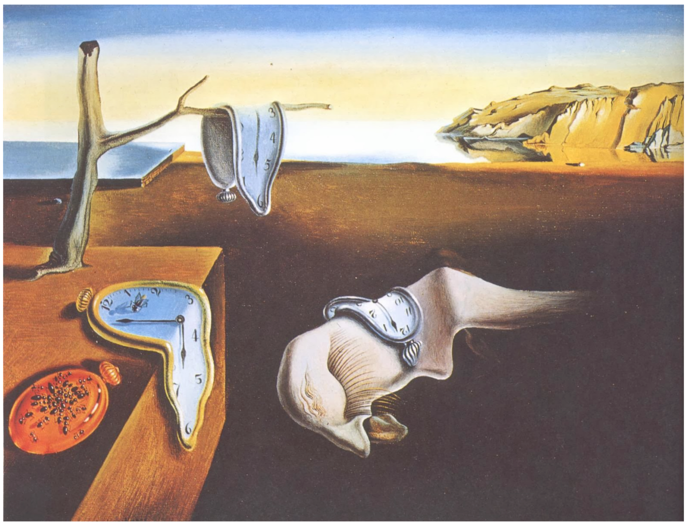
Master Computational Neurosciences and Neuroengineering

Supervised Project : Cognition and Brain Dynamics, Neurospin

Bayesian Modelling of the unfolding of time prediction

Mathias Vigouroux

Supervisor : Dr. Sophie Herbst



2021

Contents

1

Abstract

The timing of forthcoming events is predicted by the brain in order to guide our behaviour. Indeed, integrating a temporal structure in our internal representation of the environment, which is build from our multiple sensory inputs, optimize our actions and clarify our perception.[1], [2],[3] However, it still remains unclear how this temporal structure is learn. The Bayesian framework is widely used to model the way we build inferences from our sensors [4], and is gaining popularity in the time literature [5]. The goal of my supervised project is therefore to use behavioural pilot data collected previously [6] and continue the analysis by comparing the behavioural result to a simulated agent : the Bayesian Observer, who predict following Bayes' rule. The rational of the protocol is thus to compare the behavioural response of the participant, engaged in an implicit timing task, to the predictions of the Bayesian Observer trough a linear mixed-effects model. Moreover, different types of Bayesian Observer can be used, reflecting different prediction strategies of the participant. This first study show that derived quantities from the prediction of the Bayesian Observer are significant to model the participants' behaviour, giving good hope for further studies. However, we were still not able to conclude which type of Bayesian Observer should be kept.

2 Introduction

2.1 Context

With space, time is one of the fundamental characteristic of our dynamical environment. 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 implicit timing actions. However, extracting temporal predictions from our sensory environment optimize our behavior. For instance, this will allow us to catch something in the air or cross a busy road [2]. In time literature, this ability is called it is called temporal preparation [3].

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

2.2 Goal of the supervised project

A previous online experiment conducted by Elisa Lannelongue [6], former intern in the team, allowed to collect and analyse pilot data on temporal preparation. The goal of this supervised project is therefore to

- 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 is to really understand what have been done by the team and where can some improvements be made in a second part.

3

Theoretical framework : time perception and Bayesian statistics

3.1 The foreperiod

3.1.1 The fixed foreperiod paradigm

In cognitive neuroscience and psychology, the prevalent experimental setup to study temporal preparation is the "foreperiod paradigm", described in the 80's by [7]. This paradigm is schematized in A.1.

In this experimental setup, the foreperiod refers to a time interval between a cue stimulus and a target stimulus. After this target stimulus has appeared, the participant will have to respond, for instance by pressing the keyboard. This second time interval between the target cue and the response of the participant is what is refer to the Response Time (RT). A fixed foreperiod paradigm is therefore an experimental design where the foreperiod is of a fixed duration during each block.

In fixed foreperiod paradigm, it has been showed that the response time increase with the duration of the foreperiod modifiable between each block. This effect is called the "fixed foreperiod effect" [8].

3.1.2 The variable foreperiod paradigm

However, the foreperiod, that was considered fixed until now, can also be of variable duration. In that case, the experimental design is referred as the "variable foreperiod paradigm". In this paradigm, the behaviour of the participants will be explained by the hazard rate (see following chapter)[3] . Indeed, if the stimulus wasn't presented after a short duration, then the participants is likely to predict that it will be the long foreperiod. Thus, they will be more prepared to respond to the this long foreperiod and will have a shorter response time [7] .

3.1.3 The temporal orienting paradigm

Finally, an other type of foreperiod paradigm is the temporal orienting task. In this setup, a cue will orient the participant toward a specific foreperiod. For instance, the tone of the auditory cue can be different for different foreperiods. With this additional information, a shorter foreperiod will elicits faster and more accurate responses to validly-cued targets as compared to targets occurring at an unexpected moment [2].

A good modelling of the behavioural response will go beyond a description of the data, similar to a curve fitting. Good models help to break down perceptual, cognitive, or motor processes into interpretable and generalizable stages. Across domains of application, Bayesian models of decision making are based on the same small set of principles, thereby promising high interpretability and generalizability. Bayesian models aspire to account for an organism's decision process when the task-relevant states of the world are not exactly known to the organism. [4].

3.2 Bayesian inference

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

Before the experiment has even started you will make some assumption on this duration. Indeed, you know that it won't last a hundred years, as dead people rarely report an answer for this question, and that it will be above nanoseconds, because otherwise it will be impossible 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 25 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 begins, you count in your head and you receive the second stimulus 32 seconds after the first one. However, you have already done some probability and know that you don't want to draw conclusions too quickly. Thus, you decide to revise your expectations, i.e. you prior distribution, by trying to asses what is the likelihood of measuring a duration of 32 seconds 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 $p(x_{trial}|s)$, and then give an important weight to your measure, i.e. 32 seconds.

However, you don't want to completely forgot what you have heard in the waiting room, and keep some importance to the 25 seconds assumptions. To that 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. In a nutshell, the posterior distribution is obtained by multiplying the prior distribution with the likelihood, schematized in A.4.

After this first presentation of the stimulus, you expect to have the second foreperiod nearer to 32 seconds. Then after, the second foreperiod and you will repeat this upgrading in order to predict the third foreperiod and so on.

More formally, the Bayes' rule is defined by :

$$p(x_{trial}|s) = \frac{p(s|x_{trial})p(s)}{p(x_{trial})} \tag{3.1}$$

4

Confronting the participant behaviour to a Bayesian observer

4.1 Rational of the protocol

In order to asses if the 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. The Bayesian model is therefore the optimal learner of the temporal statistics.

