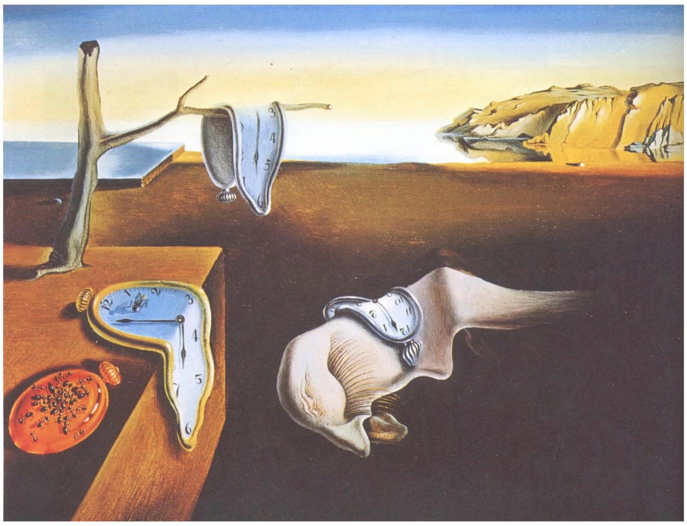
Master Computational Neurosciences and Neuroengineering

Supervised Project

Bayesian Modelling of the unfolding of time prediction

Mathias Vigouroux

Supervisor : Dr. Sophie Herbst



2021

Contents

1 Introduction

1.1 Context

With space, time is one of the fundamental characteristic of our dynamically evolving environment, underpinned by its own temporal structure. Anticipating and predicting events in our environment is thus decisive to guide our behaviour. Indeed, cognitive systems have evolved to predict the when, where, and what of the sensory environment.

If one can focus on time explicitly, usually our prediction and processing of the underlying time structure of our environment will remain unconscious, engaging in an implicit timing task. However, extracting temporal predictions from our sensory environment is mainly used to optimize our behavior. For instance, this will allow us to catch something in the air or cross a busy road for instance (Correa et al.,2006,Correa et al.,2010). In time litterature, this ability is called it is called temporal preparation.

Hitherto, understanding and modelling this extracting process of the statistical properties of our environment in order to form temporal predictions is still unknown.

Key research questions :

- How are the statistical properties of our environment stored in our brain ?
- What is the model that best fit this internal representation ?

1.2 Goal of the supervised project

A previous online experiment conducted by Elisa Lannelongue, former intern in the team, allowed to collect and analyse pilot data on temporal expectation. The goal of this supervised project is therefore to

- Familiarize with the time and bayesian litterature

- Asses which type of Bayesian Model is best fitted to describe the pilot data
- Propose improvement of the experimental setup that might be implemented in the second part of the project.

Indeed, this supervised project will be incorporated as the first part of my master thesis in the Cognition and Brain Dynamics team at Neurospin. Thus, my first objective was t really understand what have been done by the team and where can some improvements be made in a second part.

2

Theoretical framework : time perception and Bayesian statistics

2.1 The foreperiod

2.1.1 The fixed foreperiod paradigm

In laboratory settings, the main paradigm used to study temporal preparation is the "foreperiod paradigm" (Niemi and Na a ta nen, 1981). The foreperiod is the time interval between a cue stimulus and a target stimulus requiring a fast and accurate response. In this case, the response time to the stimuli is meant to reflects temporal preparation : correctly anticipated targets will elicit faster response times while unmet expectations will incur behavioral costs with lower response times. In the case of an experimental task based on a fixed foreperiod paradigm, the foreperiod duration is held constant throughout a given block of trials. For instance, we could imagine an experiment in which the target stimulus would always occurs 500ms after the cue stimulus throughout block 1, then the foreperiod would change for the next block and the target would appear 1000ms after the cue. In experiments based on a fixed foreperiod paradigm, it has been observed that participants' response times tend to be faster for the shorter foreperiods. This phenomenon is known as the "fixed foreperiod effect" (Vallesi et al., 2009). The 'fixed foreperiod effect' can be explained by the effect of scalar variability observed for temporal predictions : the uncertainty in duration estimation increases as a function of the time interval being estimated according to the scalar theory (Gibbon, 1977), which means that response times should also increase with longer foreperiods.

