–295.
- Hauer, B. J. & MacLeod, C. M. (2006). Endogenous versus exogenous attentional cuing effects on memory. *Acta Psychologica*, 122, 305–320.
- Imamizu, H., Miyauchi, S., Tamada, Sasaki, Y., Takino, R., Putz, B. m. fl. (2000). Human cerebellar activity reflecting an aquired internal model of a novel tool. *Nature*, 403, 192–195.
- Ivry, R. B. & Keele, S. W. (1989). Timing functions of the cerebellum. *Journal of Cognitive Neuroscience*, 1, 136–152.
- Jones, C. R., Rosenkranz, K., Rothwell, J. C. & Jahanshahi, M. (2004). The right dorsolateral prefrontal cortex is essential in time reproduction: an investigation with repetitive transcranial magnetic stimulation. *Ex-*

- perimental Brain Research*, 158, 366–372.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82, 35–45.
- Karlin, L. (1959). Reaction time as a function of foreperiod duration and variability. *Journal of Experimental Psychology*, 55, 185–191.
- Kawano, K., Takemura, A., Inoue, Y., Kitama, T., Kobayashi, Y. & Mustari, M. J. (1996). Visual inputs to cerebellar ventral paraflocculus during ocular following responses. *Prog Brain Res*, 112, 415–422.
- Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, 9, 718–727.
- Klemmer, E. T. (1956). Time uncertainty in simple reaction time. *Journal of Experimental Psychology*, 51, 179–184.
- Klemmer, E. T. (1957). Simple reaction time as a function of time uncertainty. *Journal of Experimental Psychology*, 54, 195–200.
- Koch, G., Oliveri, M., Torriero, S., Salerno, S., Lo Gerfo, E. & Caltagirone, C. (2007). Repetitive tms of cerebellum interferes with millisecond time processing. *Experimental Brain Research*, 179, 291–299.
- Lee, K. H., Egleston, P. N., Brown, W. H., Gregory, A. N., Barker, A. T. & Woodruff, P. W. (2007). The role of the cerebellum in subsecond time perception: evidence from repetitive transcranial magnetic stimulation. *Journal of Cognitive Neuroscience*, 19, 147–157.
- Los, S. A., Knol, D. L. & Boers, R. M. (2001). The foreperiod effect revisited: Conditioning as a basis for nonspecific preparation. *Acta Psychologica*, 106, 121–145.
- Los, S. A. & van den Heuvel, C. E. (2001). Intentional and unintentional contributions to nonspecific preparation during reaction time foreperiods. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 370–386.
- Mackay, W. A. & Crammond, D. J. (1987). Neuronal correlates in posterior parietal lobe of the expectation of events. *Behavioral Brain Research*, 24, 167–179.
- McLachlan, G. J. & Krishnan, T. (1997). *The EM algorithm and extensions*. New York: John Wiley & Sons, Inc.
- Mesulam, M. M. (1981). A cortical network for directed attention and unilateral neglect. *Annals of Neurology*, 10, 309–325.
- Miall, R. C., Keating, J. G., Malkmus, M. & Thach, W. T. (1998). Purkinje cell complex spikes are predicted by simple spike activity. *Nature Neuroscience*, 1, 13–15.
- Miall, R. C., Weir, D. J., Wolpert, D. M. & Stein, J. F. (1993). Is the cerebellum a smith predictor? *Journal of Motor Behavior*, 25, 203–216.
- Näätänen, R. (1971). Nonaging foreperiod and simple reaction time. *Acta Psychologica*, 35, 316–327.
- Niemi, P. & Näätänen, R. (1981). Foreperiod and simple reaction time. *Psychological Bulletin*, 89(1), 133–162.
- O'Reilly, J. X., Mesulam, M. M. & Nobre, A. C. (2008). The cerebellum

- predicts the timing of perceptual events. *Journal of Neuroscience*, 28, 2252–2260.
- Posner, M. I. & Petersen, S. E. (1990). The attentional system of the human brain. *Annual Review of Neuroscience*, 13, 25–42.
- Rao, S. M., Mayer, A. R. & Harrington, D. L. (2001). The evolution of brain activation during temporal processing. *Nature Neuroscience*, 4, 317–323.
- Sakai, K., Hikosaka, O., Takino, R., Miyauchi, S., Nielsen, M. & Tamada, T. (2000). What and when: parallel and convergent processing in motor control. *Journal of Neuroscience*, 20, 2691–2700.
- Simen, P., Balci, F., deSouza, L., Cohen, J. D. & Holmes, P. (2011). A model of interval timing by neural integration. *The Journal of Neuroscience*, 31(25), 9238–9253.
- Snyder, L. H., Batista, A. P. & Andersen, R. A. (1997). Coding of intention in the posterior parietal cortex. *Nature*, 386, 167–170.
- Stuss, D. T., Alexander, M. P., Shallice, T., Picton, T. W., Binns, M. A. & al, R. M. et. (2005). Multiple frontal systems controlling response speed. *Neuropsychologia*, 43, 396–417.
- Trillenberg, P., Verleger, R., Wascher, E., Wauschkuhn, B. & Wessel, K. (2000). CNV and temporal uncertainty with 'ageing' and 'non-ageing' S1-S2 intervals. *Clinical Neurophysiology*, 111, 1216–1226.
- Vallesi, A. & Shallice, T. (2007). Developmental dissociations of preparation over time: Deconstructing the variable foreperiod phenomena. *Journal of Experimental Psychology: Human Perception and Performance*, 33, 1377–1388.
- Vallesi, A., Shallice, T. & Walsh, V. (2007). Role of the prefrontal cortex in the foreperiod effect: Tms evidence for dual mechanisms in temporal preparation. *Cerebral Cortex*, 17, 466–474.
- Wolpert, D. M., Ghahramani, Z. & Jordan, M. I. (1995). Are arm trajectories planned in kinematic or dynamic coordinates? an adaptation study. *Experimental Brain Research*, 103, 460–470.
- Wolpert, D. M. & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, 11, 1317–1329.
- Wolpert, D. M., Miall, R. C. & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Science*, 2, 338–347.
- Woodrow, H. (1914). The measurement of attention. *Psychological Monographs*, 17(5).