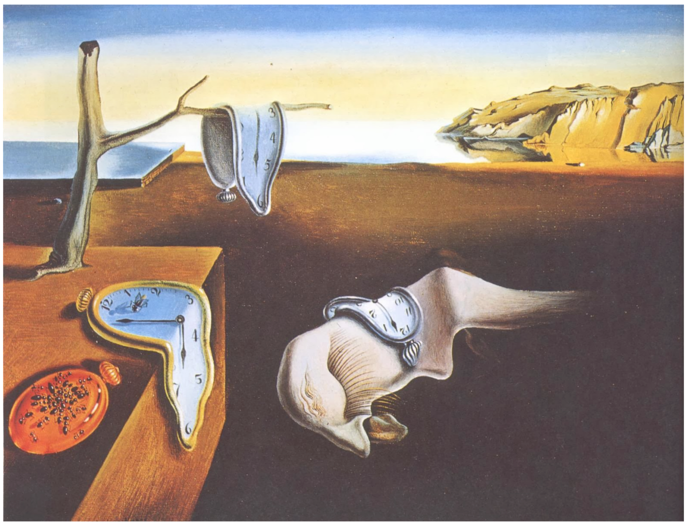
ENSAE, IP Paris

Stage d'Application at the Cognition and Brain Dynamics Team, Unicog, Neurospin, France

Bayesian Modelling of the unfolding of time prediction

Mathias Vigouroux

Supervisor : Dr. Sophie Herbst



2021/2022

Contents

1

Abstract & Key Words

The brain predicts the timing of forthcoming events in order to guide behavior. Indeed, integrating a temporal structure in our internal representation of the environment, which is built from sensory inputs, in other words making temporal predictions, optimizes actions and benefits perception [1], [2],[3]. This phenomenon is referred to as temporal preparation. Yet, it still remains unclear how temporal predictions are formed from the temporal statistics of sensory inputs. The Bayesian framework is widely used to model the way we build inferences from sensory inputs [4], and is gaining popularity in the timing literature [5],[6]. According to this framework, a Bayesian observer learns the statistics of the sensory environment by updating an internal probability distribution. While it is usually hypothesized [Elisa report, Meinderstam, Ma] that human participants similarly represent a probability distribution of the external inputs, it is unclear whether humans are learning all the statistical parameters relevant to build the distribution. In the simplest scenario in which temporal predictions can be derived from a Gaussian distribution, which is fully described by just two parameters : the mean and the standard deviation, it still remains unclear if humans learn both the mean and the standard deviation of a distribution. Indeed, the hypothesis until now was that because both the mean and the standard deviation had an impact on human behavior then these two parameters were both learned by the participant and also by the Bayesian Observer. Thus, the models used until now, particularly in timing [Elisa report, Meinderstam], were always learning both the mean and the standard deviation. The validity of this hypothesis can be assessed in the framework of implicit temporal learning by comparing models that learn different types of parameters. To our knowledge, comparing Bayesian observers that learn the mean and the standard deviation to one that learns the mean only was not done in the literature. Yet, our study showed that a Bayesian Observer that only learns the Mean better explains participants' behavior compared to an Observer that learns both the mean and the standard deviation.Then, the effect of standard deviation can also be questioned in the realm of explicit timing and memory of duration. Asking participants to reproduce a series of stimuli, separated by durations drawn from a defined distribution, and comparing the standard deviation of the series they reproduced to the one they were shown, was not done in the literature to our knowledge. However, our study showed that the standard deviation of the produced foreperiods was not directly correlated to the standard deviation of the presented foreperiod. Rather, the standard deviation of the presented foreperiod had an impact on the mean of the produced foreperiods, and then the mean of the produced foreperiods had an impact on the standard deviation of the produced foreperiods. Therefore, this study challenges the common hypothesis that humans learn perfectly distributions of temporal series of stimuli, even though all the parameters, i.e. both the mean and the standard deviation, have shown robust effect on the behavior.

Key Words : Temporal prediction, Bayesian Modeling, Standard deviation.

2 Context

2.1 Temporal preparation in the real world.

With space, time is one of the fundamental characteristics of our dynamical environment. Anticipating and predicting events in our environment is thus decisive to guide our behavior. Indeed, cognitive systems have evolved to predict the when, where, and what of the sensory environment. These predictions have an impact on our behavior and our senses.

As an illustration, let's suppose that you are (kindly) reminded to have a vaccine... the type of vaccine is up to your imagination. You enter the room and the nurses ask you preliminary questions and to pull up your sleeve. They then told you that you will receive the dose in three seconds. You start counting : 1... 2... PICK ! You receive the dose after two seconds instead of three. You are surprised but you realized that it did not even hurt that much. This real world example shows how the explicit formation of temporal predictions and expectation about events has an impact on our senses and behavior.

Even 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 the ball in the air when playing volleyball, hit the shuttlecock when playing badminton or even cross a busy road [2]. In time literature, this ability is called temporal preparation [3].

2.2 Temporal preparation in the lab.

Since the beginning of XX th century, temporal preparation has driven the attention of cognitive scientists [7]. The foreperiod paradigm is one of the most commonly used experiments to assess this behavioral ability of the making of temporal predictions for temporal preparation. In this experimental setup, participants are presented a series of stimuli, usually auditory stimuli. These stimuli are paired by two : the cue stimuli and the target stimuli. After the target stimulus, participants are asked to answer, by pressing a button as quickly as possible and their response time are recorded as a proxy for their performance. The time interval between the two tones is the parameter of interest to study temporal preparation, and is called the foreperiod. 2.1

Indeed, it has been show that the mean duration of this temporal interval and also the variability of it had an effect on the participants response time (RT): the more variable is the foreperiod i.e. the bigger is the variance of the latent distribution of the foreperiod, the longer is the Reaction Time [Klemmer 1956, Niemi 1981].

Moreover, when the foreperiod is of fix duration, the longer the foreperiod is, the bigger is the reaction time.

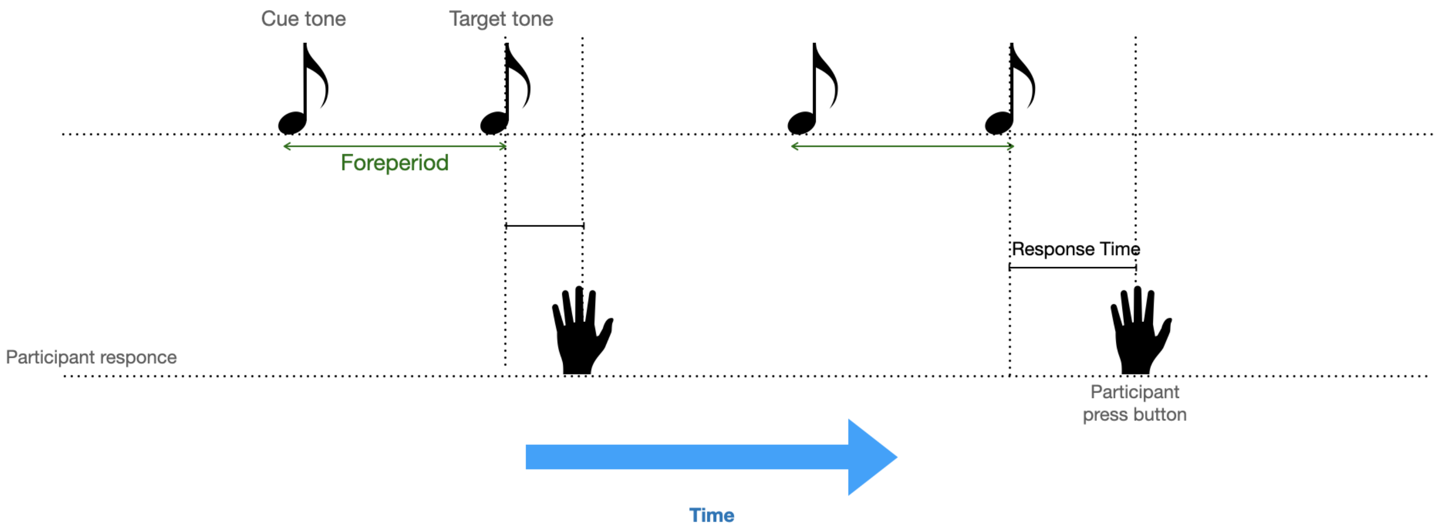


Figure 2.1: **The Foreperiod paradigm**
This psychophysics paradigm tries to asses the importance of temporal prediction on our behavior. Participants are presented a series of tones, paired by two, the cue tone and the target tone. These two tones are separated by a time interval called the foreperiod, which is the independent variable. The dependent variable is the duration of the Response Time of the participants.

This can be interpreted as the fact that participants try to learn the duration and their errors add up, participants having therefore a more biased estimation of 10 seconds compared to an estimation of 2 seconds. These two dependencies, to the mean and the variance, show that participants try to make temporal predictions, either explicitly or implicitly, of the occurring of the target stimulus after the presentation of the cue stimulus. This is materialized by longer reaction times for the participants when the learning of the foreperiod distribution is more challenging.

The remaining question is therefore how do we learn these parameters and distribution. A computational approach enables researchers to address this question more objectively. Yet, a good modeling of the behavioral 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]. In a nutshell, Bayesian models and Bayesian statistics in general try to capture the fact that we learn parameters by doing a trade off between what we used to know, our prior knowledge, and what we measure, in order to have a new representation of the world, the posterior knowledge. Indeed, because both our prior knowledge and what we have measured are noisy information, we will try to limit errors by averaging these two pieces of information.

2.3 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 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 C.1.

More formally, the Bayes' rule is defined by :

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

In other words, you will make a weighted average of what you have measured thanks to the weights of the prior distribution. 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.

The division by $p(x_{trial})$ is necessary for normalization, i.e. have a probability distribution that integrates to one.

The goal is now to assess this mathematical, theoretical, framework in the realm of psychology and human cognition. To what extent do we behave like that ?

2.4 Bayesian framework in psychology

2.4.1 Pilot experiment

The Bayesian framework started to be assessed with pilot data collected in an online experiment [thesis Elisa], using a foreperiod design. Two series of foreperiods differed by the latent distribution from which they were generated and the frequency of the cue tone in order to suggest to the participant that there were two types of foreperiods C.2. This behavioral data was then compared to the predictions of a simulated agent : the Bayesian Observer which tried to predict following Bayes' rule.

The Bayesian Observer is a model that tries to mimic the participant learning strategy. that behaves in a Bayesian way. By following Bayes rule, it will, for each foreperiod, estimate the underlying statistics of the foreperiods. Thus, it has to infer the mean and standard deviation of the distribution in order to know the distribution perfectly.

On one hand, participants, who are asked to answer as quickly as possible to the target tone, try to predict, consciously or unconsciously, the foreperiod in order to be ready as much as possible to answer to the cue tone and orient their attention in time. As a consequence, the prediction of the participants will influence their response time : this is the finding of temporal preparation, robustly assessed in the literature [8],[9],[10].

On the other hand, the Bayesian Observer statistically estimates the latent distribution of the foreperiods, and then computes predictions of the foreperiod trial-by-trial. However, it is impossible to know the prediction of the participants and therefore impossible to know how they form these predictions.

The idea is thus to use the prediction of the Bayesian observer as a substitute for the participants' predictions and compute from the predictions of the model measures of surprise and preparation that will then be used as a regressor of the reaction times of the participants. This first study shows that derived quantities from the prediction of the Bayesian Observer are significant to model the participants' behavior, giving good hope for further studies. Two main quantities caught our attention.

2.4.2 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 :

$$err = p_{n-1} - \gamma_n \tag{2.2}$$

2.4.3 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

$$I(FP_n) = log(f(FP_n)) = -log(f(FP_n)) \tag{2.3}$$

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, hence 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.

2.5 Limits of the previous studies

2.5.1 Hypothesis on the Bayesian Observers

In the continuity of the previous hypothesis in the literature, these Bayesian Observers were "perfect" : they were learning everything to describe completely, and perfectly, the latent distribution.

This can be broken down to two main hypotheses.

The first one is that the model, and the participant, know that the latent distribution is Gaussian. This hypothesis is reasonable to the extent that the Central Limit Theorem shows that all distributions tend to be Gaussian. Therefore, as this result is valid at all physical scales it could be already built somewhere within us, or at least, mature humans, and therefore participants, are likely to have built a heuristic on this result, consciously or unconsciously. The Bayesian Observer is therefore solely a parametric estimator.

The second hypothesis is that the model learns all the parameters to describe the latent distribution perfectly, i.e. the mean and the standard deviation. However, the importance of the learning of these two parameters are still unclear in implicit timing tasks.

Effects of the mean of the foreperiod on participants' response time were already depicted in the literature without taking into account any Bayesian reasoning. However, an other effect of the mean, the regression to the mean effect, which describes the tendency to bias the reproduction of a stimuli toward the mean of the presented stimuli, seems to be an argument in favor of the Bayesian brain hypothesis since it shows the importance of the previously learned information [11]. However, this phenomenon could be explained only by a dependence to the mean, for instance a linear regression on the mean, without the formation of an internal probabilistic distribution for the participant. Then, the effect of surprise, and more generally temporal preparation, shows the importance of predictability in the participant behavior. Indeed, Naantanen in 1981 [8] revealed that temporal uncertainty has an effect on the response time of participants. On the same note, the standard deviation has already shown an effect in behavioral study with rodents [12]. However, these findings do not question the learning of the standard deviation. Yet, this parameter is more challenging to be learned, even statistically, compared to the mean. Therefore, it seems that the effect of the learning of the standard deviation in temporal preparation has to be better explored. As a consequence, the experimental setup will focus mainly on the role of the standard deviation in temporal predictions and preparation.

2.5.2 Complexity of the task

Moreover, different types of Bayesian Observers have been used during the previous work done by Elisa Lannelongue reflecting different prediction strategies of the participant, who could be either aware or not that there are two foreperiod distributions for each of the cue tones. Indeed, in this previous experimental settings, there were two sets of tones for each block and each tone corresponded to a particular foreperiod distribution. Thus, the participants had to learn the foreperiods distributions and also the fact that there were two distributions. However, the type of Bayesian Observer that best describes the participant behavior could not be found unambiguously. This raises the question of the formation of categories that might follow an Occam Razor : participants might not do the effort to learn that there are two distributions for each of the two auditory cues[13], and is embedded in a structural learning process.

Yet, the learning of the distribution of even one foreperiod distribution is still unclear. Therefore, this new experimental paradigm focuses solely on the Bayesian Brain hypothesis for the unfolding of temporal predictions and tries to assess if the brain learns temporal predictions similarly to a Bayesian observer and what statistics it learns. Therefore, for the current study, the number of parameters of the experimental paradigm is decreased. Thus, only one foreperiod distribution will be used per block, and it is only the standard deviation of these distributions that will vary between each block.

2.6 New types of Bayesian Observer

Following the previously described rationale [6] which compared two types of Bayesian observers, this new experiment will also compare two types of Bayesian observers. Once again, this would help better understand the learning strategies of the participants.

The first one is the Mean/Std Observer : this Bayesian observer updates both the mean and the standard deviation of the latent distribution of the presented foreperiods, this is the type of observer used in the literature [6], [11],[4].

The second one is the Mean Only Observer : this Bayesian observer only updates the mean of the foreperiods series. In other words, the first Bayesian observer is aware that there are both a mean and a standard deviation that have to be learned for the foreperiods and the second one is only aware that there is a mean to learn.

Therefore, if the prediction of the Mean/Std Observer is better to explain the behavior of the participant engaged in the implicit timing task than the Mean Only Observer then it means that the participant is likely to internalize the variance of the latent distribution of the foreperiods.

The hypothesis until now was that because both the mean and the standard deviation had an impact on human behavior then these two parameters were both learned by the participant and also by the Bayesian Observer. Thus, the models used until now were always learning both the mean and the standard deviation. The validity of this hypothesis can be assessed in the framework of implicit temporal learning by comparing models that learn different types of parameters. To our knowledge, comparing Bayesian observers that learn the mean and the standard deviation and the mean only was not done in the timing literature.

2.6.1 The mean and Std observer

The first type of Bayesian Observer is the Mean and Std 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.[14]

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

2.6.2 The Mean Only Observer

The mean will be updated by :

$\mu_{t+1} = \frac{\mu/\tau_t^2 + y/\sigma_t^2}{1/\tau_t^2 + 1/\sigma_t^2}$

where τ_t is a scaling parameter, that is updates by : $1/\tau_{t+1}^2 = 1/\tau_t^2 + 1/\sigma_t^2$

In other words : $\theta|y \sim \mathcal{N}(\mu_1 \tau_1)$

As it can be seen, a value for the standard still has to be defined. Our idea was therefore to define the standard deviation using the Weber law, i.e. formula $\sigma_{t+1}^2 = WeberFract.\mu_{t+1}$.