2.1.2 The variable foreperiod paradigm

Another branch of the foreperiod paradigm involves varying foreperiods within blocks as it is the case in this project for the second experimental task. If varying foreperiod durations are uniformly interleaved across trials, participants will tend to be faster for targets appearing after longer intervals. For instance, we could imagine a task in which there are two possible foreperiods of 500ms and 1000ms and for each trial one of the two foreperiods is randomly presented to the participant. As time goes by within a trial and the target has not yet appeared after the estimated timepoint at which the shorter foreperiods should have occured, participants will infer that it will occur after the longest ones. This phenomenon is known as “the variable foreperiod effect” (Niemi and Na a ta nen, 1981). Classically in the literature, the mechanism explaining the variable foreperiod effect involves the monitoring of the hazard rate of stimuli appearance to optimize temporal preparation (Nobre and van Ede, 2018).

2.1.3 The temporal orienting paradigm

Another useful paradigm to study the effects of temporal preparation is the temporal orienting task. This paradigm is based on Posner's spatial orienting task (Posner et al., 1980) adapted to temporal preparation. In the case of a task based on the temporal orienting paradigm, if there are two possible foreperiods of 500ms and 1000ms, an explicit cue would predicts with a high probability the specific foreperiod duration after which the target stimuli will occur. Temporal orienting effects are typically such that a shorter foreperiod elicits faster and more accurate responses to validly-cued targets as compared to targets occurring at an unexpected moment. In the case of longer foreperiods, the benefits of temporal orienting effects are usually not as marked because participants would tend to reorient their attention to the long interval if the target has not appeared as early as expected, which counterbalances the cost of an invalid temporal expectation (Correa et al., 2006; Coull and Nobre, 1998).

In this project, the second task is based on a variable foreperiod paradigm. However, the main difference with much of the literature on experiments using the same paradigm is that foreperiods are drawn from continuous Gaussian distributions. This allows us to investigate in greater details the influence of the hazard function and the effect of surprise on behavioral responses.

2.2 Generalities on Bayesian Statistics

One of the most widely use mathematical framework to study how we use dynamical sensory information in order to make inference and update them is the Bayesian frame work

2.3 Example of Bayesian inference

Let suppose that you enter a neuroscience laboratory and the experimenter ask you to predict the duration of a foreperiod between two stimuli that he will present you. The duration of the stimuli will be what we call the "state of the word", s.

Before the experiment has even started you will make some assumption on this duration. Indeed, you know that it won't long a hundred years, as dead people rarely report an answer for this question, and that it will be above femtoseconds, because otherwise it will be impossbile for you to report that there were even two stimuli. Thus, you create a first distribution of the duration. Moreover, you also overheard the conversation of the prior participant who was saying that the duration was of 30 seconds, so you give some credit to what you heard. This first assumption is what we call the prior probability, $p(s)$: prior probability on the state of the world s.

Then the experiment begin and you receive the second stimulus two minutes after the first one. However, you have already did some probability and know that you don't want to draw conclusions too quickly. What you decide is to revise your expectation by trying to asses what is the likelihood of measuring a duration of two minutes if I am in a given state of the world s. You will start, with extreme values of 100 years or femtoseconds and compute a low likelihood of the measure for this given states $p(x_{trial}|s)$.

However, you don't want to completely forgot what you have heard in the waiting room, and keep some importance to the 30 seconds assumptions. To add end you will multiply this newly calculated likelihood with you prior probability. From this updated, you got the posterior probability $p(s|x_{trial})$, which correspond to your new belief on the state of the world (the foreperiod duration) knowing what you have measure.

More formerly, let's take the example of Inference under Measurement Noise of Sound Localization from[**?**].

We have a :

- s : "state of the world" (here the position of the sound source)
- $p(s)$: prior probability of s
- $p(x_{trial}|s)$: likelihood of the measure for a given state
- $p(s|x_{trial})$: posterior probability of s knowing that we measure x.
- x_{trial} : "measurement from the ear" (our noisy measurement)

- Bayes' rule :

$$p(x_{trial}|s) = \frac{p(s|x_{trial})p(s)}{p(x_{trial})} \tag{2.1}$$

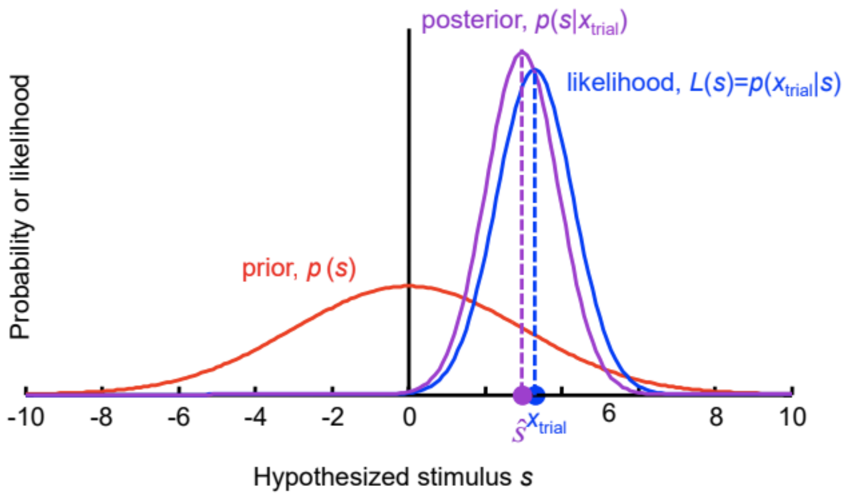


Figure 6. The Posterior Distribution Is Obtained by Multiplying the Prior by the Likelihood Function