This section will therefore present in a first part the behavioural experiment and develop the Bayesian Observer model.

4.2 Behavioural task

The behavioural task is a combination of both a variable foreperiod paradigm and a temporal orienting paradigm. More precisely, participants heard two auditory cues separated by a foreperiod. An auditory cue was selected because it has been shown that implicit temporal predictability enhances pitch discrimination sensitivity [1]. There were two distinct set of cues presented. For this two set of cues, the tone were different and the foreperiods were drawn from distinct Gaussian distributions.

However, participants were not aware that there were two distinct latent distributions.

Participants were only asked to respond, by pressing the keyboard, as quickly as possible to the target tone.

The paradigm is schematized in 4.1.

There were 4 block for each participants, in which the parameters of the latent Gaussian distribution, the mean and the standard deviation, varied A.2. For instance, the two distribution can be closer to each other, and more or less widespread.

4.3 The response time

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 A.3.

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.

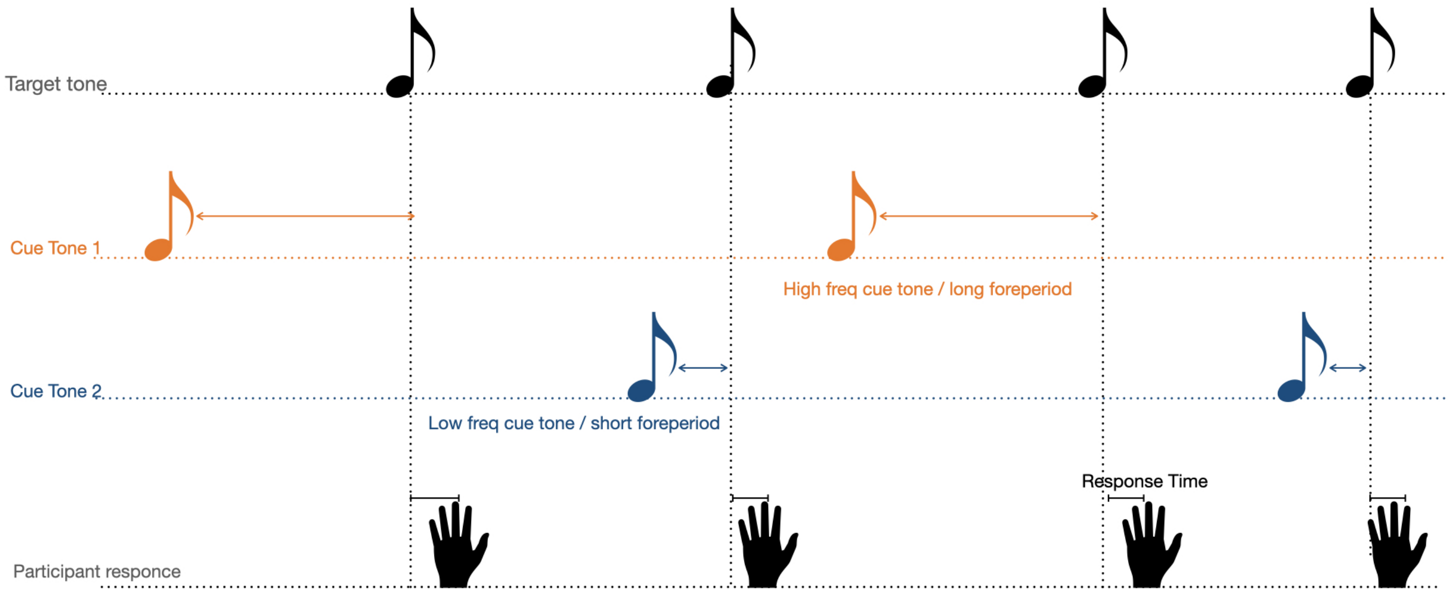


Figure 4.1: Presentation of tones

4.4 The Bayesian observer model

The Bayesian Observer is a theoretical observer that behaves in a Bayesian way. By following Bayes rule, it 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, it only have to infer the mean and standard deviation of the distribution. The mean and standard deviation of the two latent distribution have been chosen in order to be easily estimated by the Gaussian observer [6].

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.[9]

- $\mu|\sigma^2 \sim \mathcal{N}(\mu_0 \sigma_0^2/\kappa_0)$ 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.

4.5 Hybrid vs selective observer

In the experimental paradigm, there is one distribution for each of the tones, leading to a potential temporal orientation. However, the participant and thus the Bayesian observer can be aware or not 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 of the mixed 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, i.e. has a more accurate representation of the underlying temporal statistics. 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.

4.6 Measure of predictions.

On one hand, from this statistical estimation of the latent distribution the Bayesian observer will compute predictions of the foreperiod trial-by-trial. On the other hand, participants, who are asked to answer as quickly as possible to the target tone, will also try to predict, consciously or unconsciously, the foreperiod in order to be ready as much as possible to answer to cue tone (temporal preparation). As a consequence, the prediction of the participants will influence their response time. However, it is impossible to know the prediction of the participant.

The idea is thus to use the prediction of the Bayesian observer as a substitute of the participants predictions.

4.6.1 The prediction error

One of the most strait forward indicator to asses the prediction 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 prediction error is defined by :

$$err = p_{n-1} - \gamma_n \tag{4.1}$$

4.6.2 The surprise

On top of this first measure, the surprise of the event can also be relevant. They key assumption, being that, if the surprise of an event is big, i.e. a foreperiod that was longer or shorter than expected, then the participant isn't prepared enough for this event and take make more time to answer the target tone.

The surprise for the Bayesian observer can be measured with the Shannon's information. This quantity is defined as

$$I(FP_n) = log(f(FP_n)) = -log(f(FP_n)) \tag{4.2}$$

where f is the distribution of the foreperiod.

The aim of this quantity is to grow inversely to the probability of an event, and to add the surprise of two independent variables, hense the logarithm.

In time literature, [5] showed that in visual stimulation paradigm, the model-based surprise variable predicted trial-by-trial variations in reaction time more strongly than the externally observable interval timings alone.

4.7 The hazard function

Until now, the described statistical tools didn't took into account the fundamental asymmetry of time. Indeed, if we know that an event will occur and hasn't occurred yet, we have more certainty that it will occur in the future. However, the previously defined surprise function don't take this dynamical property into account, it consider as surprising as very short foreperiod than a very long foreperiod. Yet participants are likely to be less surprised by the extremely long foreperiod as they know that it won't be the very short foreperiod quite quickly.