Weber discovered at the beginning of the 19th century that the just noticeable difference when participants are asked to measure something (a distance for example) is not absolute but relative. For instance, people will easily notice the difference of distance between a stick of 2 cm and 4 cm, yet the length of a car of 4 m and 4.2 will look exactly similar. In other words, "the increase in estimation bias for larger magnitudes is accompanied by a linear increase in the standard deviation of the estimated magnitude with its mean. That is, estimates of larger magnitudes are noisier than the estimates of smaller magnitudes" [11]. The proportionality factor between these two variable is know as the Weber Fraction. Usually, in the temporal domain this Weber fraction is around 0.1This proportionality factor, known as the Weber fraction, will be chosen to 0.1 (0.17 for [5], 0.13 for [11]).This Weber fraction was first chosen to be around 0.1 in order to follow the literature. Yet, this parameter was then optimized (see subsection 5.4.3).

2.7 Exploratory study for the reproduction of a foreperiods series

The learning of the Std of the foreperiod was also assessed in the realm of explicit timing and memory of duration. Indeed, at the very end of the previously mentioned implicit task participants were asked to reproduce a few sets of cue and target tones from the last block of foreperiod that they heard. As the order of the block was randomized between each participant, participants needed to reproduce different variability. Participants were however not explicitly asked to focus on the variability. Inferring a variance is more difficult than a mean even in a statistical point view : if only one measure is needed to estimate the mean of a series, at least two measures are necessary to code the variance, as the variance is by definition the error to the mean. Temporal uncertainty plays a role in the participants' behaviors and thus favors the importance of standard deviation, the difficulty of estimating it might go against the learning of this complex parameter.

For instance, if you are asked to make tones spaced by a two seconds interval you might not have any problem creating this temporal series. However, the question becomes much more challenging if you are asked to create a similar temporal series with still a mean of two seconds, but this time it has to have a 1 seconds standard deviation !

Asking such a question might be extremely hard as a majority of people are not familiar with statistics. However, the effects of the variance of the presented stimuli on the variance of the reproduction task were not developed to our knowledge. This reproduction paradigm is interesting as it bypasses the ignorance of the exact definition of standard deviation and focuses more largely on the role of learning and memory of variability and uncertainty.

3 Research Questions and Working Hypothesis

3.1 Research Questions

- Do the reactions time and reproduction of a series of foreperiods depend on the variance of the foreperiods presented when the participants are engaged in an implicit timing task followed by an explicit timing task ?
- Does the behavior of participants engaged in an implicit timing task depend on the variance of the temporal statistics of the environment ?
- First : Are participants' response times better explained by a Bayesian observer that also represents the variance of the external world ? In other words, is the variance of the latent distribution perceived and updated by the human observer ?
- Second : If asked to explicitly reproduce the previously presented foreperiods, does the reproduction of the foreperiods distribution depend on the variance of the latent distribution ? In other words, is the "stored " variance recruited for explicit motor command by the participant ?

3.2 Rational of the study

The protocol was composed of one main experiment to which we added an explicit assessment of the representation of temporal variability. In the main experiment, referred to as the implicit timing task (since participants were not asked explicitly to track time during this experiment),composed of four blocks, participants were just asked to press space as quickly as possible after the target cue. The additional experiment, is referred to as the explicit timing task, since participants were asked explicitly to remember temporal durations, was immediately following the main experiment and participants were asked to reproduce the last foreperiods that they heard.

3.3 Working Hypothesis

The main experiment will check two findings : first verify if there is an effect of the Std on the Reaction Time.

WH1: *Human observers represent the variance of the temporal regularities in the sensory environment during an implicit timing task. Therefore their surprise, indicated by prolonged reaction times, is larger when an observation diverges from a narrow versus a broad distribution. Yet, as participants are engaged in an implicit timing task, it is impossible to have access to their predictions so we rely on temporal preparation.This results in a larger reaction time if the foreperiods are less predictable.*

Then, assess if the Bayesian framework is valid . In order to do so 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 temporal statistics.

WH2: *Human learning can be described by a Bayesian observer. This is indicated by a correlation between the theoretical parameters of a Bayesian observer model (e.g., surprise) and human behavior.*

The additional experiment will try to assess to what extent participants had memorized the variability of the presented foreperiods.

WH3: *Humans are learning the std of a presented foreperiod series. Therefore, when participants are asked to reproduce this series of foreperiod the std of their reproduced series is correlated with the std of the presented std.*

4 Material and Methods for the Main Experiment

This study was conducted with 31 participants. This sample size is commonly used in the temporal literature [5],[10]. The behavioral testing was done outside of Neurospin at the Institut du Cevreau et de la Moelle Epiniere, on the PRISME Plateform, where they have a room enabling us to have up to twelves participants in parallel including eye tracking. Eye tracking data will be useful for a further study where the impact of variance will be assessed at a different biological scale. The study was conducted under an ethical authorization from the Comité d'Éthique pour le Recherche (CER) de l'Université Paris-Saclay (Dossier N° 371).

4.1 Experimental Design

As it was mentionned in the introductory part, the "foreperiod paradigm" enable scientist to study the role of temporal predictions. This paradigm is schematized in 2.1. 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].

Participants were placed in front of a computer, put the brand of the computer), where they were autonomously following the instructions D. During these blocks, the participant were only asked to answer as quickly as possible to the Target stimuli. Two Bayesian observers also learned the distribution during these blocks. The experiment started with a calibration of the eye tracker, an EYE Tribe tuned to a 60 Hz sampling frequency. Participants also received the auditory stimuli through headphones. The experiment was back end by a Matlab running code, the Psychtoolbox library for the GUI. The analysis conducted in Python.

4.2 Foreperiod distributions

The foreperiod were drawn in order to study the effect of the variability of the foreperiods on our behavior 4.1

Before the beginning of the experiment there was a training phase. In this training phase, participants are presented with a foreperiod distribution that was distinct from the foreperiod used in the rest of the experimental paradigm. Indeed, a distribution with a mean of 2.5 seconds and 0.25 second for the standard deviation was chosen. A Weber Fraction of 0.1 was therefore selected for the distribution of the foreperiods during the training phase. Thus, this distribution was likely to be easily learned.

In the main task, four foreperiod distributions were selected, all Gaussian, with a mean of 2.5 s and varying standard deviation. The mean duration of the foreperiods is 2.5 seconds in order to have 120 trials leading to

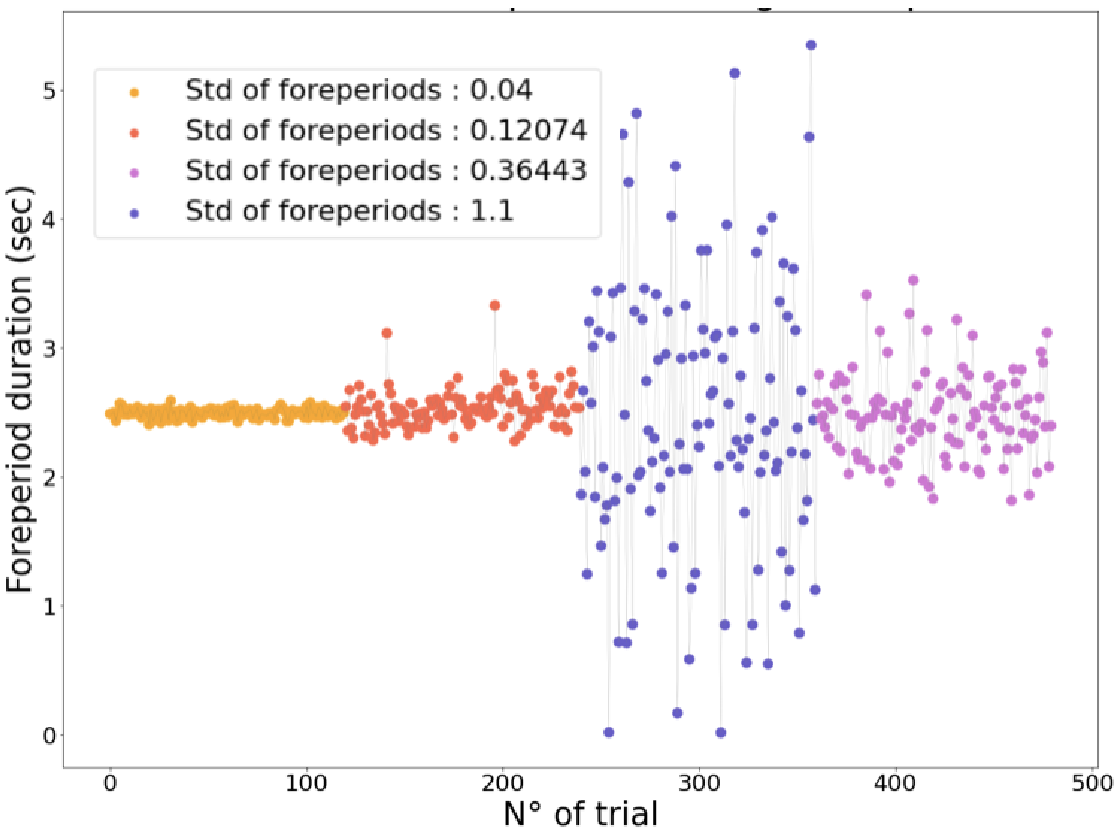


Figure 4.1: **Example of foreperiods presented to one participant**
The variation of the foreperiod distribution between block is the variance. The foreperiods can be either really predictive, in the orange block, or not predictive at all, in the blue block. The order of the blocks were randomized between each participants.

14 min of trials and 4 blocks . Indeed, for the experimental design we have decided to use four blocks. For these decisions, a major determinant was the block duration (<15min) and duration of the experiment (<1.5h).

The standard deviations were drawn on a log scale between 0.04 and 1.1. Their histograms are depicted in C.3. The limiting factor for the upper limit of this range is the learning of the std by the Mean Std observer in a finite amount of time. On the other end, the inferior limit was chosen to be almost 10 times smaller than the std gave by a Weber fraction of 0.1 (classical value [11],[5]) and the mean of 2.5. Moreover, human participants have shown to have a limit for their temporal resolution and won't differentiate temporal intervals separated by only just a few milliseconds. The choice of the log-scale goes once again with the findings of Weber. Indeed Fechner derived a logarithmic relation between absolute physical and sensed magnitudes, the Weber-Fechner law [15].

4.3 Inter Trial Interval

The foreperiods are separated by an Inter Trial Interval (ITI). Inter Trial Interval were drawn following a uniform distribution between 0 and 6 seconds in order to have a mean of 3 seconds. Thanks to that the participants can't predict the onset of the cue stimulus. At the beginning of the block participants wait for 5 seconds in order for them to be more focused at the appearance of the stimulus.

4.4 Auditory Cues

Moreover, in order to better distinguish the different blocks, the tone of the cue tone was also varying from 400 HZ to 1100 Hz, still on a logarithmic scale. The tone of the target tone remained the same through the entire experiment.

4.5 Feedbacks

Participant were only instructed to press the space-bar as quick as possible after the target tone and to gaze at a fixation cross in the center of the screen. The fixation cross at the center of the screen would ensure the same condition for attention for all of the participants. In order to motivate participants to have better performance, they were instructed that depending on their performance the color of the fixation cross would change for a few seconds (it changed for 1 sec). If their reaction time was really good, i.e. below 0.2 second, the fixation cross would turn green for 1 second. If it was intermediate, between 0.2 second and 0.4 second, it would turn orange for 1 second. If it was long, above 0.4 second, it would turn red. These thresholds were defined empirically in order to have from time to time green feedback if really focused, orange most of the time, and red only if completely distracted. Participants were also informed if they answered too early, i.e. pressing the space-bar between the cue tone and the target tone. Participants were also informed if they did not answer after 2 seconds after the target tone and the code passed to the next foreperiod.

5

Analysis and Results of the Main Task

5.1 Preprocessing

Two participants were only shown one fourth of the experiment because due to an error in the code. These two participants were removed. Therefore, there were 29 participants in total. Afterward, trials where participants answered two soon or did not answer were removed. Moreover, trials where the Reaction Time of participants were above 2.5 std of the average response time (RT) of this participant were also removed.

5.2 Result of the Behavioral Effect

5.2.1 Assessing the quality of the data.

A first quality check of the data consisted of an inspection of the distribution of the RT. The distribution of the RT is skewed. Yet, the log(RT) is not skewed 5.1. This is coherent with the literature and goes along with psychophysics models such as the diffusion drift model.

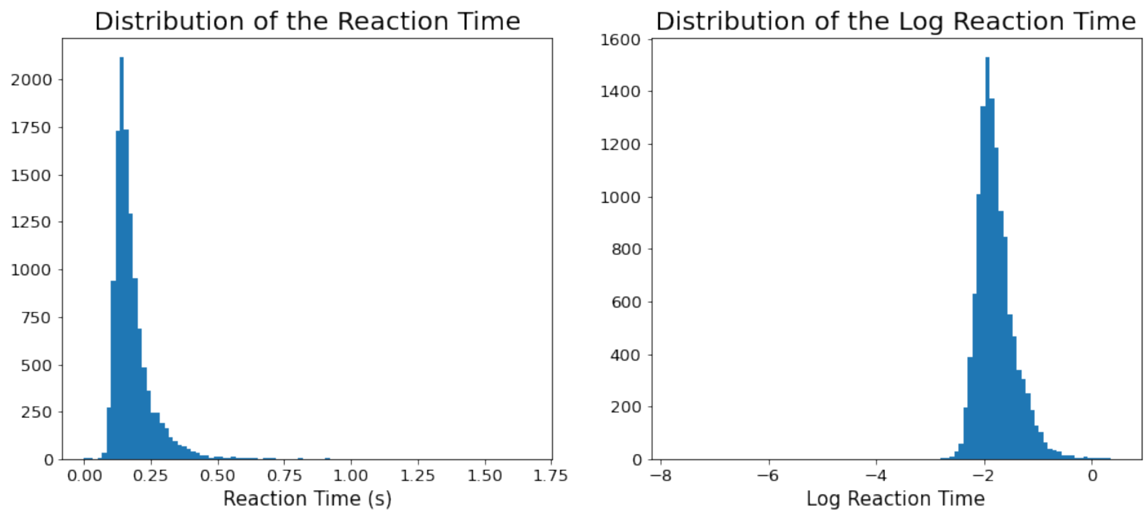


Figure 5.1: **Response times of the participants**
The left panel shows the raw distribution of the Response Time. The distribution is skewed. The right panel shows the distribution of the log transformed RT. This changed in the skewness of the RT is a good indicator of the quality of the data.

5.2.2 Checking participant by participant

The idea was to check if there was an effect of the foreperiods on the participants behavior by checking participants one after the other. There is a significant effect of foreperiods durations (p-value <= 0.05) for 74 % of participants. On the same note, Std had a significant effect (p-value <= 0.05) on the RT for 74 % of participants.

5.2.3 Doing a Linear Mixed Effect Model with the experimental setup

To check more rigorously the effect of the experimental design on the behavior, here the reaction time, a multivariate analysis, with all the parameters of interest, was conducted. However, the multivariate analysis cannot be "just" a simple multiple regression because there isn't enough participant in this pilot data (only 32). A Linear Mixed Effect Model was therefore selected.

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 C.4. In our case, the regression is defined as :

$$Log(RT) = 1 + Block + NbOfTrial + PresentedFP + Std + (1|Participants) \tag{5.2}$$

Instead of the Response Time directly, it was the Log(RT) that has been chosen as it not skewed, hence preventing statistical artifacts.The results can be found in 5.2.

Model:	MixedLM	Dependent Variable:	Log_RT
No. Observations:	12158	Method:	REML
No. Groups:	29	Scale:	0.0930
Min. group size:	345	Log-Likelihood:	-2916.2387
Max. group size:	432	Converged:	Yes
Mean group size:	419.2		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	-1.697	0.041	-41.229	0.000	-1.777	-1.616
Block	0.018	0.002	7.059	0.000	0.013	0.022
Nb_of_trial	0.000	0.000	5.379	0.000	0.000	0.001
Presented_FP	-0.073	0.005	-14.787	0.000	-0.083	-0.063
Std	0.098	0.007	14.851	0.000	0.085	0.111
ITI	-0.001	0.003	-0.223	0.823	-0.007	0.006
Group Var	0.041	0.036				

Figure 5.2: **Regression of the Log(RT) on the experimental parameters** As expected, the ITI has no effect on the behavior. However, both the mean and the std show a significant effect. **This validate the working hypothesis n°1.**

The number of the block (variable : $Block$) was conserved since it captures the fatigue, or the learning of participants during the experiment. The positive sign of the coefficient for this regressor shows that the learning might be more important.

The trial number within one block (variable : $NbOfTrial$), showed a too small effect to be interpreted.

The Inter Trial Interval, the duration between two foreperiods, had no significant effect. This favors the good setting of our experiment as we wanted by definition that the ITI would not impact the RT ($p - value \simeq 0.8$).