Figure 2.1: Caption

3

Experimental design

Experimental goal In order to asses if this bayesian framework is valid we have to make controlled variations in the statistics of the environment and then compare the behavioral responses to a bayesian model.

3.1 Behavioural task

The behavioural task is a variable foreperiod paradigm with a temporal orienting paradigm. Participants heard two auditory cues separated by a foreperiod. There was to distinct set of cues presented. For this two set of cues, the tone were different and the foreperiods were drawn from distinct Gaussian distributions.

There were 4 block for each participants. In each of these blocks the latent distribution of the foreperiods were differents, for instance they were closer to each other.

In parallel to this behavioral paradigm we have to build a Bayesian observer that will try to mimic the behavior of the participant in this restraint environment.

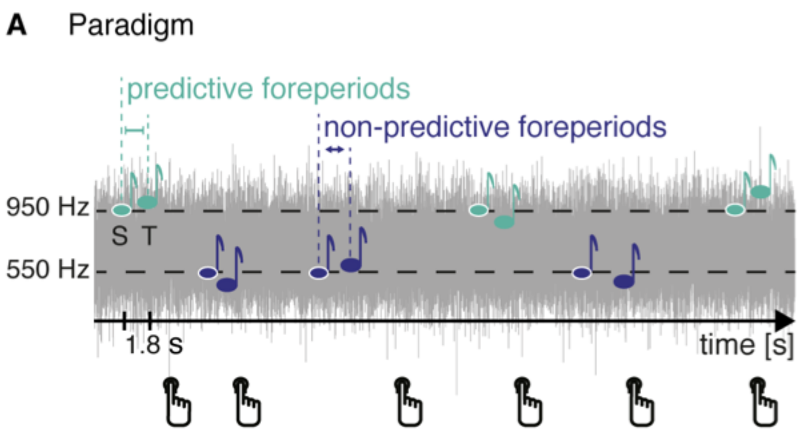


Figure 3.1: Presentation of tones

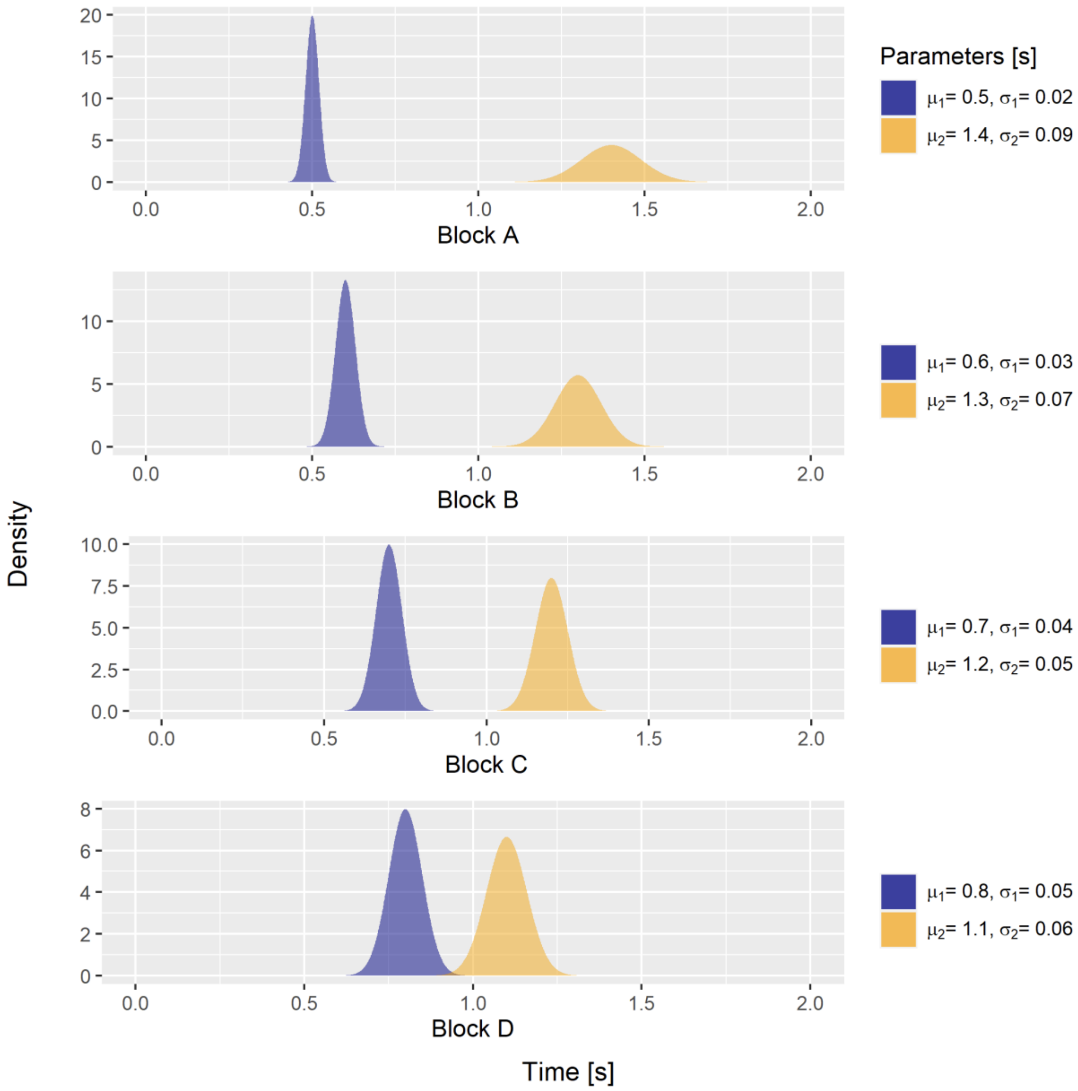


Figure 2.1: Graphic representation of the distributions used to generate the stimuli for the interval task for each block

Figure 3.2: Presented foreperiods

4 Building a bayesian observer model

4.1 The bayesian observer model

The Bayesian Observer is a theoretical observer that behave in a Bayesian way. By following, Bayes rule, he will, at each time step, thus before each foreperiod, estimate the underlying statistics of the foreperiods. As a consequence of the law of large numbers, the Bayesian observer will make the assumption that the latent distribution will follow a Gaussian distribution. Thus, he only have to infer the mean and standard deviation of the distribution.

Here we are confronted to a model where both parameters of the Gaussian distribution, the mean and standard variation, are unknown. In this case, this two quantities will follow for the Bayesian observer the following distribution.

- $\mu|\sigma^2 \sim \mathcal{N}(\mu_0 \sigma_0^2/\kappa_0)$ and
- $\sigma^2 \sim Inv\chi^2(\nu_0, \sigma_0^2)$

  where : σ_0^2 is a scaling factor, ν_0 the degree of freedom, κ_0 the number of measurements and μ_0 the prior mean.

4.2 Hybrid vs selective observer