4.7.1 The role of hazard function

Hazard function is therefore trying to encapsulate the effect of time passing on. Time literature reveals that response time is correlated with the hazard functions of the latent foreperiod distributions, showing that temporal preparation involves both the prior knowledge about foreperiod duration and the elapse of time. Moreover, human physiological studies [10] show that the EEG signals are modulated by the hazard function in a foreperiod paradigm.

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

$$h(FP_n) = P(FP_n > t + dt | FP_n > t) \tag{4.3}$$

Or equivalently :

$$h(FP_n) = \frac{f(FP_n)}{1 - F(FP_n)} \tag{4.4}$$

where f is the distribution and F the cumulative distribution.

4.7.2 Hazard function and surprise

The hazard function is a probability distribution, thus once can compute the surprise deriving from it. Therefore it is possible to define the hazard function surprise as :

$$I_{HF}(FP_n) = -log(h(FP_n)) \tag{4.5}$$

where f is the distribution and F the cumulative distribution.

5 Model fitting

Let's recall that the goal of the experiment is to try to find a Bayesian model that best explain the participant's behaviour.

5.1 Assessing the participant behaviour

The only behavioural measure that we have is the Response Time (RT) of the participant. In classic time literature [7], it has been shown that the RT depends on the duration of the foreperiod. However, it is the logarithm of the response time that follow a normal distribution. Ensuring that our variable of interest follows a normal distribution facilitate the regression and diminish statistical biases in its analysis.

Moreover, the RT of the participant will likely be correlated to his previous response time (auto-regressive model). Indeed, if the participants responded quickly to the target tone is it quite likely that they are focused and thus will answer quickly to the new foreperiod.

Moreover, participants can become more tired during the experiment, and thus increase their RT. However, one can argue that along the experiment the participant will get more used to the task and thus decrease his RT during the experiment. As a consequence, it is very likely that the trial number will have an effect on the RT.

On the same note, if the participants haven't heard a lot one particular tone, they will be less used to the duration of the foreperiod for this particular tone. Therefore, the number of trial per distribution, which depends on the block and the tone, is likely to have an effect on the RT.

5.2 Assessing the Bayesian observer prediction

The main novelty of this study is to add regressors from the Bayesian observer in addition to usual behavioural regressors cited in the previous paragraph.

The two types of Bayesian observer will estimate the mean and standard deviation of the foreperiods A.5. From this statistical estimation they will predict the foreperiods. From this predictions, it is possible to derived the quantities introduced in the previous chapter. A.6.

What we can clearly see is that there are some difference between the two types of Bayesian observer. Therefore it seems impossible to do a univariate study in order to asses which type of observer is the most relevant one, and a multivariate analysis is needed.

5.3 Linear Mixed-Effects Models

However, the multivariate analysis cannot be "just" a simple multiple regression because there isn't enough participant in this pilot data (only 32).

Idea : Having various intercepts and slope for each participant to control for inter-individual variability.

Notation : Let's denote the variable of interest Y and the regressor X. The the mixed-effects model is given by :

$$Y = \alpha + \beta X + (1 + X|participant) \tag{5.1}$$

The first part of the equation, $\alpha + \beta X$ is called the fixed effect, as it fixed for each of the participants. The second term, $(1 + X|participant)$ is the random effect, which is an effect that varies "randomly" for each of the participant A.7.

5.3.1 How to select the fixed effect and the random effect variables ?

Choosing the fixed and random effect is not an easy task. However, in order to have the more precise model and to take into account inter-participant variability, it is commonly suggested to put all the fixed effect variable as a random effect variable. However, in the case of this experiment there are too many regressor and not enough participants to make such an analysis.

Thus, it was decided to first model the participant reaction time with all regressor as fixed effect and only use the intercept as a random effect. Therefore the first estimated equation is :

$$\log_{10}(RT) = ForePeriod + \sum_i Surprise_i + \sum_j PredictionError_j + Trial + TrialPerDist + PreviousRT + (1|P_{id}) \tag{5.2}$$

where i can be (selective,selective-hasard-function,hybrid,hybrid-hasard-function) and j can be (selective, hybrid).

This first regression was then updated by an iterative algorithm, the step-wise regression. The aim of this algorithm is to suppress the regressor from the previous equation until there is no more statistically significant regressors left. After this step, the estimated regression have less fixed effect regressors than previously.

Once these fixed effects have been selected, they are also uploaded as random effect.

5.4 Final Model

As a consequence, the final model will be a linear mixed-effects regression with as much fixed effect as random effect variables.

In the case of the uninformative prior (see following section), the final model will be given by :

$$log_{10}(RT) = FP+Surp_{select}+\sum_j PrE_j+Tr+TPD+PRT+(FP+Surp_{select}+\sum_j PE_j+Tr+TPD+PRT|participant_{id}) \quad (5.3)$$

with $j \in (selective, hybrid)$

6

Informative vs Uninformative priors

6.1 Informative vs uninformative priors

In this Gaussian framework, the participant, and thus the Bayesian observer start with a prior probability. There are 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 hyperparamters.

In the studied experimental paradigm, computing the uninformative priors implies 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, the informative priors implies 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.

6.2 Why choose the type of prior ?

First, this two type types of prior reflect two different types of behaviour. Naturally,the regressors kept by the final model for this two types of prior are not identical. However, in both of the paradigm variable from the Bayesian model remain statistically significant A.8, A.9.

Second, it is not possible to have the prediction from the uninformative and informative prior in the same equation. Indeed, the participants will only choose once, at the beginning, if they have an informative or uninformative prior. Moreover, because it is only the prior that is different between these two models, the further predictions of the Bayesian observers with different type of prior are likely to be highly correlated. This would led to a spurious regression.

6.3 Akaike Information Criterion

As a consequence, an other criterion for the two final linear mixed-effects models has to be implemented.

A common criterion for comparing two model for which we know the likelihood is the Akaike Information Criterion (AIC). This criterion tries to capture a trade-off between the number of parameters, that one wants to keep as low as possible in order to avoid overfitting, and the likelihood, thus the precision, of the model.

Thus, the AIC can be defined as

$$AIC = 2k - 2ln(L) \tag{6.1}$$

where k is the number of regressors and L the likelihhod.

6.4 Comparing the informative and uniformative priors

The best model is the one with the lowest AIC. Computing them for the two types of prior gives :

- AIC for informative prior : 13101
- AIC for uninformative prior : 13093

As a consequence, the model with uniformative priors seems to be the best one, as it has the lowest AIC. However, this difference is relatively very small as it only correspond to 0.06% of relative difference.