The Std has indeed a significant and positive effect. As a consequence, the **Working Hypothesis 1 is validated**.

On the other hand the foreperiods had a significant and negative effect. This go along with the variable foreperiod paradigm. 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 [8] .

Then the goal is to understand if participants learn the standard deviation. In order to do so, the behavioral responses will be compared to different Bayesian models, one that learn the mean only and one that learn both the mean and the standard deviation. The Bayesian model is therefore the optimal learner of temporal statistics.

5.3 The Bayesian Observers

5.3.1 A difference in the learning of the parameters...

The Mean / Std Observer learns both the Mean and the Std "independently". Therefore, this observer will really learn the distribution. On the other hand the mean only observer will just learn the mean and not the std. This Observer will thus have a distorted learning of the distribution. The difference in the learning of the statistical parameters can be found in 5.3.

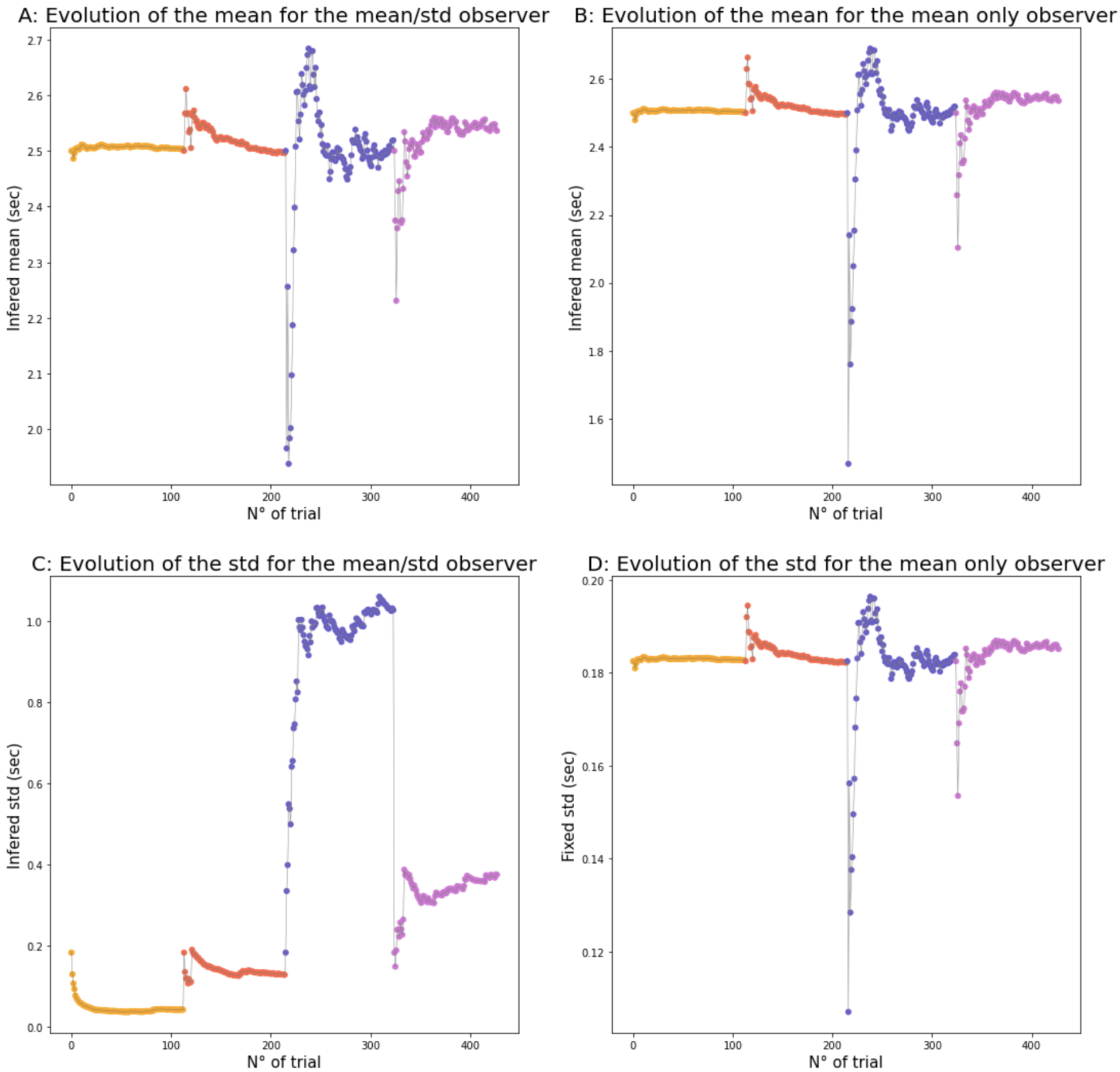


Figure 5.3: **Learning of the statiscial parameters for the two types of Bayesian Observers**
The color represent the different types of block. On the left side, the panel A and C depict the learning of the Mean/Std Observer. This two parameters seem to vary independently. On the right side, the panels B and D depicts the learning of the Mean and Std for the Mean Only Observer. The Std of the foreperiods for the Mean Only is only a fraction of the Mean.

5.3.2 ...that results in a small difference in the prediction errors...

The prediction errors for the two types of observer are highly correlated and similar (upper panel of C.5, or 5.4 to see for each Std separately). Indeed, the prediction error just takes into account the mean for the two observers. However the two observers are learning almost exactly the same meaning. This measure does not take into account the distribution learned by the observer but just the simplest parameter of it : the mean.

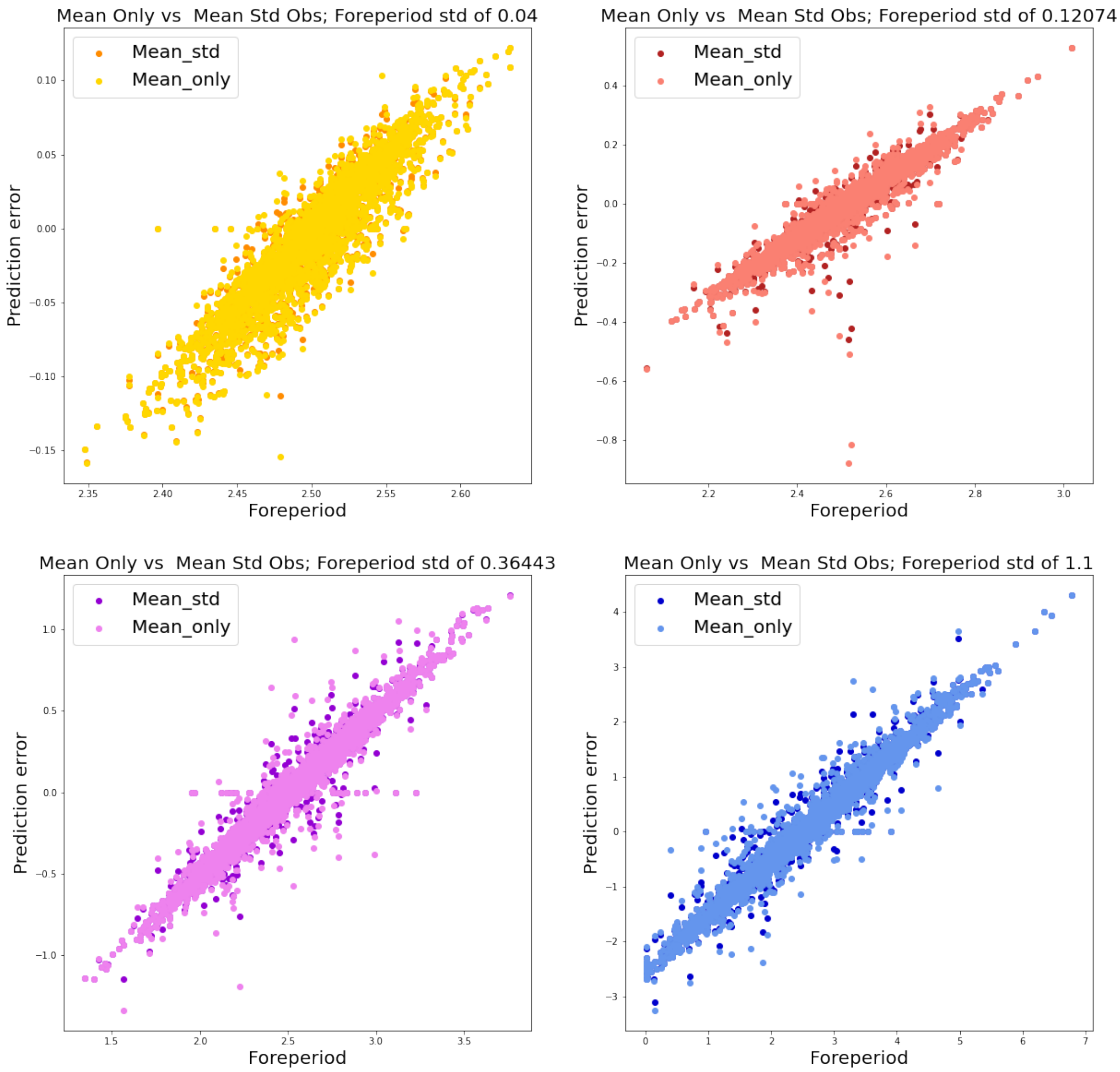


Figure 5.4: **Comparison of the Prediction Errors between the 2 Observers, block-by-block**
Whether in the really predictive case (yellow case) or in the non predictive case at all (blue case) the prediction errors of the two types of observers are highly correlated.

5.3.3 ... and a bigger difference in the surprises of the two Observers.

However, the surprise do take into account the distribution learned by the Observer and might therefore be more different from one another. By looking at the surprise, the measure of surprise for the two types of observers that differ a lot. This difference can be also inspected for each type of block, i.e. each std of foreperiod as it is depicted in 5.5.

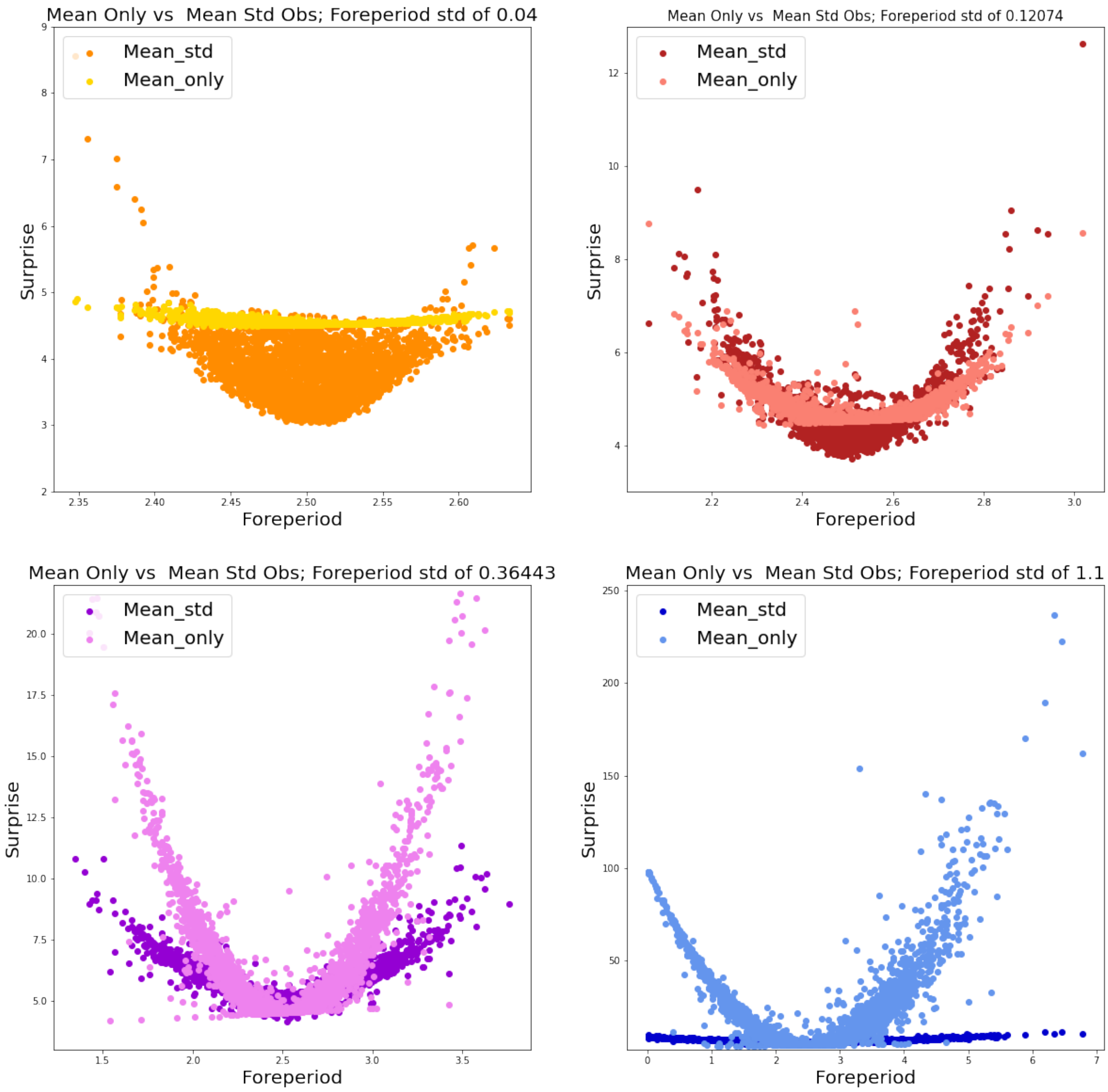


Figure 5.5: **Comparison of the Surprise between the 2 Observers, block-by-block**
Unlike the prediction errors, the comparison of surprise show a dependency on the type of block. When the block is really predictive, orange case, the surprise of the Mean Only Observer seems to be flat compared to the surprise of the Mean/Std. When the std becomes bigger, it is the opposite : it is the surprise of the Mean / Std Observer that seems flat compared to the suprise of the Mean Only Observer. The differences for the surprises of these two observers can be also visualized in the newly called Phoenix Plot C.

The key research question for this implicit timing task is therefore : **Which of the two Bayesian Observers, the Mean Only Observer and the Mean Std Observer is the best one to explain behavioral data ?**

5.4 Finding the good Bayesian Observer.

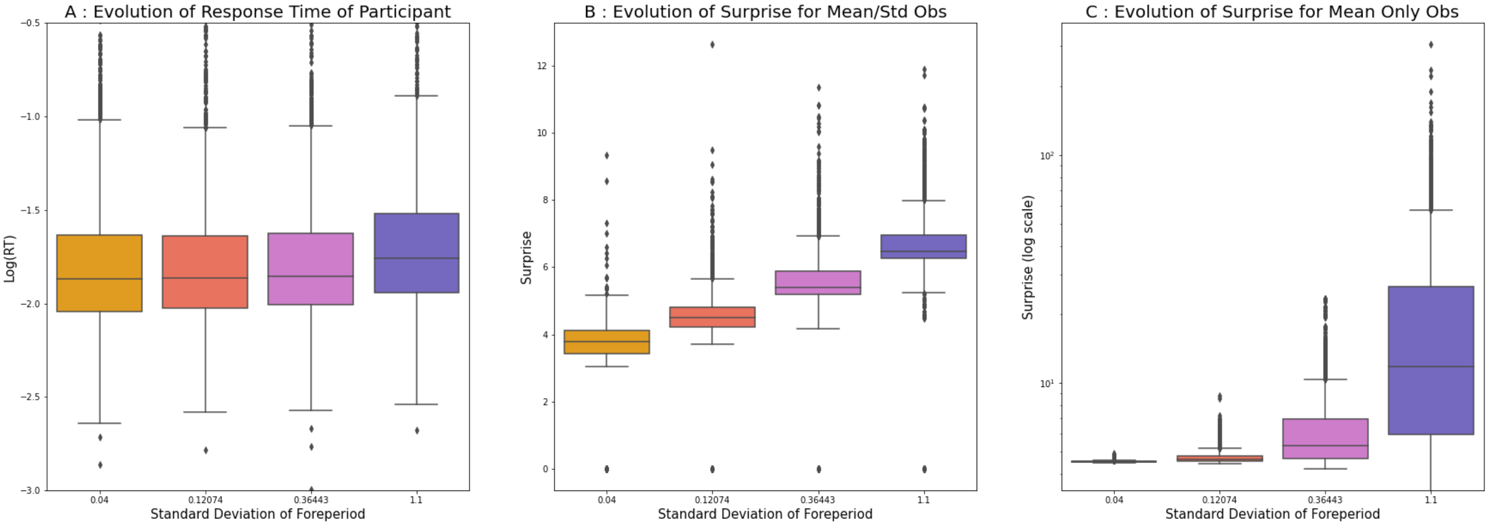


Figure 5.6: **Comparing the participant behavior to the predictions of the two models**
Panel A represents the increase of the Response Time with the Std of the foreperiod distribution. It seems that this increase is not linear, since it is mainly for the biggest Std that we see a difference compared to the other condition. Panel B gives the evolution of the surprise for the Mean/Std Observer with the Std of the foreperiod distribution. This increase seems to be linear. Panel C gives the evolution of the surprise for the Mean Only Observer with the Std of the foreperiod distribution. This increase seems to be non linear and more exponential.