In the experimental paradigm, there are two distribution for each of the tones, leading to a potential temporal orientation. However, the participant and thus the Bayesian observer can be not aware of this distinction. Therefore the participant could be not aware that there are two latent distribution for the foreperiods of the two tones, and as a consequence estimate only one latent distribution for the two tones. In this case, the observer is updating his belief on the underlying distribution at each foreperiod, thus making a mix, or hybrid, model. This behaviour will be called the hybrid observer. In this case the Bayesian observer will update only two values; the mean and the standard distribution.

On the other hand the participant can be modelled as selective observer who use the difference in the auditory tone in order to create two distinct latent distributions. In this case the Bayesian observer will have to update four different values; the mean and standard deviation of the foreperiods for the low tone, and the mean and standard deviation of the foreperiods for the high tone.

4.3 Informative vs uninformative priors

Once this behavior of hybrid or selective observer is fixed, the participant can have different types of prior. In the general theoretical framework of Gaussian statistics, there is two types of priors. The informative prior, which biases the parameters towards particular values and the non-informative prior, which doesn't influence the posterior hyper-paramters.

In the studied experimental paradigm, in the uninformative priors, we assume that the participant has no precise expectation for the foreperiod. Therefore, the prior distribution is model as a broad gaussian with the mean being the average of the defined time axis, and the standard deviation being the same as the mean.

On the other hand for the informative priors, we assume that there is no prior on the first observation and that updating starts after the second observation. The informative prior will be then a narrow Gaussian which mean is the foreperiod of the first observation and the standard deviation is a fixed small fraction of this first foreperiod, 10% of the foreperiod in order to follow Weber law.

Thus, we have four different types of models, the selective observer with uninformative priors, the hybrid observer with uninformative priors, the selective observer with informative priors and the hybrid observer with informative priors.