7 Conclusion and future work

7.1 Methodological summary

The methodological study that emerge from Elisa's previous work is quite dense. Yet, it can be divided in :

- Artificially create variation in the statistical properties of the environment and feed the Bayesian observer the stimulus. The participant is not aware about the statistically properties of the environment and so does the Bayesian observer.
- The Bayesian observer can be : selective (one distribution per tone), or hybrid (only one common distribution).
- The Bayesian observer can have two types of prior : informative and uninformative.
- Compute measure of the prediction of the Bayesian observer (prediction error, surprise, hazard rate surprise), and regress the Response Time of the participant on it.
- Use statistical analysis to determine the best model.

7.2 Significant effect of the Bayesian observer and effect of surprise

In line with the results from [8], the measure of surprise has an effect on the reaction time. Finally, this first result reveals that an information theory framework is relevant to study cognition.

Moreover, the effect of the surprise on the Reaction Time is positive. Thus, it shows that the surprising event for the Bayesian observer will also surprising event for the participant, and are harder to encode for the participant who therefore has a longer reaction time A.10.

7.3 Limits of the study

The previous section, show that the Bayesian observer and the information theory framework that it rely upon is relevant. However, it is impossible for the moment to asses which type of Bayesian observer fit the best the behaviour.

First, the updating behavior of the Bayesian observer, i.e. either selective or hybrid, cannot be assessed. Indeed, only the surprise, or the hazard function surprise, for the selective observer remain. However, the prediction error for both type of learning behaviour, i.e. the prediction error for the selective observer and the prediction error for the hybrid observer.

Second, the good type of prior, i.e. informative or uniformative prior, can not be assess neither. Indeed, the difference of the Akaike Information Criterion for the two different prior only varies by $\Delta AIC = 0.06\%$. Moreover this information criterion is favoring the uninformative prior. However, it is the informative prior that makes more sens on a psychological standpoint.

7.4 Future work

7.4.1 Improvement in the experimental paradigm

Clearly, the first improvement that has to be made is to increase the number of data points in order to gain some statistical power in the analysis.

The previous experiment also missed data to study the underlying biological process that might responsible for the time preparation. A recent study [11], showed that in Bayesian observer model, the pupil size is correlated with surprise and confidence. Moreover, pupil size is a good indicator of the level of neuromodulators, for instance noradrenaline. However, the temporal dynamics of pupil size might be too lo to be useful in this experimental paradigm.

7.4.2 Improving the Bayesian observer

On a theoretical standpoint, in the current model, the priors of the Bayesian model are reset at the beginning of each of the four blocks presented to the participant. Such behaviour is not very likely to occur in human subjects who might use their knowledge from the previous block.Therefore, adding a volatility term can be added to control this effect.

A change in the behaviour of the participant could be also reflected by the evolution from a hybrid observer at the beginning of the experiment, when the participant hasn't inferred yet that there are two underlying foreperiod distribution for each tone, to a selective observer for the latter trials.

7.5 Using other type of statistical algorithm to find the best model

The step-wise regression was the only algorithm that tried to suppress regressors. It would be also interesting to use other type of penalized regression, for instance the Lasso or Ridge regression, to keep as few regressor as possible. Indeed, by augmenting the penalization parameters, it would be possible to see for how long the Bayesian observer predictions still remain as regressors.

Finally, the study [12] showed that a Q-learning reinforcement learning algorithm can be used to better fit the behavioral data in a dual-learning task. The Q-learning reinforcement learning algorithm is a model free reinforcement learning algorithm that can be therefore used at the beginning of the experiment when the participant still doesn't have a clear idea, in other words a model of the environment.

A
Figures

List of Figures

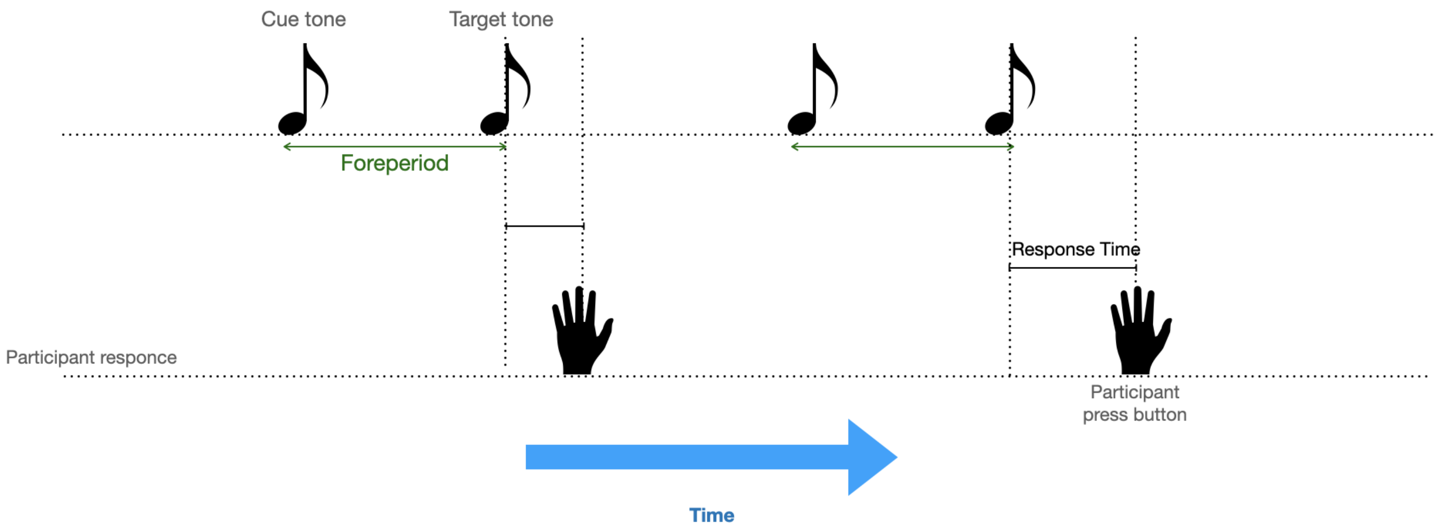


Figure A.1: Distinction between the foreperiod and the response time.