On one hand, the reaction time of the participant increases with the Std. This increase does not not seems to be linear (Panel A of 5.6). Indeed, the increase in RT is more important for the biggest std compared to the other one.

On the other hand, the increase of the surprise for the mean std observer is linear (Panel B of 5.6) and the increase for the mean only observer is more exponential (Panel C of 5.6). This mean that, the Mean / Std Observer tries to learn something from the presented foreperiod. When the variance of the foreperiod is large, the Mean /Std Observer will lower its expectations and therefore will not be much surprise when it encounters very short and very long foreperiods. However, the Mean Only Observer, just try to learn the mean, which is the same in all the experiment. It therefore struggle to optimize its predictions, and is really surprise by very short and very long foreperiods, independently of the block.

5.4.1 Comparing the prediction errors

The different regressors computed from the predictions of the two models are correlated with each others.The difference

In other words, the prediction error of the Mean Only Observer is correlated with the prediction error of the Mean/Std Observer C.5. Our first approach was therefore to regress the prediction errors of the Mean/Std

Observer (resp. prediction errors of the Mean Only Observer) on the prediction errors of the Mean Only Observer (resp. Prediction error of the Mean/Std Observer). This regression gives a residual, that we called the Innovation of Prediction Error of the Mean/Std observer (resp.Innovation of Prediction Error of the Mean Only observer).

More formally :

$$PredErrorMeanStd = \alpha_1 + \beta_1.PredErrorMeanOnly + InnovationPredErrorMeanStd \tag{5.3}$$

$$PredErrorMeanOnly = \alpha_1 + \beta_1.PredErrorMeanStd + InnovationPredErrorMeanStd \tag{5.4}$$

These innovation were uncorrelated with the Prediction Error from the other type of observer and captured what was not linearly explain by the Mean Only Observer Prediction Error in the Mean Std Observer Prediction Error (resp.the Mean/Std Observer Prediction Error in the Mean Only Observer Prediction Error). Then, the participants' Reaction Times were regressed on the Prediction Error of the Mean Only Observer plus the Innovation of the Innovation of Prediction Error of the Mean/Std Observer (.resp Reaction Times were regressed on the Prediction Error of the Mean/Std Observer plus the Innovation of the Innovation of Prediction Error of the Mean Only Observer).

More formally, by doing two Linear Mixed Effect models defined by :

$$Log(RT) = 1 + PredErrorMeanOnly + InnovationPredErrorMeanStd + (1|Participant) \tag{5.5}$$

$$Log(RT) = 1 + PredErrorMeanStd + InnovationPredErrorMeanOnly + (1|Participant) \tag{5.6}$$

These two regressions show that the Innovations, therefore the difference between the two types of participants were not significant C.11. This already goes against the literature as only the Mean/Std Observer was used until now. Indeed, following an Occam Razor Principle, if two solutions explain the same phenomenon then the easiest one has to remain.

5.4.2 Comparing the surprises

As it was shown before, prediction errors from the two observers are really similar, which would explain why the difference between the two was not significant. One could argue to do the exact same thing with the surprise. However, the correlation between the surprises of the two types of observers is oddly distributed. It is made of a multi-linear correlation, with a slope depending on the block. Therefore by removing one linear factor the residual still remain correlated. This multilinearity is depicted in the lower panel of C.5.

Therefore, it is impossible to do the same thing with surprise.

5.4.3 Comparing separately the two observers to the participant behaviour

In order to find out which of the two model is better to explain participant behavior a new approach composed of two linear mixed effect models was conducted. In this model, each of the regression are solely composed of either regressors from the Mean Only Observer or solely composed of regressors from the Mean Std Observer.

More formally, we defined :

$$Log(RT) = 1 + PredErrorMeanOnly + SurpriseMeanOnly + (1|Participant) \tag{5.7}$$

$$Log(RT) = 1 + PredErrorMeanStd + SurpriseMeanStd + (1|Participant) \tag{5.8}$$

For these two models, all the regressors are statistically significant, showing that the Bayesian Observers are indeed good models of the participants behavior 5.4.3. Therefore our **Working Hypothesis n°2 is validated**.

Model:	MixedLM	Dependent Variable:	Log_RT
No. Observations:	12158	Method:	REML
No. Groups:	29	Scale:	0.0928
Min. group size:	345	Log-Likelihood:	-2900.1754
Max. group size:	432	Converged:	Yes
Mean group size:	419.2		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Mean_only_pred	-0.082	0.005	-16.548	0.000	-0.091	-0.072
Intercept	-1.873	0.038	-48.955	0.000	-1.948	-1.798
Mean_only_surpr	0.003	0.000	16.001	0.000	0.003	0.004
Block	0.019	0.002	7.613	0.000	0.014	0.024
Nb_of_trial	0.000	0.000	5.135	0.000	0.000	0.001
Group Var	0.040	0.036				

Figure 5.7: Result of the regression on the mean only predictions

Model:	MixedLM	Dependent Variable:	Log_RT
No. Observations:	12158	Method:	REML
No. Groups:	29	Scale:	0.0935
Min. group size:	345	Log-Likelihood:	-2943.3084
Max. group size:	432	Converged:	Yes
Mean group size:	419.2		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Mean_std_pred	-0.074	0.005	-15.067	0.000	-0.083	-0.064
Intercept	-2.049	0.042	-49.318	0.000	-2.131	-1.968
Mean_std_surpr	0.037	0.003	12.796	0.000	0.031	0.042
Block	0.017	0.002	6.769	0.000	0.012	0.022
Nb_of_trial	0.001	0.000	6.273	0.000	0.000	0.001
Group Var	0.040	0.036				

Figure 5.8: Result of the regression on the mean std predictions

Figure 5.9: **Trying to find the best model** The regressors extracted from both of the Bayesian Observer remain statistically significant. **This validate the working hypothesis n°2.**

The likelihood of these two models were compared in order to find out which type of Bayesian observer best describes the participants' behavior. The likelihood of a model is a measure of the fitness of the model on the behavioral data. By iterating through different Weber fractions, the likelihood of the model with quantities derived from the Mean Only Observer goes above the likelihood of the with quantities derived from the Mean/Std Observer. Indeed, it is the case for the Weber fraction between the minimum that has to be imposed to insure the convergence of the surprise, and around 0.2 as it is depicted in the figure below : 5.10.

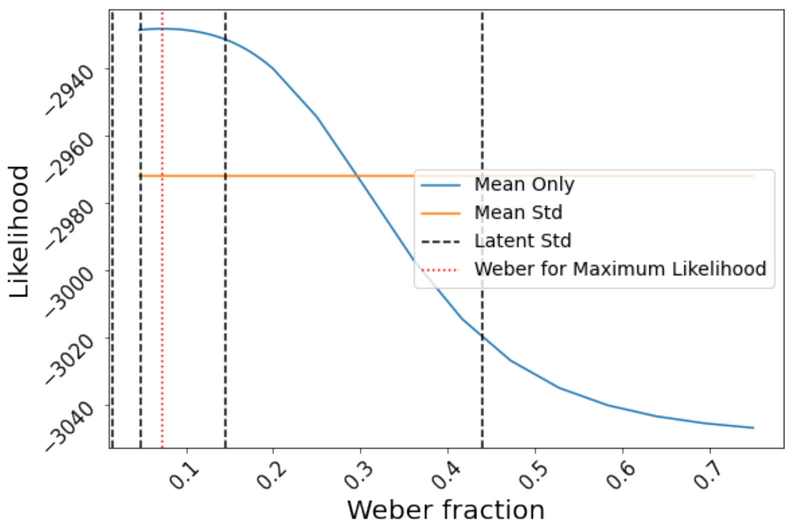


Figure 5.10: **Optimizing the Weber Fraction**
The blue line represent the evolution of the likelihood of the regression of the participants RT on derived quantities from the Mean Only Observer. The orange line represent the evolution of the likelihood of the regression of the participants RT on derived quantities from the Mean Std Observer, which is independent of the Weber fraction, which is a parameter of the Mean Only Observer solely. The black dashed line represent the Weber fractions associated with the four latent Std. The red dashed line represent the Weber Fraction which gives the best model. Since, the Weber fraction is a independent parameter it can be optimized freely. Finding Weber fraction for which the behavioural data is better explained by the Mean Only Observer compared to the Mean/Std parameter thus challenges the classical hypothesize that Bayesian Models had to learn both the mean and the std.

6 The reproduction task

6.1 Rational

The second task is the reproduction task. At the end of the fourth block, participants were asked to reproduce the last set of foreperiod that they heard. The goal was to see how the standard variation of the last block that they heard affects the standard variation of the series of foreperiods that they created.

6.2 Methods

Participants were instructed to terminate the interval started by the cue tone. Indeed, cue tones, with the same tone as the one they heard in the last block (reminder : the tone varies between each block, hence with the variance), were provided to them. After that they had to remember the duration between the cue tone and the target. Then participants pressed on the space bar and this emitted a tone with the same tone as the target tone. After this tone, there was an Inter Trial Interval, drawn from the same distribution as the rest of the experiment. And then again a cue tone, and participants had to remember the duration of the foreperiod, etc. This was done in order to have a series of 30 reproduced foreperiods by the participant.

6.3 Analysis

For each series of foreperiods created by the participants, the mean and standard deviation were computed. These two parameters were exhaustively compared to all the parameters of the experiment : the mean of the presented foreperiod, the mean and std of the ITI, and also to their produced mean.

More formally :

$$ProducedStd = RealMean + MeanITI + StdITI + RealStd + ProducedMean \tag{6.1}$$

The $RealStd$, is the mean duration of the presented FP and plays here the role of an intercept since it was always fixed at 2.5 s. The $MeanITI$ and the $StdITI$ represent the effect of the Inter Trial Interval, and are expected to have no effect. The $RealStd$ refers to the standard deviation of the last block presented to the participants. The $ProducedMean$ correspond to the average foreperiod created for each participant.

6.4 Results

Participants were able to remember the mean of the presented foreperiods. Indeed, the average of the mean of the foreperiod series that they created was similar to the mean of the presented foreperiods (2.6 sec for the reproduction for a real mean of 2.5 sec).

As expected we found no effect of the ITI, both of the mean and the standard deviation, on the produced standard deviation C.12.

Yet against our prediction, the Std of the produced series of foreperiods was unaffected by the presented Std (p-value of 0.08 if only regressor C.13 and quickly become not significant at all when other regressors are added up C.12. As a consequence, the **Working Hypothesis n°3 is NOT validated**.

Classically, again in agreement with the finding of Weber, the std of the produced series were correlated to the produced mean. The Weber fraction is equal to 0.74, hence a 10 factor with what we have found during the implicit task (Optimal Weber of 0.073) C.15. More interestingly, the mean of the produced series was affected by the standard deviation of the presented foreperiods. This time the Weber fraction is smaller (0.012) and more in agreement with the literature [5], [11].

The produced standard deviation was not directly impacted by the presented standard deviation. However, it was impacted indirectly through the mean of the produced series.

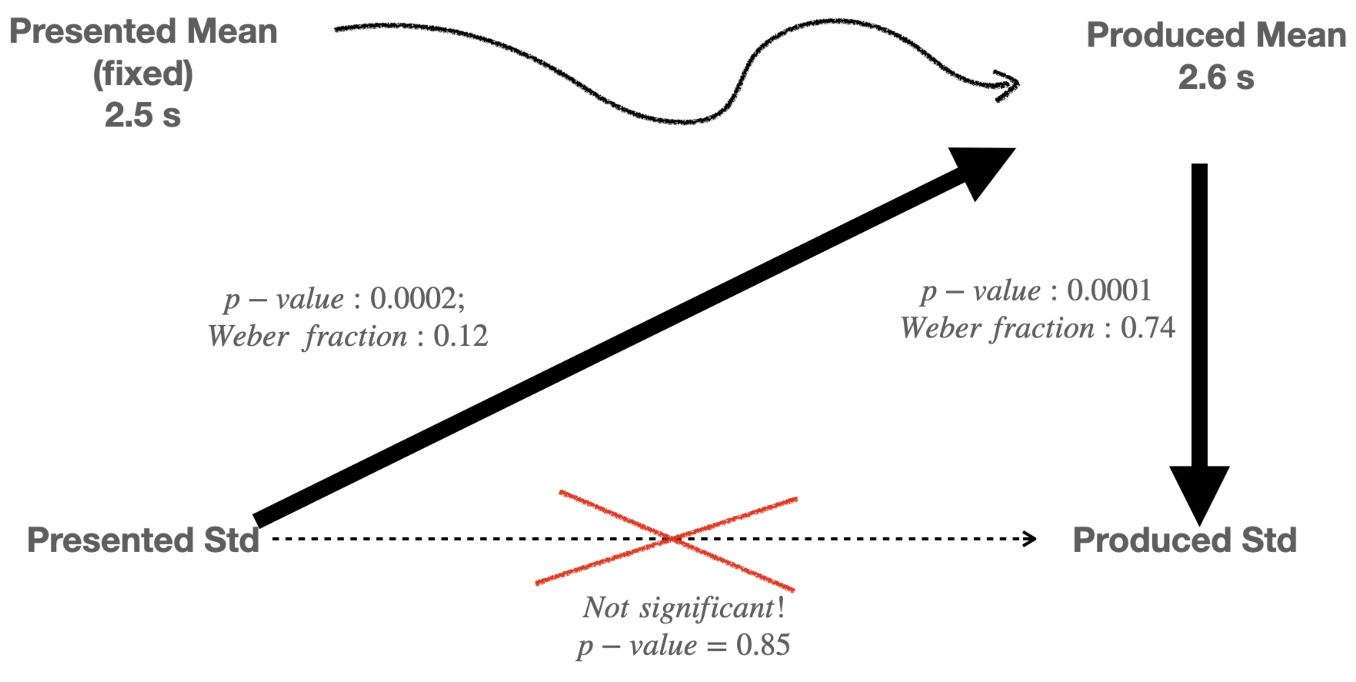


Figure 6.1: **Summary of the results for the reproduction task**
Participants were able to reproduced the same mean duration (2.6s for reproduction, 2.5 real). Against our predictions, participants were not able to directly recreate the Std of the presented foreperiods. In accordance with Weber law, the produced Std was impact by the produced mean. More interestingly, the produced mean was impacted by the presenetd std.

7

Discussion

7.1 Summary of the main results

This study was therefore trying to assess to what extent the variance is relevant. In order to do so, three levels of analysis were done until now. First, in the main task, the impact of the variance directly on the participant behavior, their reaction time was assessed. The standard deviation had shown a significant effect on the reaction time of participants. Second, still during the main task, a Bayesian Observer was trying to learn the foreperiod distribution. Two types of Bayesian observers were compared, one that learned both the mean and the standard deviation and one that learned the mean only, and for which the standard deviation was just a fraction of the mean. The best model to describe participants' behavior was the one that learned the Mean Only. Third, in order to assess to what extent participants have memorized the variance of the presented foreperiods, they were asked to reproduce the last series of foreperiods that they heard. The produced std was not affected by the presented std directly.

7.2 Discussion about the variance directly on the behavior

In accordance with literature, [7],[16],[8], the variance of the foreperiod distribution had an effect on participants' behavior. Yet, the unclear question remained : to what extent is it really the variance that matter. Indeed, necessarily, when the variance is huge then there is more short and long foreperiods. Yet, the duration of one foreperiods also has an effect. By looking just at the behaviour it is therefore impossible to disentangle the effect of the standard deviation of the distribution and the effects of the foreperiods duration.

7.3 Discussion about the Bayesian Observers

Since, the Weber fraction is a independent parameter it can be optimized freely. Finding Weber fraction for which the behavioural data is better explained by the Mean Only Observer compared to the Mean/Std parameter thus challenges the classical hypothesize that Bayesian Models had to learn both the mean and the std. Interestingly the range of Weber fractions for which the Mean Only Observer is better than the Mean/Std Observer is similar to what is usually found in the literature when participants are asked to measure or create different magnitudes [11]. Therefore, this first implicit task goes against classical literature. Indeed, it shows that participants' behaviors are better explained by a model that does not learn perfectly the underlying distribution, and learns just the mean, from which, via an heuristic, the Weber Fraction, it can create a distribution of the foreperiods.