5

Finding the good model

5.1 Rational of the protocol

As it is mentioned in the previous chapter the aim of this study is to find the type of observer that is the most similar to human behavior. In order to do, it is necessary to find statistical indicators measuring the behaviour of the Bayesian observer and regress them of the actual behaviour of the participants.

5.2 The prediction error

One of the most strait forward indicator to measure the behaviour of the Bayesian observer is the prediction error. At the time step n, let's define p_{n-1}, the prediction of the foreperiod, and γ_n, the actual foreperiod presented. Then , the prediciton error is defined by :

$$err = p_{n-1} - \gamma_n \tag{5.1}$$

5.3 The hazard function

In this project, the hazard function models the effect of passing time on an observer's expectations : as time goes by, the expectation that an event will occur given it has not yet occurred will grow.

However, the time literature reveals that the monkeys' response times and the firing rate of neurons in the lateral intraparietal area were correlated with the hazard functions associated with the latent foreperiod distributions, which shows that temporal preparation involves both the prior knowledge about foreperiod duration and the elapse of time. These findings have been replicated in humans in both fMRI (Bueti et al., 2010) and EEG studies such as Herbst et al. (2018) which showed that the EEG signal obtained from three different foreperiod distributions was modulated by the associated hazard function.

More formally, the Hazard rate is the probability that an event will occur given that it has not yet occurred. It is defined by

$$h(FP_n) = P(FP_n > t + dt | FP_n > t) \tag{5.2}$$

Or equivalently :

$$h(FP_n) = \frac{f(FP_n}{1 - F(FP_n} \tag{5.3}$$

where f is the distribution and F the cumulative distribution.

5.4 The surprise

Shannon's information: Measure of surprise.

Intuition : big surprise when rare event (inversion), the surprise of two independent event is the sum of their surprise.

$$I(FP_n) = log(f(FP_n)) = -log(f(FP_n)) \tag{5.4}$$

Adding the hasard function inside

$$I_{HF}(FP_n) = -log(h(FP_n)) \tag{5.5}$$

where f is the distribution and F the cumulative distribution.

6 Model fitting

From this experimental paradigm, it is possible to have a first check of the response times. As expected they follow a skewed gaussian distribution. This skewness can be corrected by taking the logarithm of the RT.