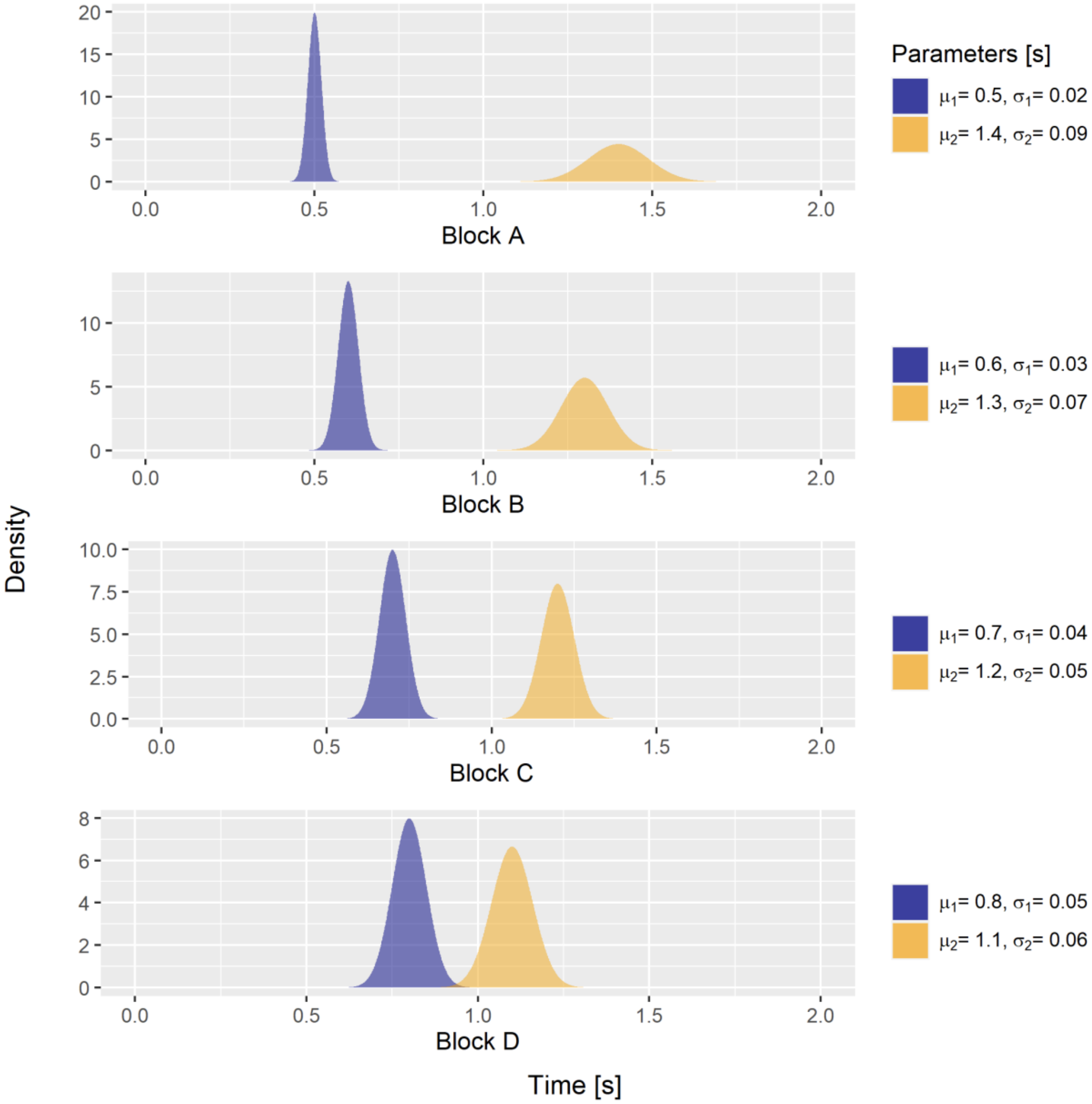


Figure A.2: Presented foreperiods

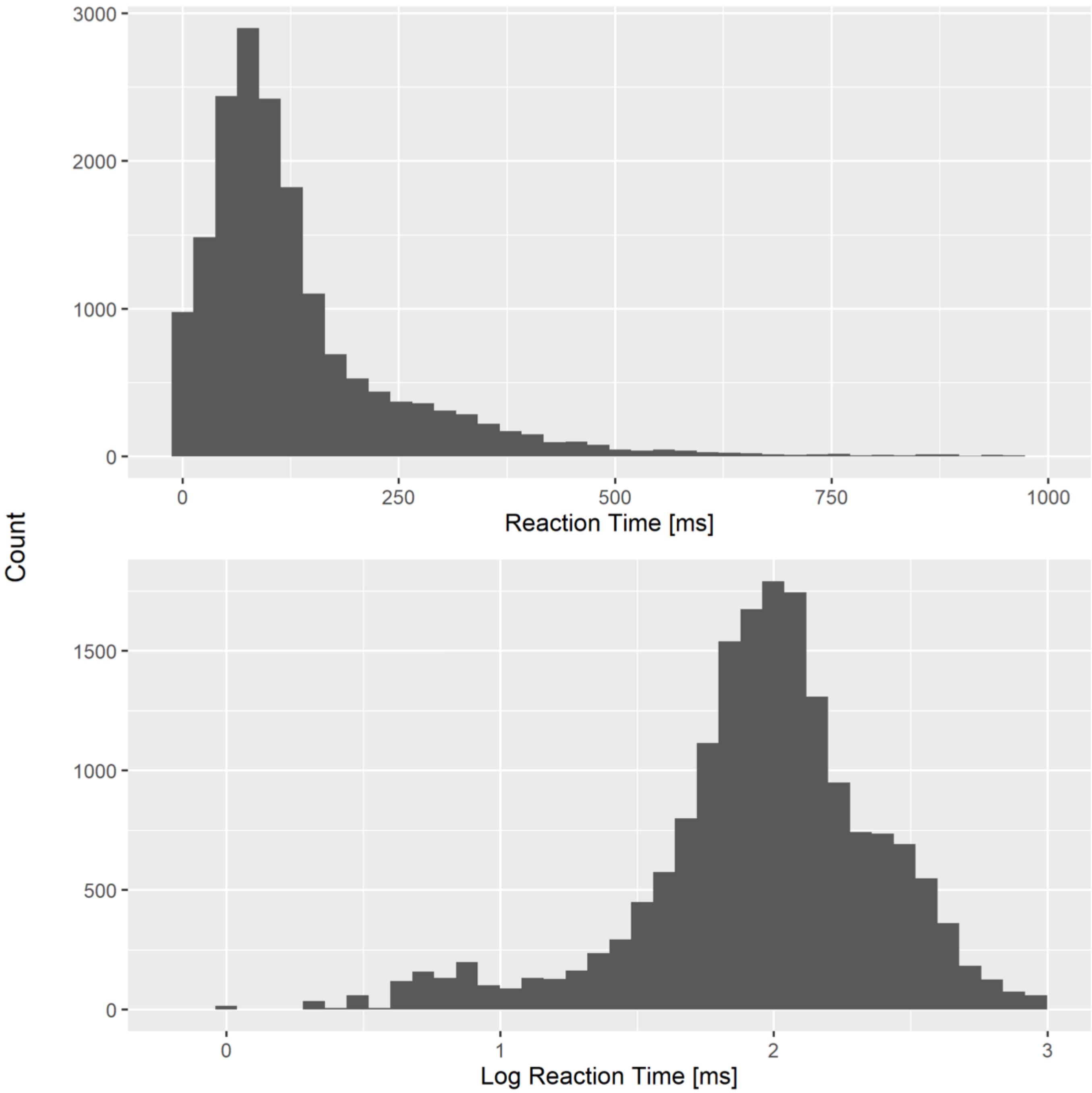


Figure A.3: Response times of the participants

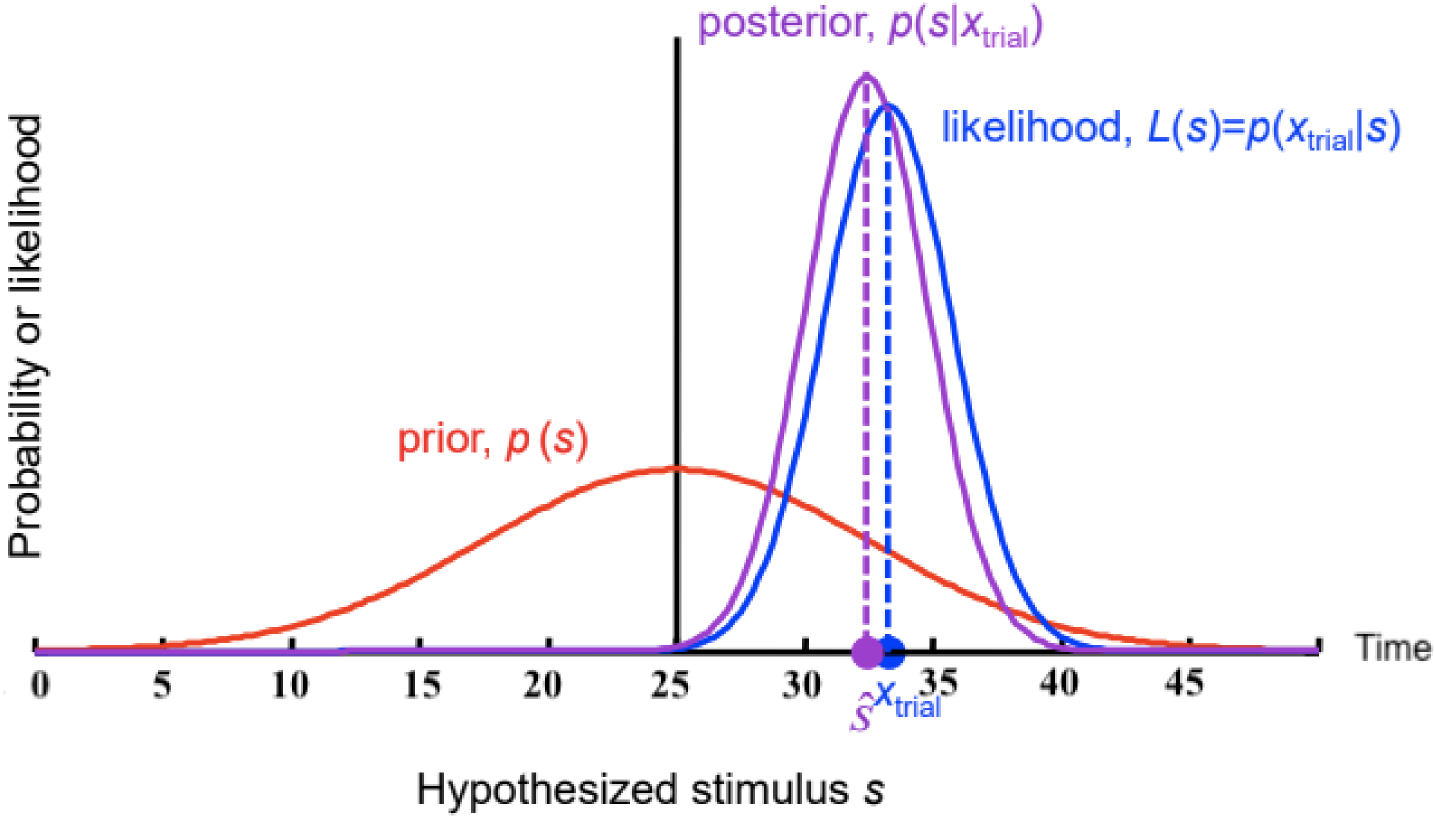