7.4 Discussion about the explicit task

The result shows that maybe participants would encode the variance directly in the mean and from this mean create a variance directly. This is quite interesting because once again it shows that we are not perfectly learning the distribution, here Gaussian, represented by two "independent" parameters, the mean and the standard deviation, but rather just create one big parameter, a mean, from which we can create a distribution. This goes along with the findings of the first experiment where it has been shown that the participants are better explained by an observer that learns the Mean Only, for which the Standard Deviation is just a function of the mean, compared to the observer that learns both the mean and the standard deviation.

Idea : maybe not compare only to the last block and define a discount factor and use a discounted mean of the last updates of the Bayesian Observers, or of the blocks.

If the variance of the presented foreperiods series was correlated to the variance of the produced series of foreperiods, this would have show that the variance has been learned, memorized, in the auditory cue and then recruited for a motor command. This would have meant that there is communication between the different regions (the participants perceive a sound and execute a motor command). In other words, this statistical distribution would be stored in a higher order memory than just auditory memory and therefore our brain would be confronted with a cross-modal parameter. This would go along with the hypothesis of a global prior, that could even be even at the heart of the explanation of illusions [11]. In current studies the cross modality means that there is a transfer from one sensory area to the other, for instance from visual to auditory. Yet, in our experiment the change of modality is a transfer from an implicit timing task to an explicit timing task.

Since there was no direct effect, the remaining question is therefore, have we store a distribution somewhere and it is to hard to recruit or have we never build this distribution ?

7.5 Discussion about the interaction between the two task

In order to have an effect of the std (i.e. the Mean/Std Observer better explain the RT than the Mean Only) in the implicit timing task (when we present the foreperiods and the participant only have to answer as quickly as possible) the participant has to learn that the standard deviation is meaningful in order to adjust his surprise accordingly. In order to have an effect in the reproduction task of the std, the participant has to have learned that the std is meaningful (actually during the presentation task) and then recruit this internal representation for the motor command. Therefore the reproduction task, which is an explicit timing task, is somehow built on top of the presentation task, which is an implicit timing task. As a consequence, there could be an effect of the std in the implicit timing, which does not transfer to explicit timing.

In other words, it was not very likely to have the effect in the reproduction task and not in the implicit timing task of the variance. Indeed, for the variance to be significant in the implicit timing task it only need the internalization, whereas in the reproduction task it uses both the internalization and the recruitment of the internalized memory for the motor command. In a nutshell, because the explicit timing task is built on top of the implicit timing task, there could be an effect of the std in implicit timing, which does not transfer to explicit timing, however it would have been odd to have an effect in the explicit timing task and not in the implicit timing task.

7.6 Importance of the study for clinical application

Reproductions of a single stimuli already show an inherent variability [11]. Understanding the importance of the variability in temporal structures on motor responses remains therefore an unanswered question. Distortions of time estimation are found in numerous psychiatric diseases and conditions : ADHD, autism, schizophrenia, Parkinson's. The case of schizophrenia is of particular interest because several contemporary theories of this disease postulate that a disturbance in the computation of Bayesian inference is at the core of the abnormality of this condition [17]. Defining an indicator and creating a model that assess the behavior of neurotypical subjects might help diagnose earlier pathological conditions, and is the first step to measure the efficiency of upcoming treatments to cure these pathologies.

7.7 Future work

7.7.1 More participants

As the result is quite novel and goes against the literature (even though until now it was never tested, it was just hypothesized) more participants are planned to be incorporated into the study (10 more at least), in order to gain statistical power.

7.7.2 Use of 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.

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 [9] 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

$$h(FP_n) = P(FP_n > t + dt | FP_n > t) \tag{7.1}$$

Or equivalently :

$$h(FP_n) = \frac{f(FP_n)}{1 - F(FP_n)} \tag{7.2}$$

where f is the distribution and F the cumulative distribution.

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 :

$$I_{HF}(FP_n) = -log(h(FP_n)) \tag{7.3}$$

where f is the distribution and F the cumulative distribution.

7.7.3 Pupillary Data

It would be still possible that the variance impact responses at a lower biological scale than the "macroscopic" / "psychological" behavior of the participants. Indeed, in the literature, it has been shown that an autonomous response, the pupillary dilatation, is a good indicator of surprise [18],[19]. Indeed, the results [18] show that the pupil size was increased sooner when target stimuli were expected earlier. This autonomous response is also a clear marker of more biological signals such as noradrenaline [19]. Comparing the physiological reaction to more or less variability in the stimuli, might assess the scale at which the selectivity for the variance takes place, and therefore better understand underlying biological effects and neural mechanisms [20].

As it was mentioned during the introduction, during the experiment pupils of the participants were also followed. Because of a lack of time, these data were not fully analyzed until now.

The trajectories of the pupillary dilatation will be then compared between the different types of blocks. Yet, the pipeline for the analyses of this data is already setup. Pupillary data from the training phase and the inter block sequences will be removed. Then, a confounding on the pupillary diameter has to be taken into account. Indeed, the heads of the participants were not fixed so they could move back and forward from the screen and the eye tracker. This naturally makes variation on the pupillary diameter. Yet, this artifact can be easily corrected. First, the inter-eye distance can be computed from and this is a very good estimator of the distance between the participants and the screen. Then, the raw pupillary data can be regressed on the inter-eye distance, giving therefore, from the residual of that regression, a measure of the pupillary dilatation corrected from the screen distance artifact. Then, blinks can be interpolated using a cubic spline interpolation [21]. This method gives an interpolation method that is smoother and has smaller error than some other interpolating polynomials such as Lagrange Polynomials and Newton Polynomials. Once this preprocessing is done, the pupillary data will be cut into epochs centered on the target tone. These epochs will play the role of "Event Related Pupillary Dilatation", depicting how the target tone had an impact on ocular responses. These Pupillary Epochs will be then compared between the different standard deviations in order to study impact of the predictability of the stimuli at this autonomous scale.

7.7.4 Assessing participants subjective experience

To know if this recruitment is conscious or not participants were asked at the very end if they were aware of the variability and tried to reproduce it or not. This task of a reproduction of a series of foreperiods is new to the literature. Thus, for the moment, it is unclear if the building of a representation of the variance and recruitment of this representation as a memory during the reproduction task is conscious or not. Some Natural Language Processing (NLP) techniques might be used in order to cluster he responses of participants.

7.8 Open Questions

7.8.1 Discovering new modalities

However, a big question and important choice remain for this section : should the cue tone be the same as the last block ? If so, what does it change if we change the cue of the reproduction section ? What happens if we give a totally new cue ? A new frequency for the cue tone ? A visual cue instead of an auditory cue ? If we use the tone of another block will that be influenced ? Are participants aware that it was the tone of a block that they have heard ?

Indeed, the study of categorization is mainly done in the visual field. Yet, auditory perception is better to track time, whereas visual perception is more suited for categorization[1]. Thus, it might be interesting to select different types of stimuli, being auditory or visual and compare them.

This raises the question of the importance of the cue tones ? To what extent the participants remember the different tones, and to what extent this has an effect on their behavior ? In other terms, is there a cross modality association and memory?

This question goes along with an even more fundamental question which is : to what extent our findings on the (not so big) importance of the variance could be transposed to other modalities !? These results being completely novel, knowing how the variance of stimuli affect our senses and more importantly to what extent do we make the effort to learn a complex parameter such as the variance is a completely open question.

7.8.2 How are probabilities stored in our brain ?

Finally, these studies have challenged the current hypothesis that participants could learn perfectly the underlying distribution of a series of foreperiods. The big problem is that we only have a proxy of their predictions. Thus, we might never know if humans can represent a bell shaped probability density function somewhere in their head. Indeed, our study will not answer the question of how the distribution is stored i.e. if the distribution is stored as a function directly or only the parameters are stored [17].

A

... as Acknowledgements

I would like to finish this report by expressing my immense gratitude to my supervisor, Dr. Sophie Herbst. From the beginning of this project, in October 2021 to currently in the end of June 2022, she always supported my ideas and encouraged me in the discovery of new research question. This experience at Neurospin under her supervision really settled my fascination and passion for the field of research in cognitive neuroscience. Indeed, she trusted me enough to let me define and conduct from scratch, starting with the definition of the current problematic, conducting the experiment, and finally analyses the data. It has been an honor and pleasure to work by her side.

On the same note, I want to thanks the Principal Investigator of the Cognition and Brain Dynamics team, Dr. Virginie van Wanssenhove to have let meet be part of this team and give the necessary fundings for conducting my research.

Naturally, I thanks the entirety of the CBD team, the post docs, the research engineer, the Phd Student and the other master student. Particularly, I would like to thanks the two other students from the Computational Neuroscience and Neuroengineer Master, (futur-Dr) Andres Torrez Sanchez to have taught me dancing (since it is also a canonical example of temporal preparation). Also, (futur-Dr) Anna Razafindrahaba, for all her relevant remarks on the project, her profoundly positive mood, containing the virtues of joy, a good nature, as well as peace and love to an intense degree.

Then, I would like to thanks Dr. Antoine Chaillet and Dr. Sabir Jacquir, the two supervisor of the Master Computational Neurosciences and Nueorengiennering that I have pursued during my gap year at ENSAE, where I discover the fascinating field of neuroscience.

Also, I want to express a profound gratitude for my two mentors at the CREST : Dr. Guillaume Hollard, for all the support and encouragement he gave me to continue in the field of behavioral sciences, a field that I discovered thanks to Dr. Felix Tropf.

Finally, I would like to warmly thanks my entire family for all of their support during my studies and once again during this year.

Napunak

$\mathcal{B}$

… as Bibliography

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] Van Ede F. Nobre A. C. 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] Engel A. K. Wagenmakers E.J. Donner T. H. Meindertsma T., Kloosterman N. A. 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] H. Woodrow. The measurement of attention. *Psychological Review Publications*, 1914.

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

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

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

[11] Stephan KE. Petzschner FH, Glasauer S. A bayesian perspective on magnitude estimation. *Trends Cogn Sci.*, 2015.

[12] Dudman J.T. Yi Li. Mice infer probabilistic models for timing. *Proceedings of the National Academy of Sciences*, 2013.

[13] Niv Y. Gershman S. Perceptual estimation obeys occam's razor. *Frontiers in Psychology*, 2013.

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

[15] G.T. Fechner. *Elemente der Psychophysik*. Breitkopf und Hartel, 1860.

[16] Klemmer E.T. Time uncertainty in simple reaction time. *Journal of Experimental Psychology, Vol. 51, No 3*, 1956.

[17] Ma Wei Ji Latham P.E Pouget A., Beck J.M. Probabilistic brains: knowns and unknowns. *Nature Neuroscience*, 2013.

[18] Rijn H. Akdogan B., Balci F. Temporal expectation indexed by pupillary response. *Timing Time Perception*, 2016.

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

[20] H. Meck C.V. Buhusi, W. What makes us tick ? functional and neural mechanisms of interval timing. *Nature Neuroscience*, 2005.

[21] A. C. Nobre N. Shalev. Eyes wide open: Regulation of arousal by temporal expectations. *Cognition, V. 224*, 2022.

[22] John Gibbon. Scalar expectancy theory and weber's law in animal timing. *Psychological Review*.

[23] K.H. Norwich. On the theory of weber fractions. *Perception and Psychophysics*, 1987.

[24] Brovelli A. Girard B Viejo G., Khamassi M. 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.

C

...as Colors & Figures

List of Figures

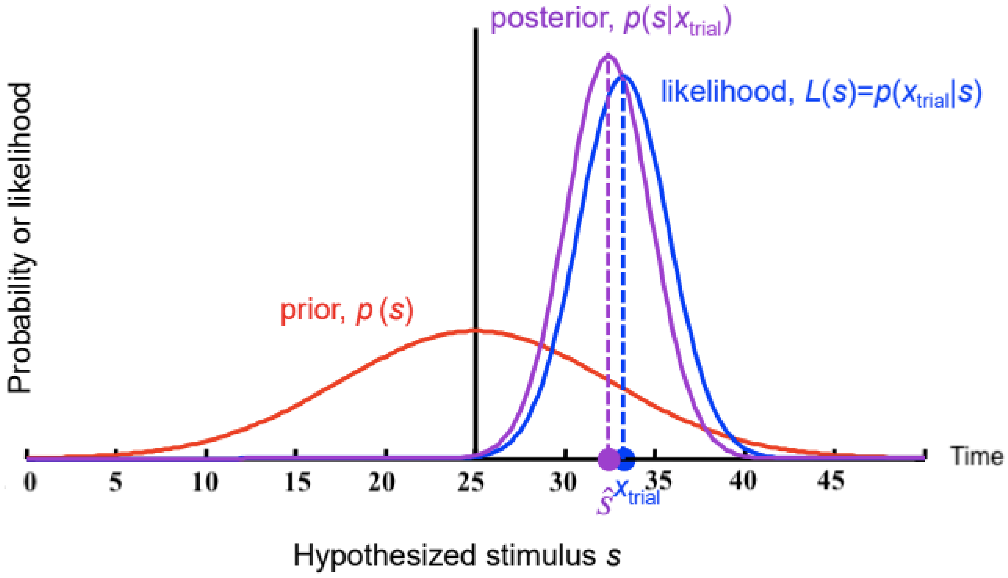


Figure C.1: **Rational behind Bayes Rule**
Multiplying the Prior Distribution, $p(s)$ which is red, with the Likelihood,$L(s)$ in blue, give the Posterior Distribution, $p(s|x_{trial})$ in purple. This can be seen as a weighted average of what we just learned, the likelihood, with weights provided by our previous knowledge, the prior.

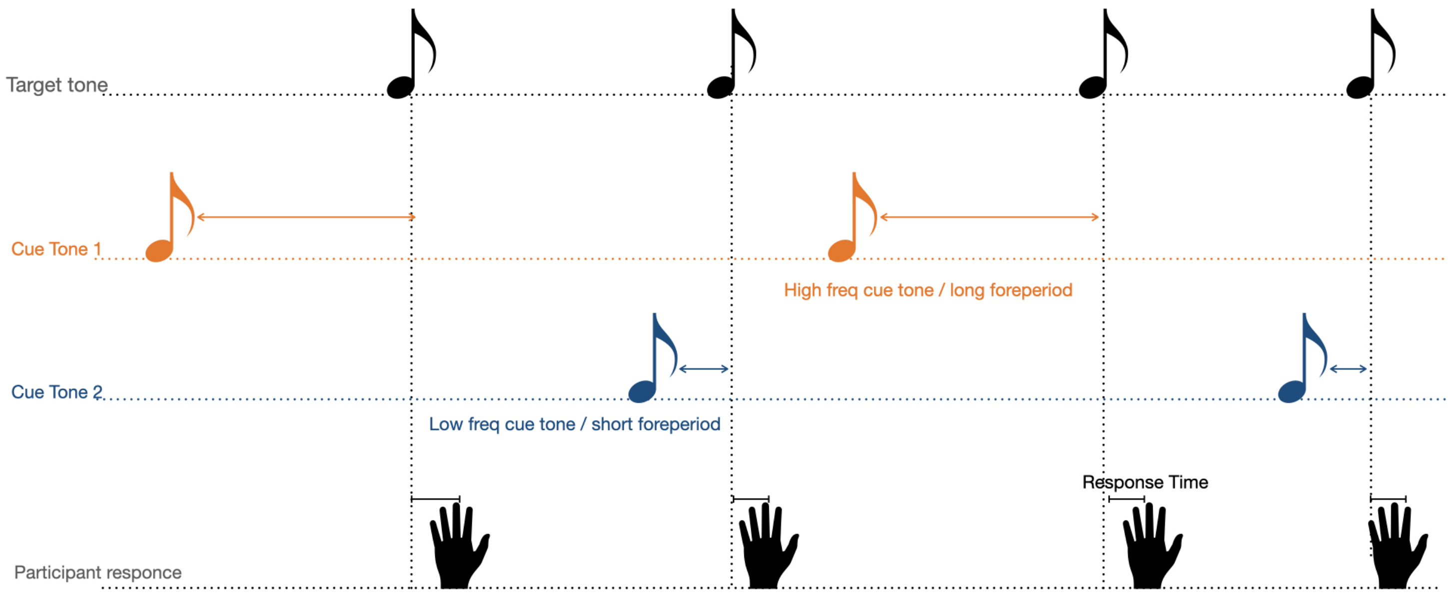


Figure C.2: **Representation of the experimental paradigm of [6]**
This paradigm was based on a foreperiod paradigm as depicted in 2.1. The difference here was that during one block there were two distribution of foreperiod that were shown to the the participants. Depending on the tone of the cue tone, the distribution of the foreperiods was different. The Response Time of participants was recorder as a proxy of their temporal preparation.