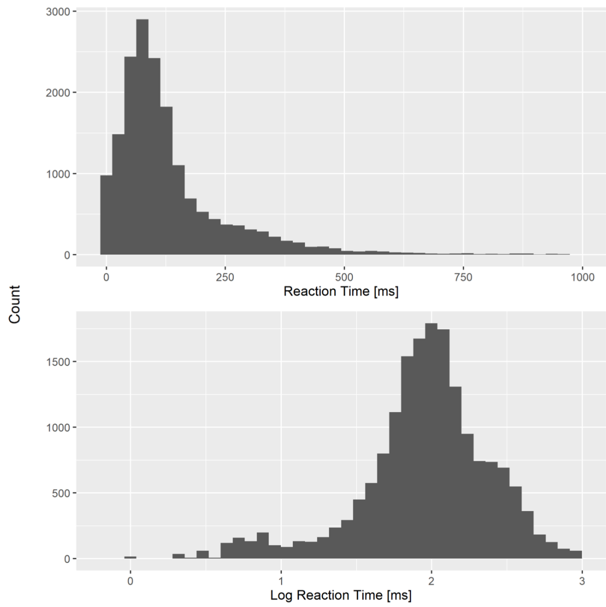


Figure 6.1: Presented foreperiods

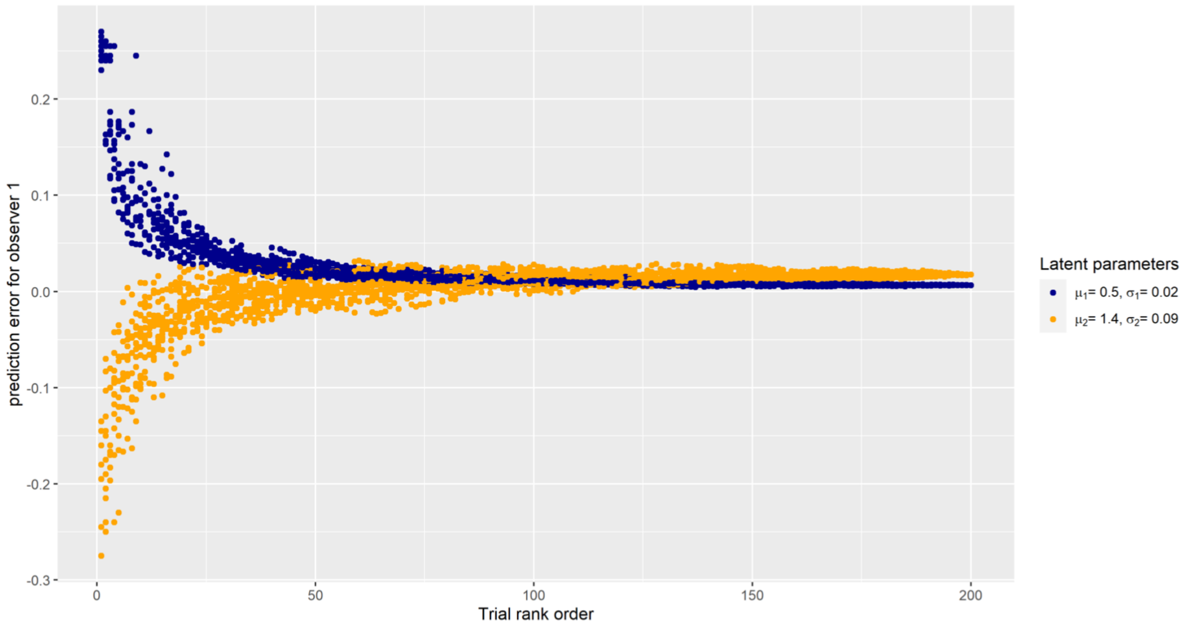


Figure 6.2: Evolution in the prediction error of the selective observer for block A.

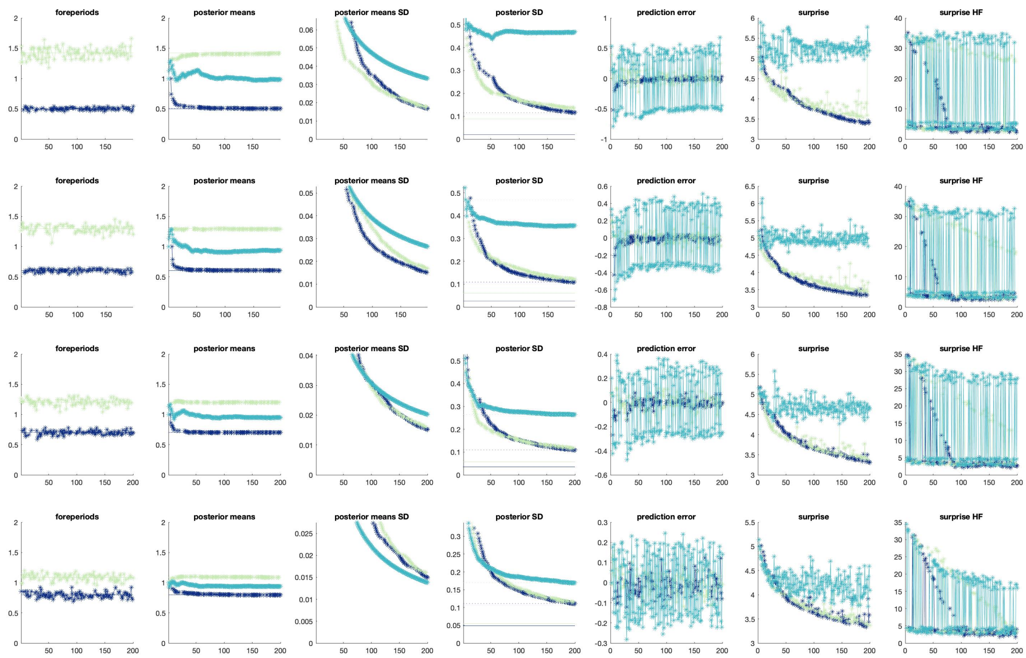


Figure 6.3: Evolution of the indicators for one participant during the four blocks