Figure A.4: Multiplying the Prior Distribution with the Likelihood, give the Posterior Distribution.

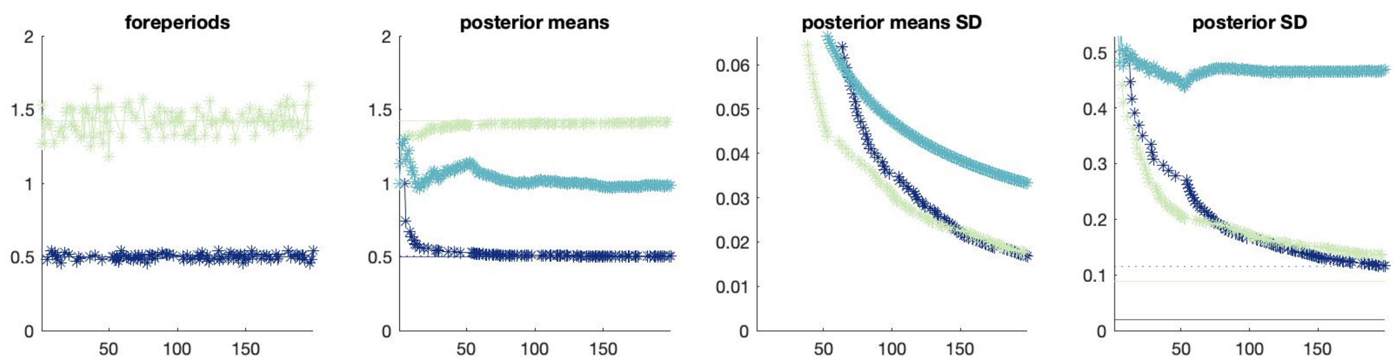


Figure A.5: Differences in estimations. Light blue : hybrid observer ; green and dark blue : selective observer