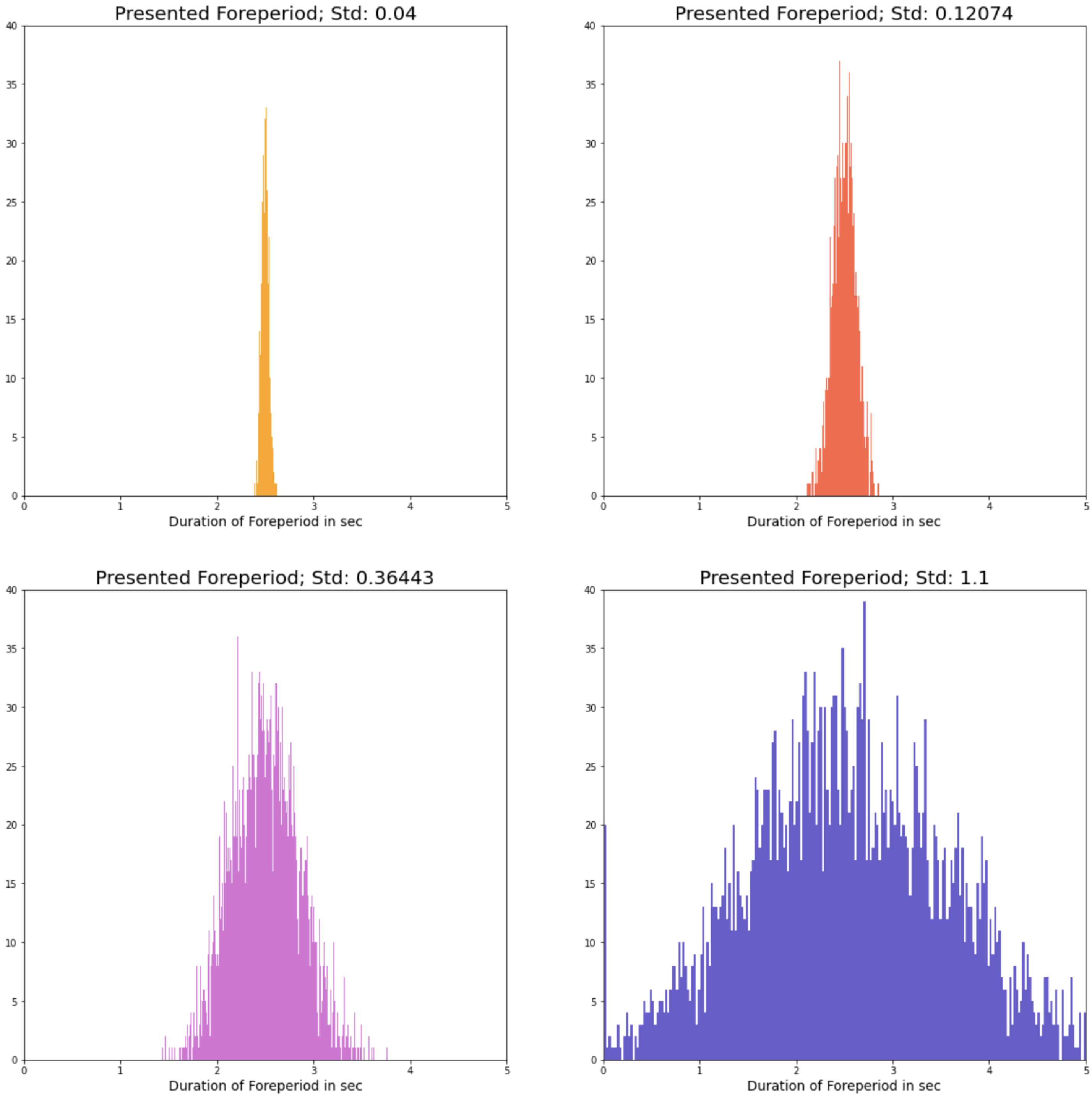


Figure C.3: **Distribution of the Foreperiod for the Current Study**
During the new behavioral paradigm of this study, the mean of the distribution of the foreperiods remained the same, always at 2.5 sec. However, the blocks, appearing in a randomized order for the participants, were more or less predictable. Indeed, the Std of the distribution could be very small (panel A, orange), resulting in a narrow distribution, or very huge (panel D, pink), resulting in a broad distribution.

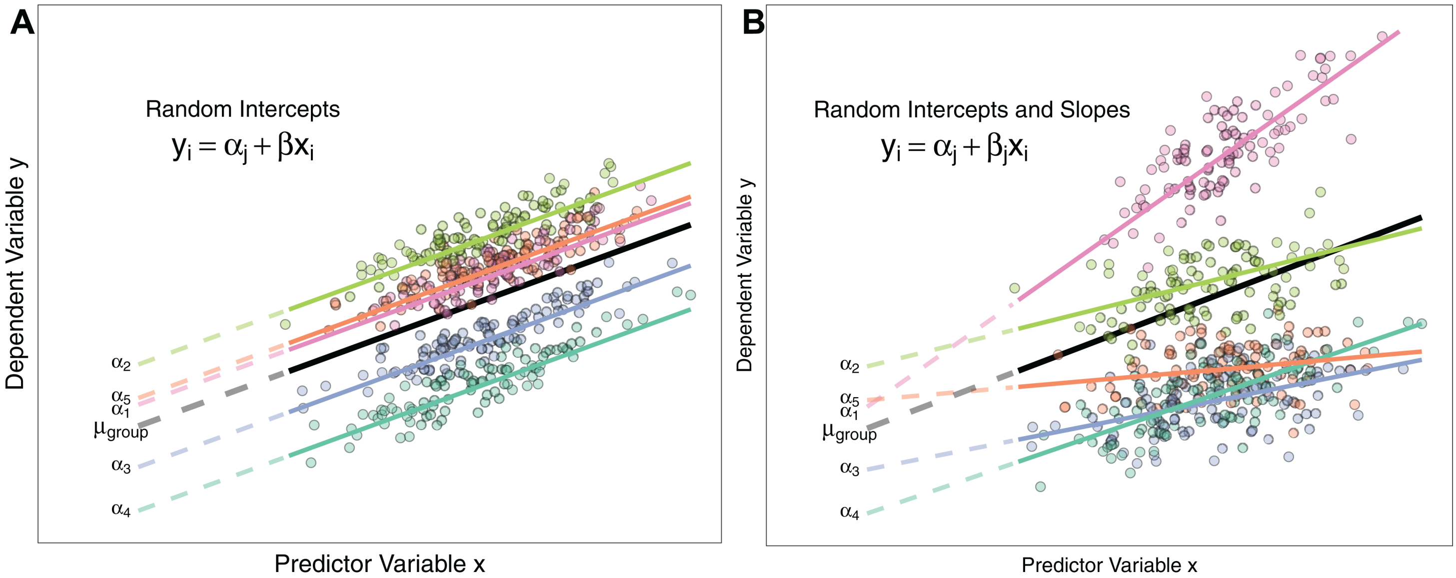
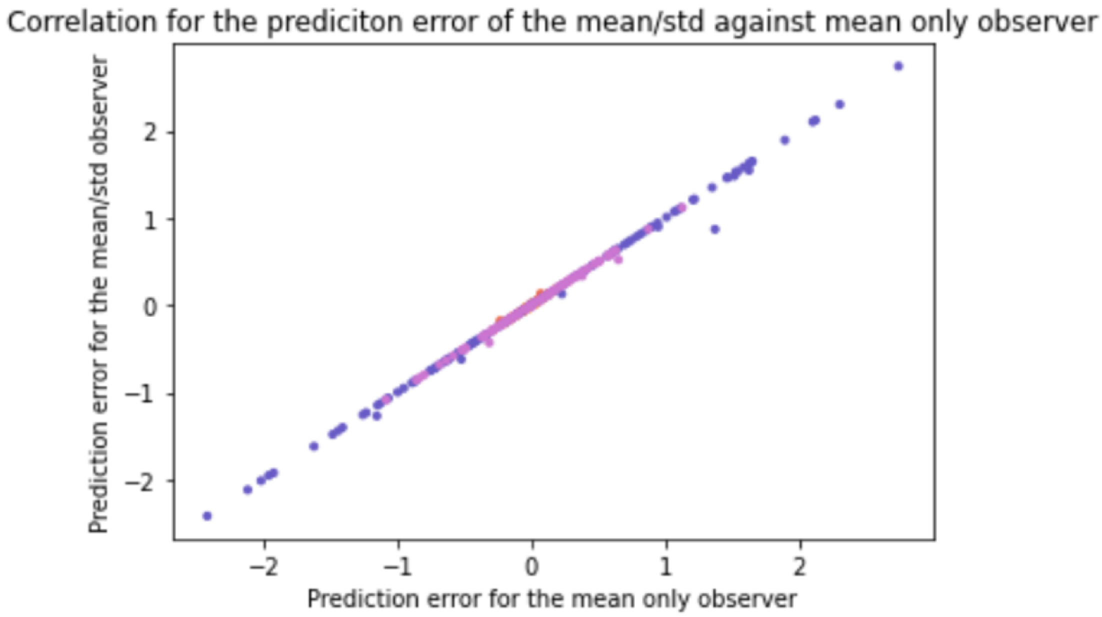
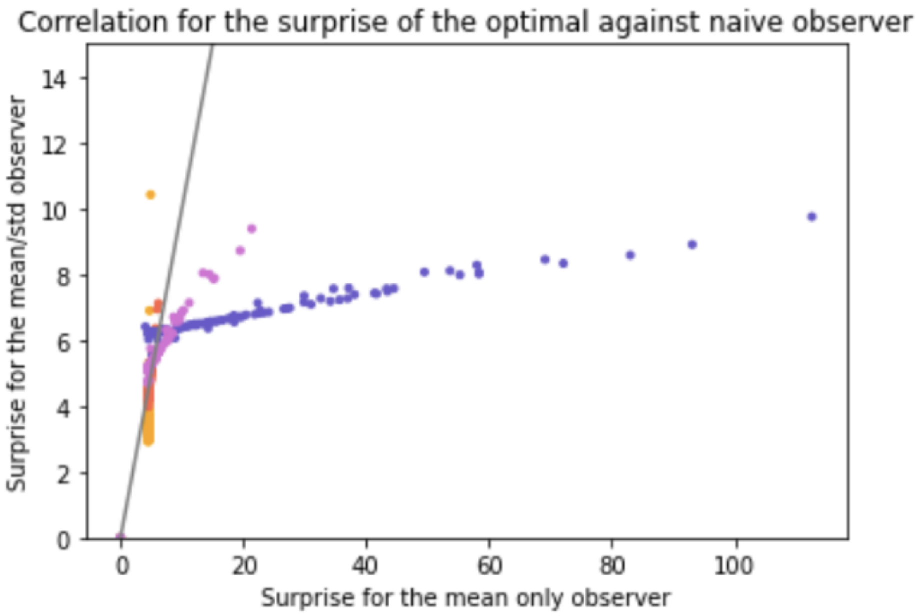


Figure C.4: **Different type of Linear Mixed Effect Models**
Linear Mixed Effect Models are regression that try to captures inter-group variability. Therefore, as it is depicted on the left panel, the linear model can have one intercept per group (generally one group is one participant). Moreover, as it is shown on the right panel, there could be also different intercept and slope for each group. *Adapted from Google*




```
Prediction correlation and p-value 0.9988397354336918 0.0
```




```
Surprise correlation and p-value 0.6150728074908386 6.713678475885793e-46
```

Figure C.5: **Correlation for the Prediction Errors and the Surprise for the 2 Observers** The upper panel depicts the correlation of the Prediction Errors. The correlation is really high. The lower panel depict the correlation for the Surprise of the two observers. Here what is interesting is that the correlation between the surprise depends on the Std of the Foreperiod Distribution. This results in a multicolinearity.

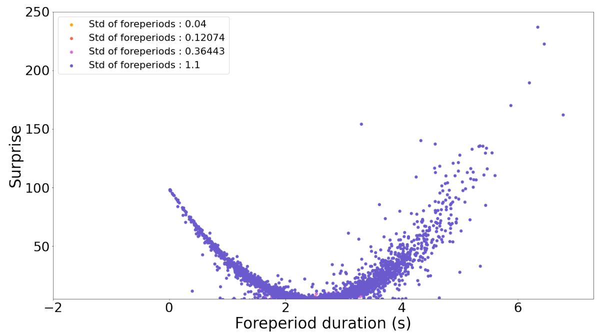


Figure C.6: Phoenix Plot for the Mean Only Observer

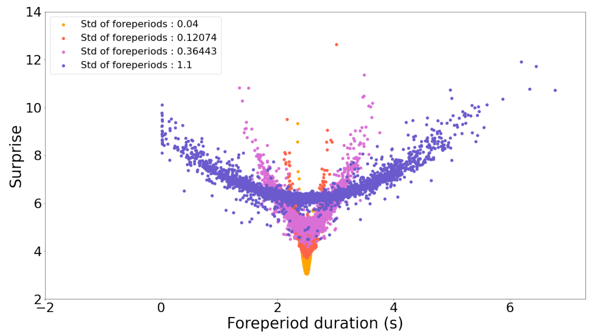


Figure C.7: Phoenix Plot for the Mean Std Observer

Figure C.8: **Evolution of the Surprise with the Foreperiod : Phoenix Plot**
What can be found here is that the surprise for the Mean STD observer is re-scaled for each block. Indeed, the mean surprise for each block is growing with the standard deviation but the range of it is not exploding, always remaining around 5 +/- 2 . Indeed, this observer is learning the distribution and the standard deviation. Therefore, even in the unpredictable blocks it won't be surprised because it knows that foreperiods can be really short or really long. On the other hand, the mean only observer does not realize that there are different distributions. Thus it's surprise just explodes with the standard deviation of the latent distribution of the foreperiods.

Model:	MixedLM	Dependent Variable:	Log_RT
No. Observations:	12158	Method:	REML
No. Groups:	29	Scale:	0.0954
Min. group size:	345	Log-Likelihood:	-3046.6224
Max. group size:	432	Converged:	Yes
Mean group size:	419.2		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	-1.771	0.037	-47.246	0.000	-1.845	-1.698
Mean_std_pred	-0.068	0.005	-13.845	0.000	-0.078	-0.058
Innovation_pred_mean_only	0.038	0.092	0.420	0.675	-0.141	0.218
Group Var	0.041	0.035				

Figure C.9: Regression with the mean only innovation

Model:	MixedLM	Dependent Variable:	Log_RT
No. Observations:	12158	Method:	REML
No. Groups:	29	Scale:	0.0953
Min. group size:	345	Log-Likelihood:	-3046.4087
Max. group size:	432	Converged:	Yes
Mean group size:	419.2		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	-1.771	0.037	-47.246	0.000	-1.845	-1.698
Mean_only_pred	-0.067	0.005	-13.804	0.000	-0.077	-0.058
Innovation_pred_std	-0.120	0.092	-1.298	0.194	-0.301	0.061
Group Var	0.041	0.035				

Figure C.10: Regression with the mean std innovation

Figure C.11: **Importance of the Innovation** The difference between the prediction errors of the two types of observer show no significant effect, independently of the direction of the regression.