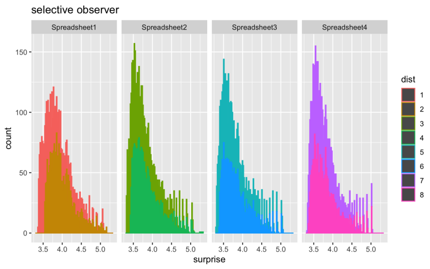


Figure 6.4: Surprise for selective observer

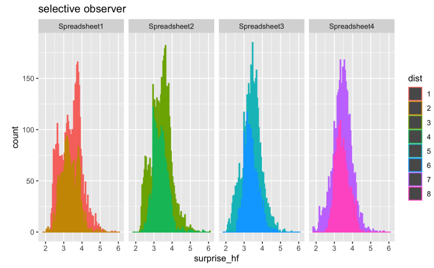


Figure 6.5: Surprise of hazard function for selective observer

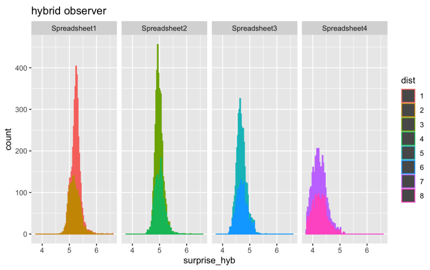


Figure 6.6: Surprise for hybrid observer

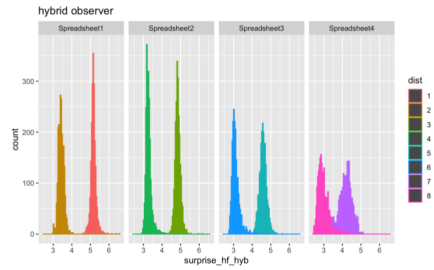


Figure 6.7: Surprise of hazard function for hybrid observer

Linear Mixed-Effects Models idea : Having various intercepts and slope for each group to control for inter-individual variability. Fixed effect and random effect.

Notation :

$$Y = \alpha + \beta X + (1 + X|participant) \tag{6.1}$$

How to select the fixed effect and the random effect variables ? Regression of the logarithm of the response time to normalize it.

- Begin with everything.

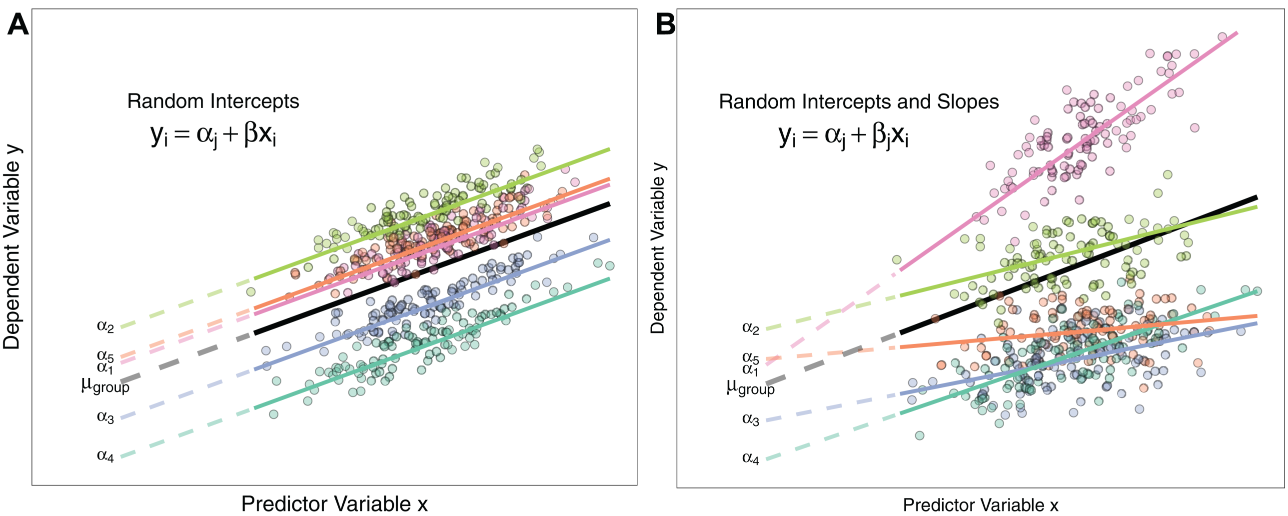


Figure 6.8: Different type of LMM

$log_{10}(RT) = FP + \sum_i surprise_i + \sum_j PredictionError_j + trial + TrialPerDist + PreviousRT + (1|participant_{id})$;

$i \in (select, select_h f, hyb, hyb_h f)$ and $j(selective, hybrid)$

- Do an iterative algorithm to select only the relevant fixed effect variables. →Step-wise regression.
- Use the found variables as random effect also.

Final Model **Final model** with as much fixed effect as random effect variables. $log_{10}(RT) = FP + surprise + \sum_j PredictionError_j + trial + TrialPerDist + PreviousRT + (FP + surprise + \sum_j PredictionError_j + trial + TrialPerDist + Previous|participant_{id})$

7 Informative vs Uninformative priors

Informative vs Uninformative priors

- Two different final models with different random effect and fixed effect.
- Two different results (R square).
- How to choose the good one ?