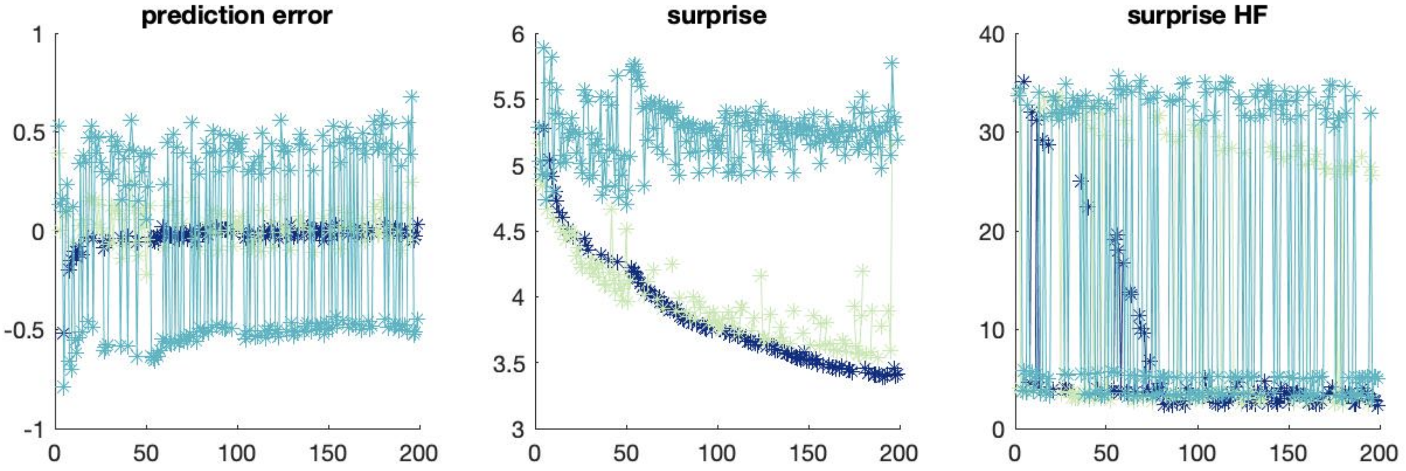


Figure A.6: Behaviour of the two types of observer

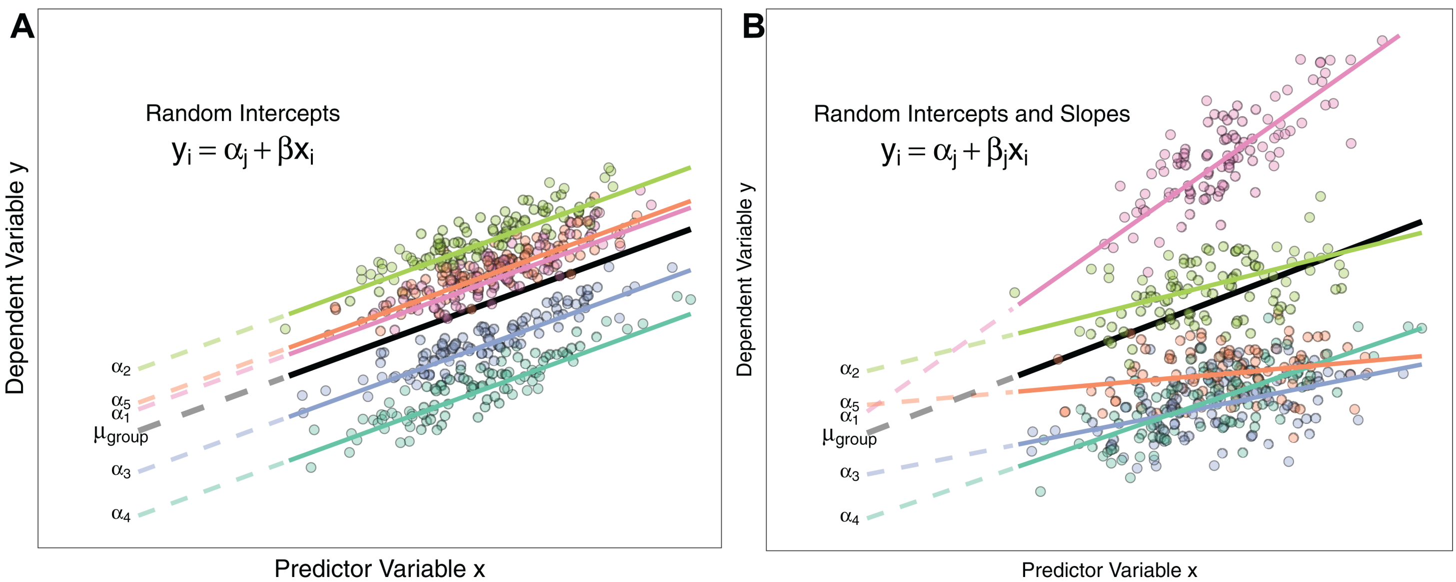


Figure A.7: Different type of LMM

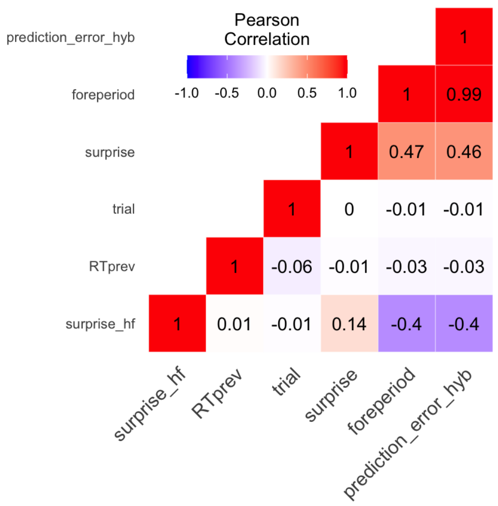


Figure A.8: Correlation of the variables selected for the informative priors

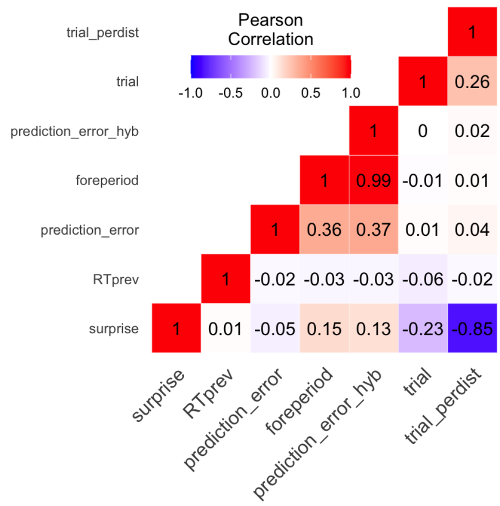


Figure A.9: Correlation of the variables selected for the uninformative priors

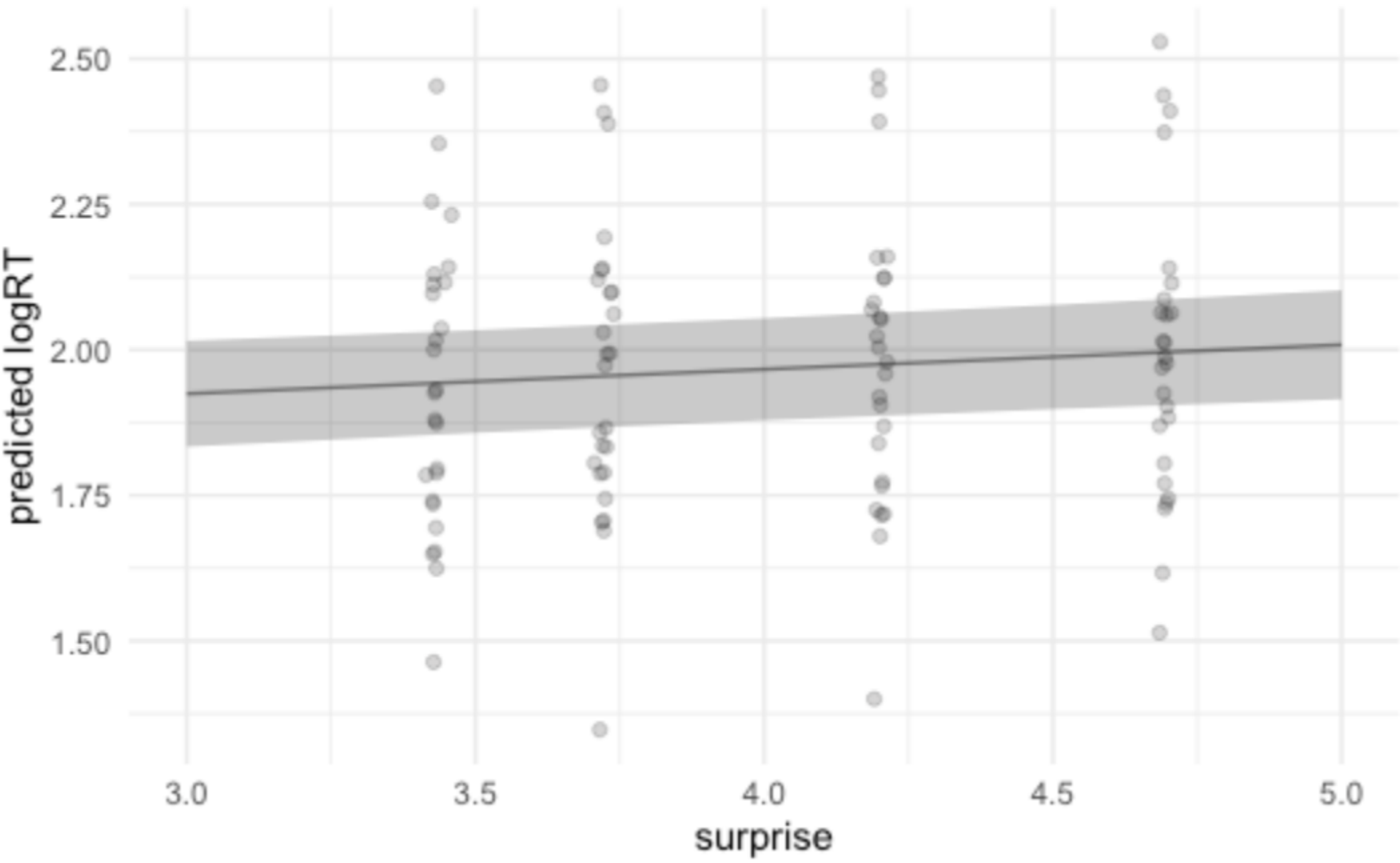


Figure A.10: Surprise has a positive effect on the RT if we use the model

Bibliography

[1] Obleser J. Herbst, S. K. Implicit temporal predictability enhances pitch discrimination sensitivity and biases the phase of delta oscillations in auditory cortex. *NeuroImage, 203, 116198.*, 2019.

[2] Correa A. Enhancing behavioural performance by visual temporal orienting. *Attention and time, pages 357–370.*, 2010.

[3] A. C. Nobre and F. van Ede. Anticipated moments: temporal structure in attention. *Nature Reviews Neuroscience, 19(1):34–48.*, 2018.

[4] Wei Ji Ma. Bayesian decision models : A primer. *Cell Press*, 2019.