Model:	OLS	Adj. R-squared:	0.405
Dependent Variable:	produced_std	AIC:	132.2490
Date:	2022-05-19 16:20	BIC:	139.4189
No. Observations:	31	Log-Likelihood:	-61.125
Df Model:	4	F-statistic:	6.103
Df Residuals:	26	Prob (F-statistic):	0.00133
R-squared:	0.484	Scale:	3.6021

	Coef.	Std.Err.	t	P>|t|	[0.025	0.975]
real_mean(intercept)	-2.0246	4.1494	-0.4879	0.6297	-10.5537	6.5045
real_std	-0.1323	0.8644	-0.1530	0.8796	-1.9090	1.6445
mean_ITI	-0.0988	2.4859	-0.0397	0.9686	-5.2087	5.0111
std_ITI	1.2775	4.5528	0.2806	0.7812	-8.0809	10.6358
produced_mean	2.2604	0.4692	4.8174	0.0001	1.2959	3.2248

Omnibus:	37.086	Durbin-Watson:	1.719
Prob(Omnibus):	0.000	Jarque-Bera (JB):	121.156
Skew:	2.481	Prob(JB):	0.000
Kurtosis:	11.317	Condition No.:	84

Figure C.12: Regression of the produced Std on everything

Model:	OLS	Adj. R-squared (uncentered):	0.069
Dependent Variable:	produced_std	AIC:	151.2435
Date:	2022-05-19 16:12	BIC:	152.6775
No. Observations:	31	Log-Likelihood:	-74.622
Df Model:	1	F-statistic:	3.283
Df Residuals:	30	Prob (F-statistic):	0.0800
R-squared (uncentered):	0.099	Scale:	7.4575

	Coef.	Std.Err.	t	P>|t|	[0.025	0.975]
real_std	1.5727	0.8680	1.8119	0.0800	-0.1999	3.3454

Omnibus:	47.787	Durbin-Watson:	1.893
Prob(Omnibus):	0.000	Jarque-Bera (JB):	220.092
Skew:	3.251	Prob(JB):	0.000
Kurtosis:	14.319	Condition No.:	1

Figure C.13: Link between the produced Std and the real Std

Figure C.14: **Trying to asses the effect of the presented Std on the produced Std** There is no significant direct effect of the presented Std on the produced Std. **The working hypothesis n°3 is not validated.**

Model:	OLS	Adj. R-squared (uncentered):	0.433
Dependent Variable:	produced_std	AIC:	125.5144
Date:	2022-05-30 18:19	BIC:	126.8466
No. Observations:	28	Log-Likelihood:	-61.757
Df Model:	1	F-statistic:	22.41
Df Residuals:	27	Prob (F-statistic):	6.24e-05
R-squared (uncentered):	0.453	Scale:	5.0014

	Coef.	Std.Err.	t	P>|t|	[0.025	0.975]
produced_mean	0.7359	0.1555	4.7334	0.0001	0.4169	1.0549

Omnibus:	52.556	Durbin-Watson:	1.863
Prob(Omnibus):	0.000	Jarque-Bera (JB):	299.072
Skew:	3.723	Prob(JB):	0.000
Kurtosis:	17.174	Condition No.:	1

Figure C.15: Effect of the produced mean on the produced std

Model:	OLS	Adj. R-squared (uncentered):	0.381
Dependent Variable:	produced_mean	AIC:	122.9965
Date:	2022-06-01 17:28	BIC:	124.3287
No. Observations:	28	Log-Likelihood:	-60.498
Df Model:	1	F-statistic:	18.26
Df Residuals:	27	Prob (F-statistic):	0.000214
R-squared (uncentered):	0.403	Scale:	4.5713

	Coef.	Std.Err.	t	P>|t|	[0.025	0.975]
real_std	3.3713	0.7889	4.2735	0.0002	1.7526	4.9899

Omnibus:	5.284	Durbin-Watson:	1.169
Prob(Omnibus):	0.071	Jarque-Bera (JB):	3.639
Skew:	-0.835	Prob(JB):	0.162
Kurtosis:	3.572	Condition No.:	1

Figure C.16: Effect of the presented std on the produced mean

Figure C.17: **Trying to asses the indirect effect of the presented Std on the produced Std** The results on the left panel show a classical Weber fraction fining. More interestingly, the right panel show that the produced mean is impacted by the presented std. There is therefore an indirect effect of the presented std on the produced std via : PresentedStd -> ProducedMean -> ProducedStd

D

… as Diapositives d'instructions

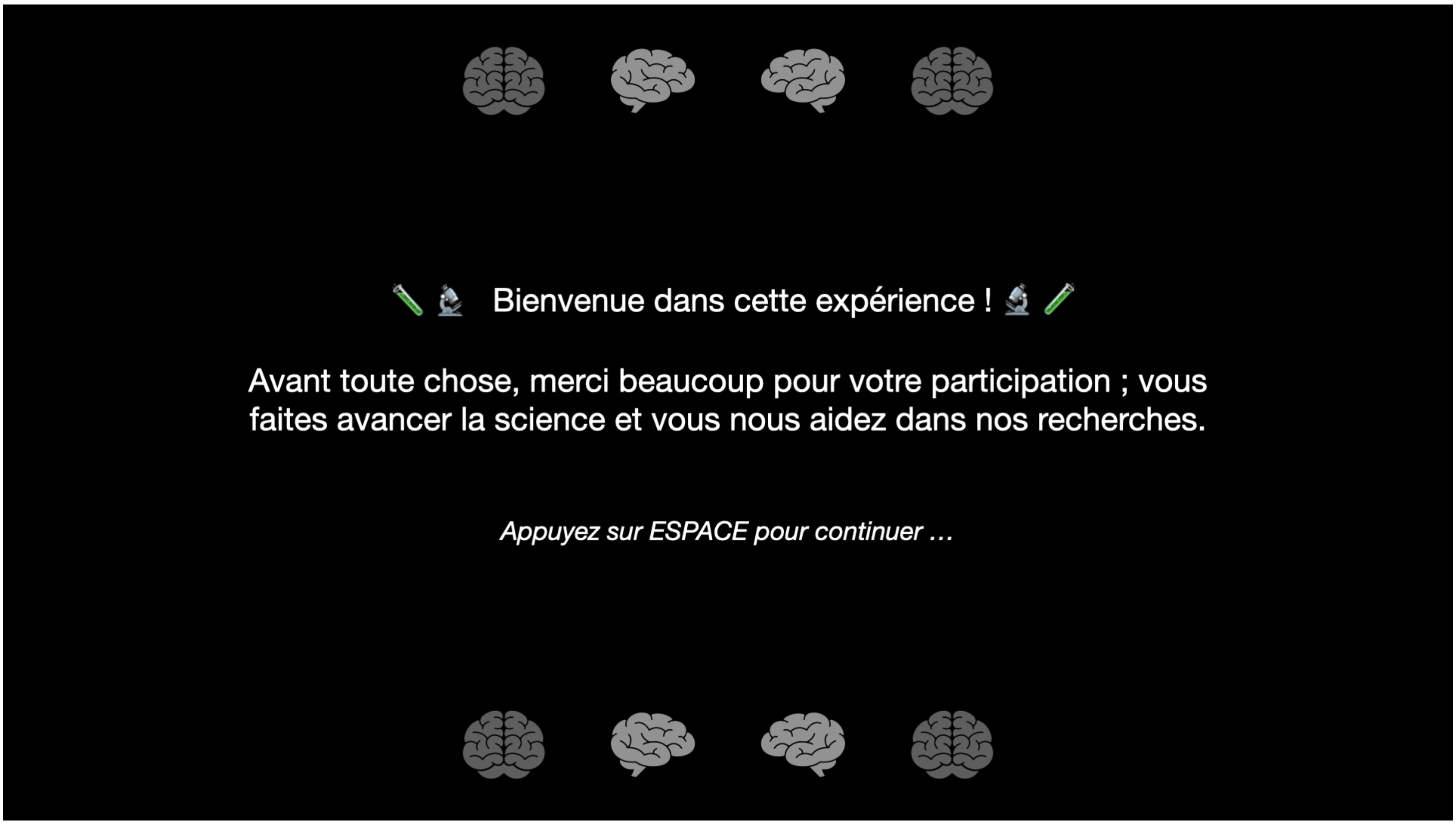


Figure D.1: Instruction n°1,a for the participant, *this was the initial screen in the room*

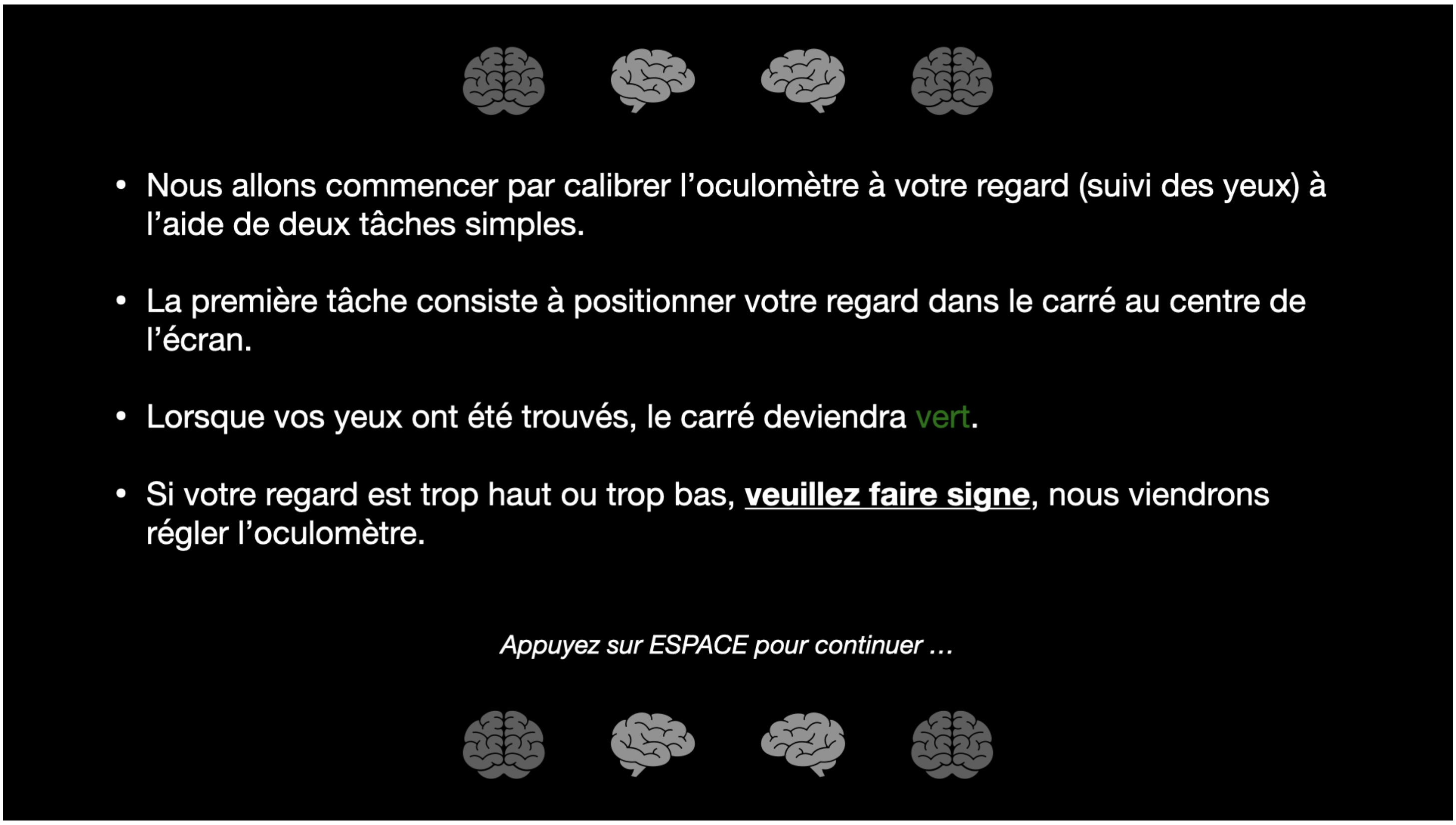


Figure D.2: Instruction n°1,b for the participant, *Start of the experiment with the calibration*

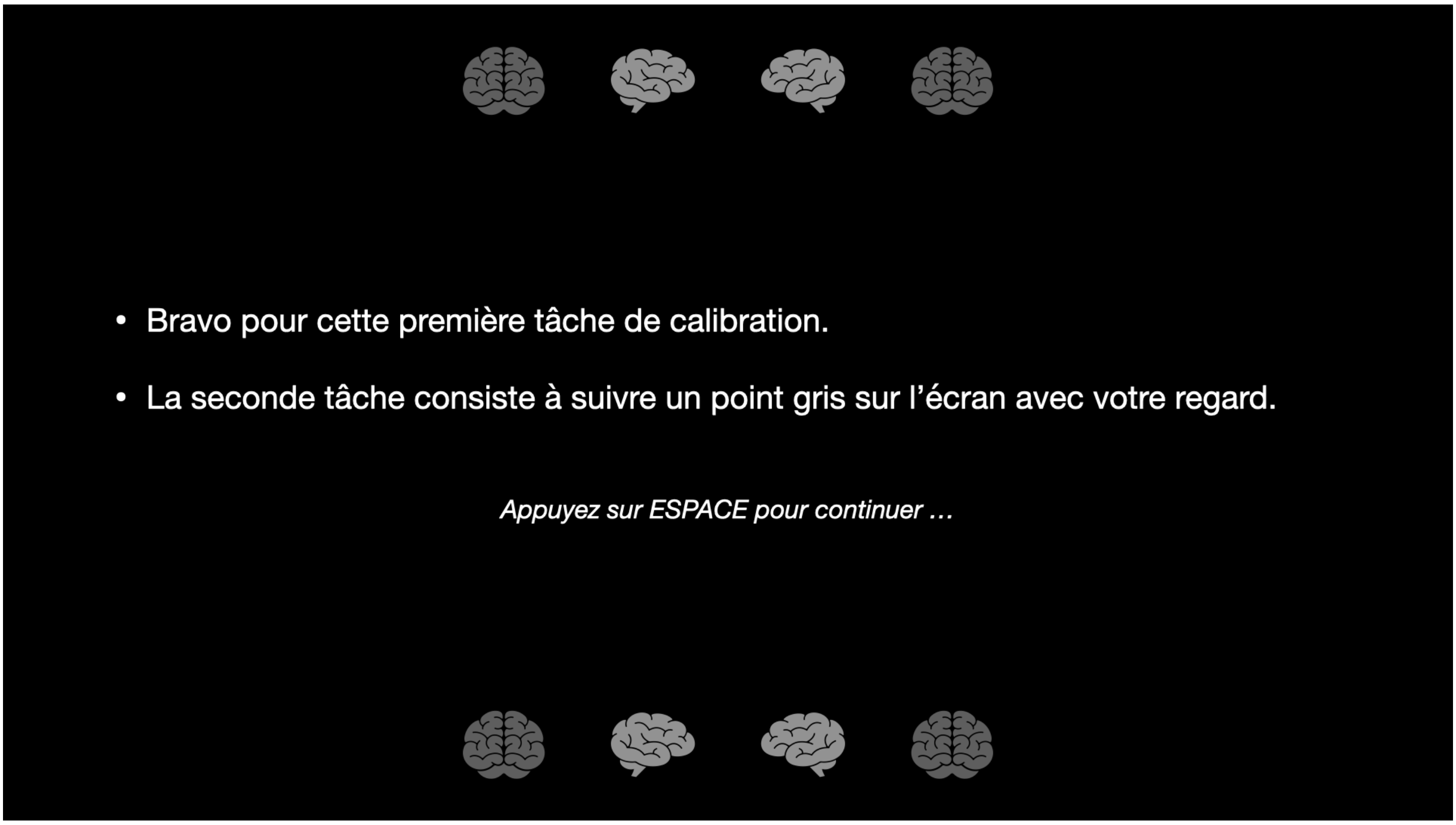


Figure D.3: Instruction n°1,c for the participant, *there was a second task for the calibration of the eye tracker*