Akaike Information Criterion

- Trade-off between having a good result and minimizing the number of parameters.
- $$AIC = 2k - 2ln(L) \tag{7.1}$$

  where k is the number of regressors and L the likelihhod.
- The best model is the one with the lowest AIC.
- AIC for informative prior : 13101
- AIC for uninformative prior : 13093

$\rightarrow$The model with uniformative priors seems to be the best (not a big difference 0.06 percent)

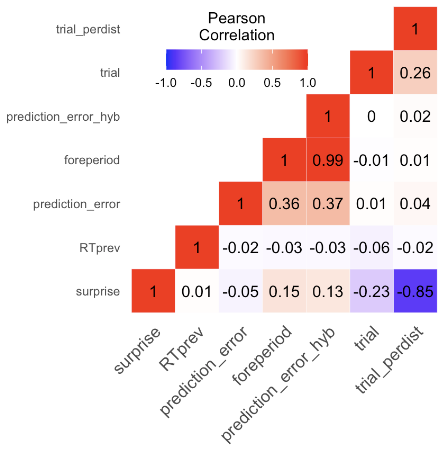


Figure 7.1: Correlation of the variables selected for the informative priors

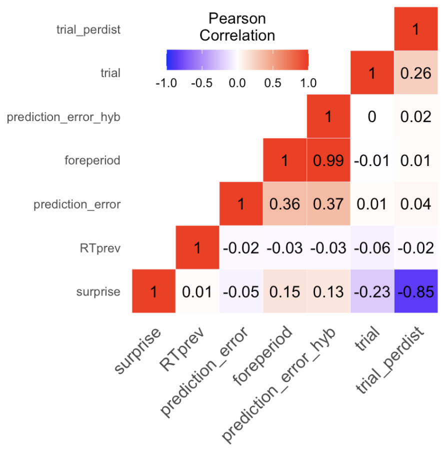


Figure 7.2: Correlation of the variables selected for the uninformative priors

- Only the surprise of the selective observer.
- Only the prediction error the hybrid observer.

8

Conclusion and future work

Methodological summary

- Artificially create variation in the statistical properties of the environment and feed the Bayesian observer the stimulus.
- The Bayesian observer can have two types of prior : informative and uninformative.
- The Bayesian observer can be : hybrid (one distribution per tone), or hybrid (only one common distribution).
- Compute indicator (prediction error, surprise, hazard rate), and regress behavioral data on it.
- Use statistical tools to determine the best model.

A non conclusive study Interpretation Consistently with results from O'Reilly et al. (2013); Visalli et al. (2019), behavioral results have shown an effect of surprise measures on participant's reaction times. This could be interpreted as a reflection of the behavioral cost of surprise since reaction times can be considered an index of information encoding and cognitive effort. This would be consistent with an information theoretic account of cognition ; unexpected events are require more effort to encode, which should incur a behavioral cost in terms of reaction times. Furthermore, the fact that model-based measures of surprise have a significant effect of partici- pant's reaction times implies that event considered surprising by participants were also considered 'surprising' by the Bayesian Observer to some extent. This would tend to show that the Bayesian Observer can provide reliable estimates of the participant's temporal expectations.

- Only the surprise of the selective observer.
- Only the prediction error of the hybrid observer.

- The model with uniformative priors seems to be the best ($\Delta AIC = 0.06\%$).
- The informative model make more sens psychologically.

Future work

- Adding volatility in the model.
- Try to see if the the behavior of the participant change from hybrid to selective during the experiment.
- try other type of variable selection (Lasso regression for instance).
- Addition of a Q-Learning algorithm (model free Reinforcement Learning).
- Use of eye trackers in a behavioral setup in order to have biological markers (good tracker of noradrenaline)
- ... but eye reaction might be to slow for the paradigm.

Bibliography

[1] E. Rudolph, P.Graubmann, and J. Grabowski. Tutorial on message sequence charts (MSC'96). In *FORTE*, 1996.

[2] ITU-TS. ITU-TS Recommendation Z.120: Message Sequence Chart (MSC). Geneva, 1997.

[3] Michel Goossens, Sebastian Rahtz, and Frank Mittelbach. *The LaTeX Graphics Companion*. Addison-Wesley, 1997.

[4] L. Lamport. *LaTeX—A Document Preparation System—User's Guide and Reference Manual*. Adsison-Wesley, 2nd edition, 1994. Updated for LaTeX 2_ε.

[5] V. Bos and S. Mauw. *A LaTeX macro package for Message Sequence Charts—Reference Manual—Describing version*, June 2002. Included in MSC macro package distribution.

[6] V. Bos and S. Mauw. *A LaTeX macro package for Message Sequence Charts—User Manual—Describing version*, June 2002. Included in MSC macro package distribution.

[7] V. Bos and S. Mauw. *A LaTeX macro package for Message Sequence Charts—Maintenance document—Describing version*, June 2002. Included in MSC macro package distribution.