[5] Kloosterman N. A. Engel A. K. Wagenmakers E.-J. Meindertsma, T. and T. H. Donner. Surprise about sensory event timing drives cortical transients in the beta frequency band. *Journal of Neuroscience, 38(35):7600–7610.*, 2018.

[6] Elisa Lannelongue. Bayesian modeling of temporal predictions. *Master Thesis, Cogmaster*, 2021.

[7] P. Niemi and R Näätänen. Foreperiod and simple reaction time. *Psychological bulletin, 89(1):133.*, 1981.

[8] Ambrosini E Mazzonetto I Vallesi A. Visalli A, Capizzi M. Bayesian modeling of temporal expectations in the human brain. *Neuroimage, 202, article 116097*, 2019.

[9] Carlin J. B. Stern H. S. Dunson D. B. Vehtari A. Gelman, A. and D. B. Rubin. *Bayesian data analysis.* CRC press, 2013.

[10] Fiedler L. Herbst, S. K. and J. Obleser. Tracking temporal hazard in the human electroencephalogram using a forward encoding model. *eneuro, 5(2):ENEURO.0017–18.2018.*, 2018.

[11] Meyniel F. Brain dynamics for confidence-weighted learning. *PloS Comput Biol.*, 2020.

[12] Khamassi M. Brovelli A. Viejo, G. and B Girard. Modeling choice and reaction time during arbitrary visuomotor learning through the coordination of adaptive working memory and reinforcement learning. *Frontiers in Behavioral Neuroscience, 9*, 2015.