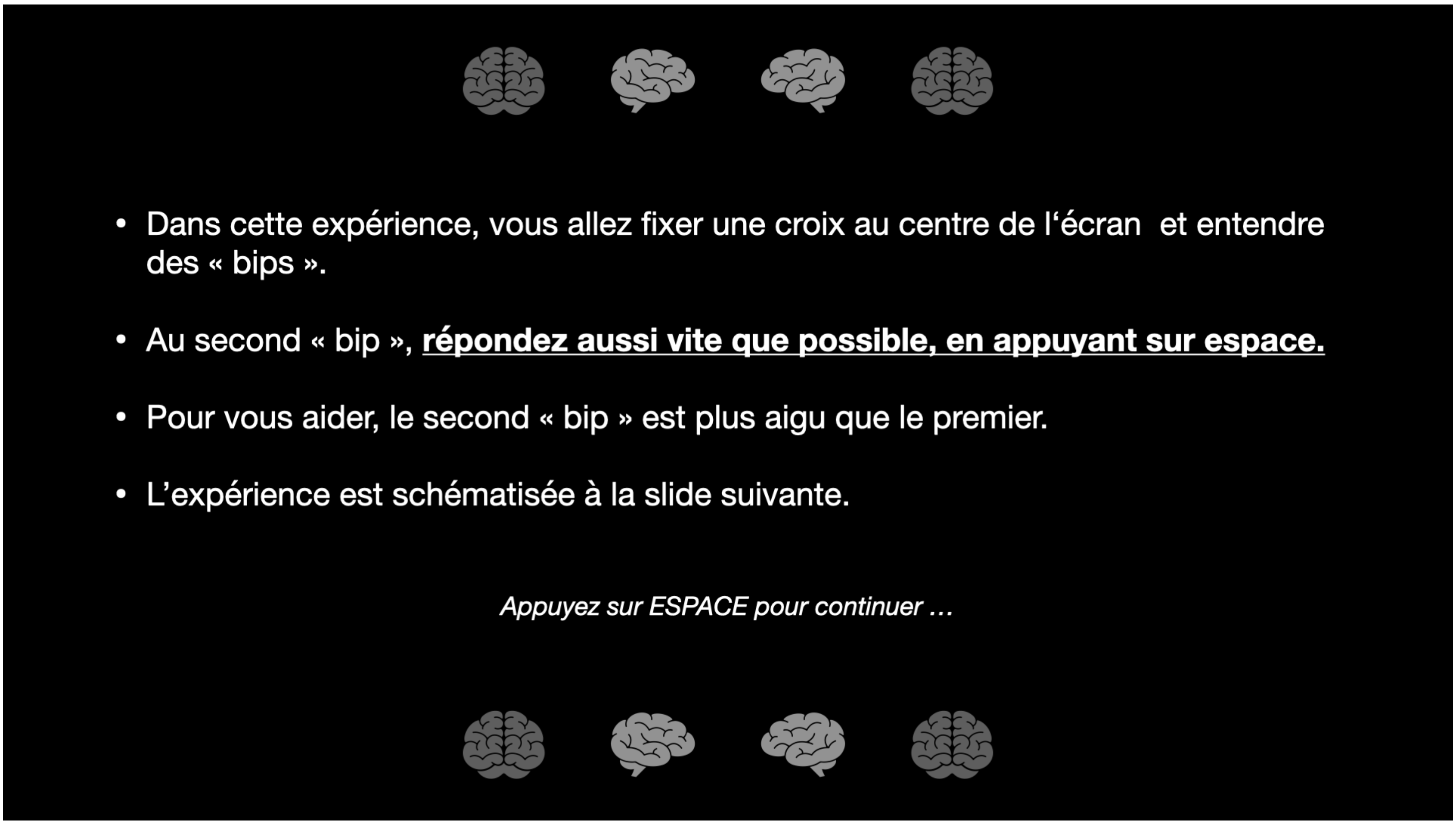


Figure D.4: Instruction n°2 for the participant

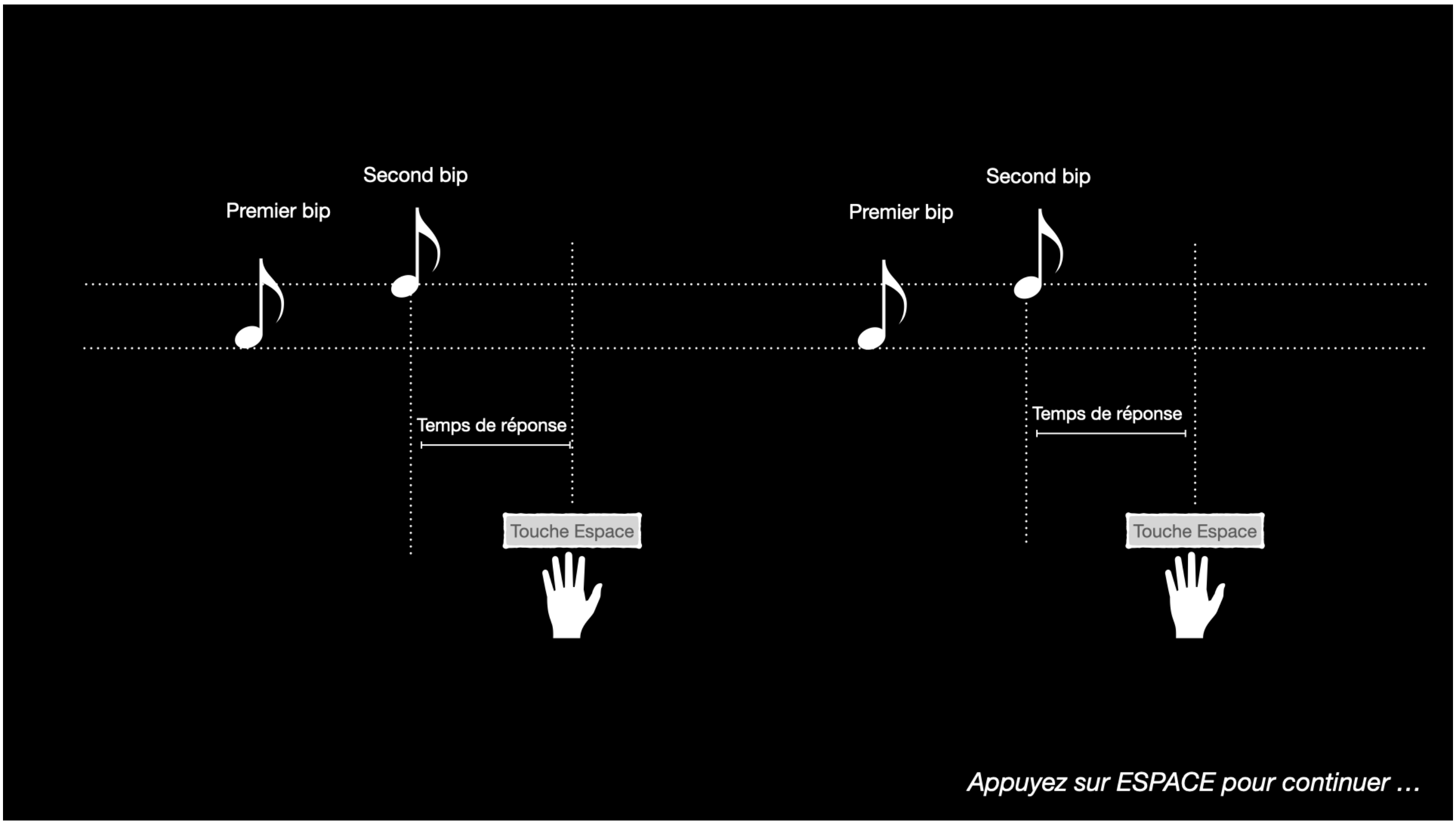


Figure D.5: Instruction n°3 for the participant

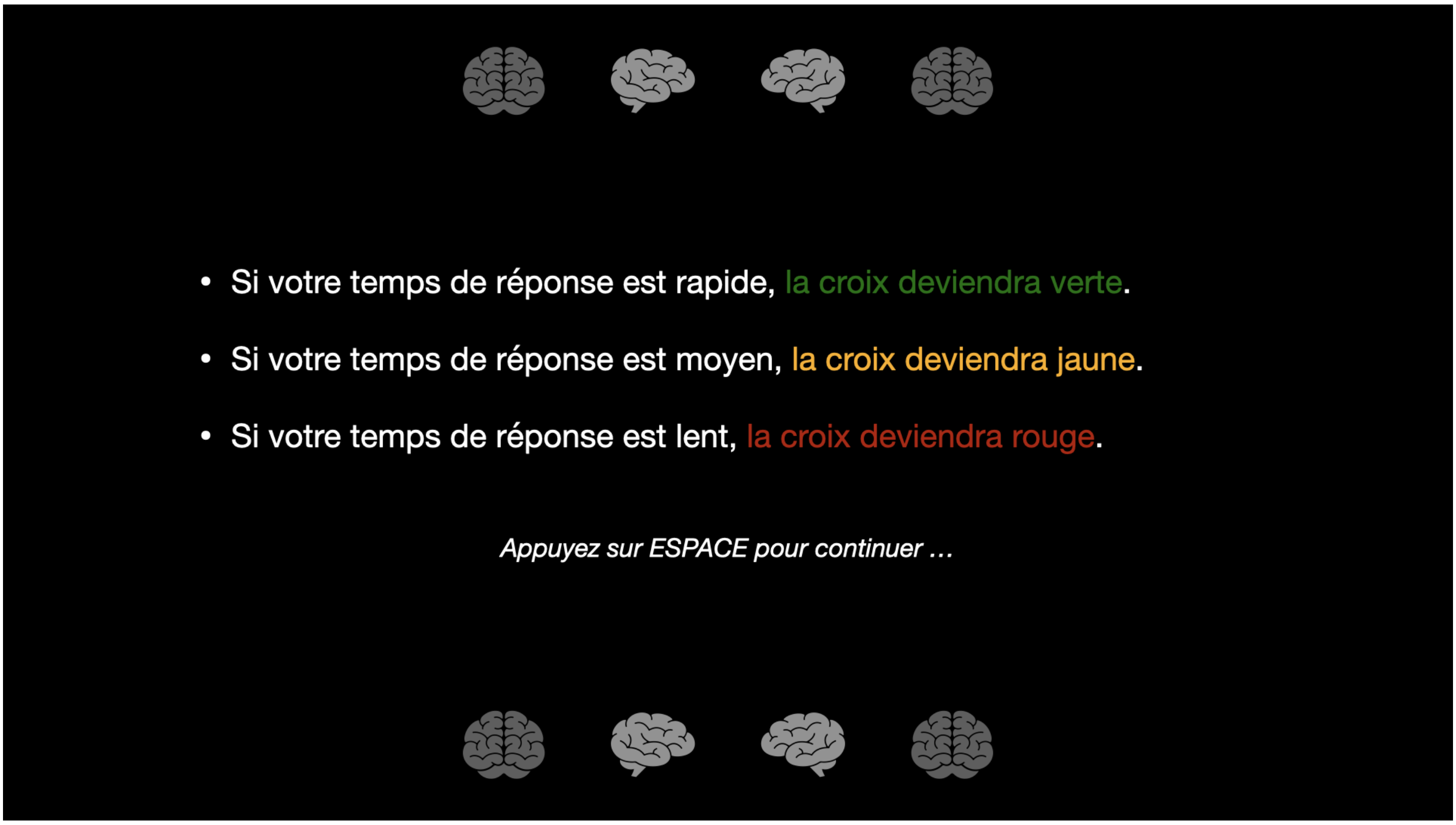


Figure D.6: Instruction n°4 for the participant

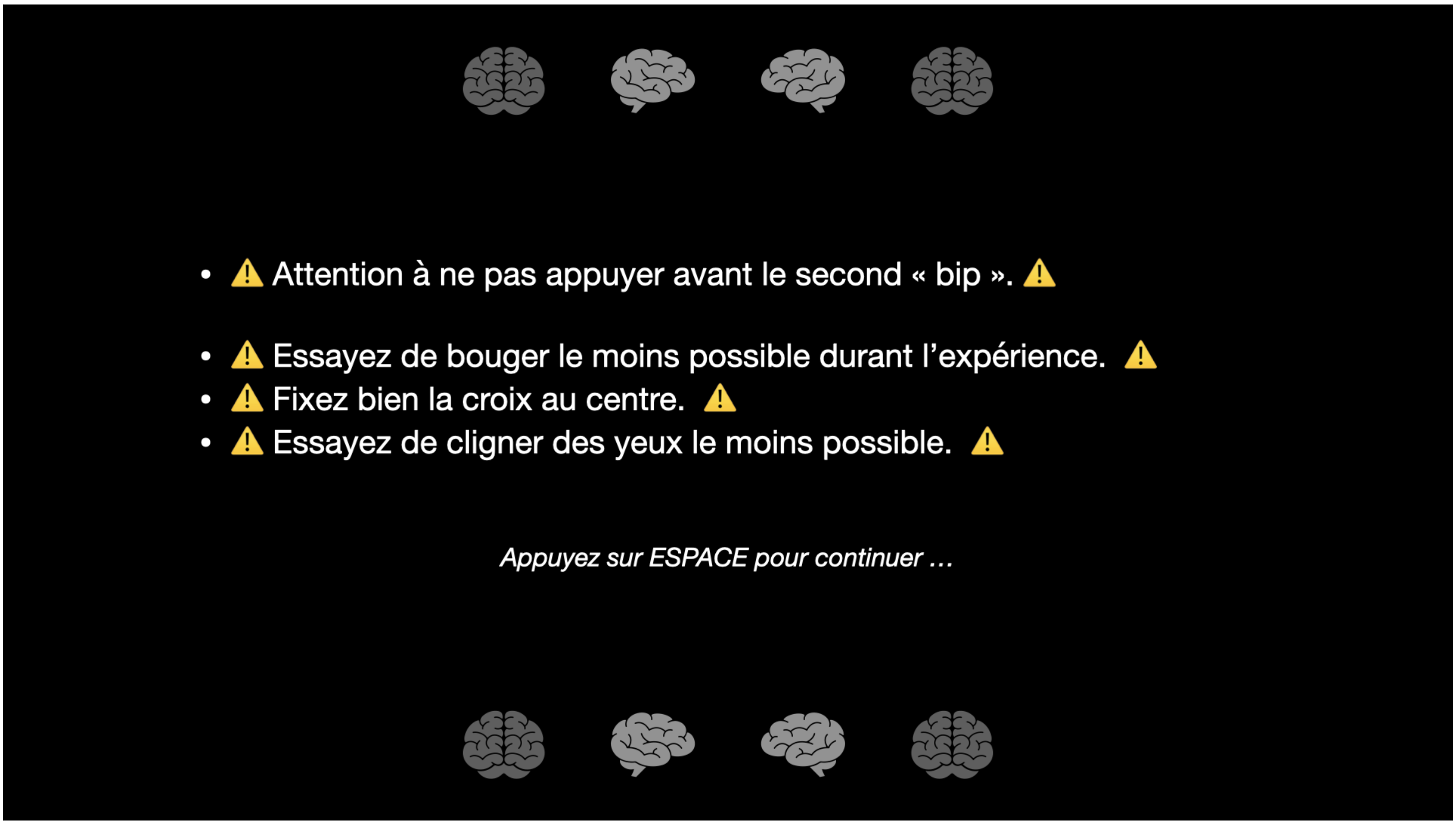


Figure D.7: Instruction n°5 for the participant

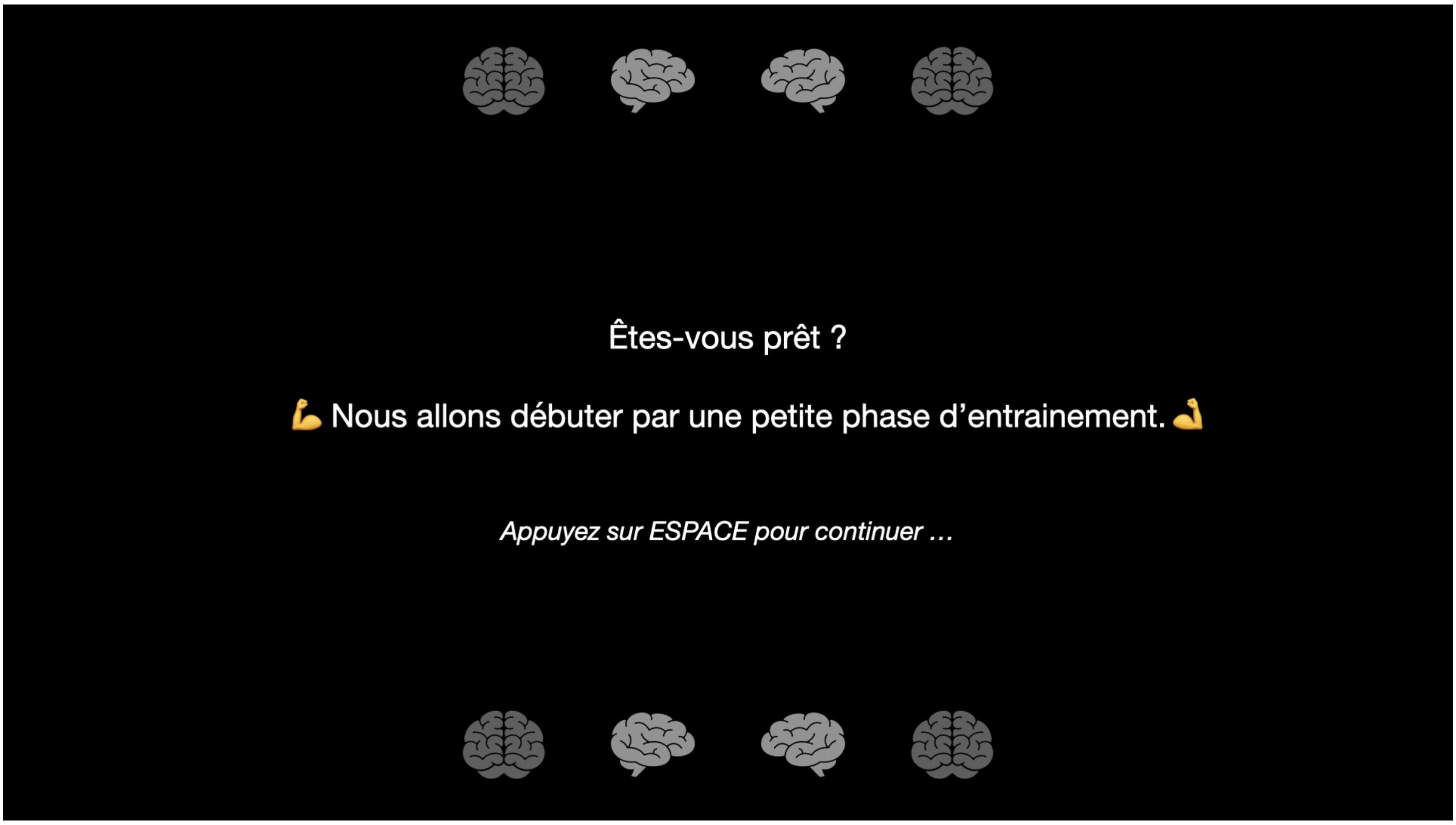


Figure D.8: Instruction n°6 for the participant

Figure D.9: Instruction n°7,a for the participant, *This slide appeared if the participants pressed too early*

Figure D.10: Instruction n°7,b for the participant, *This slide appeared if the participants didn't pressed, or waited too long*

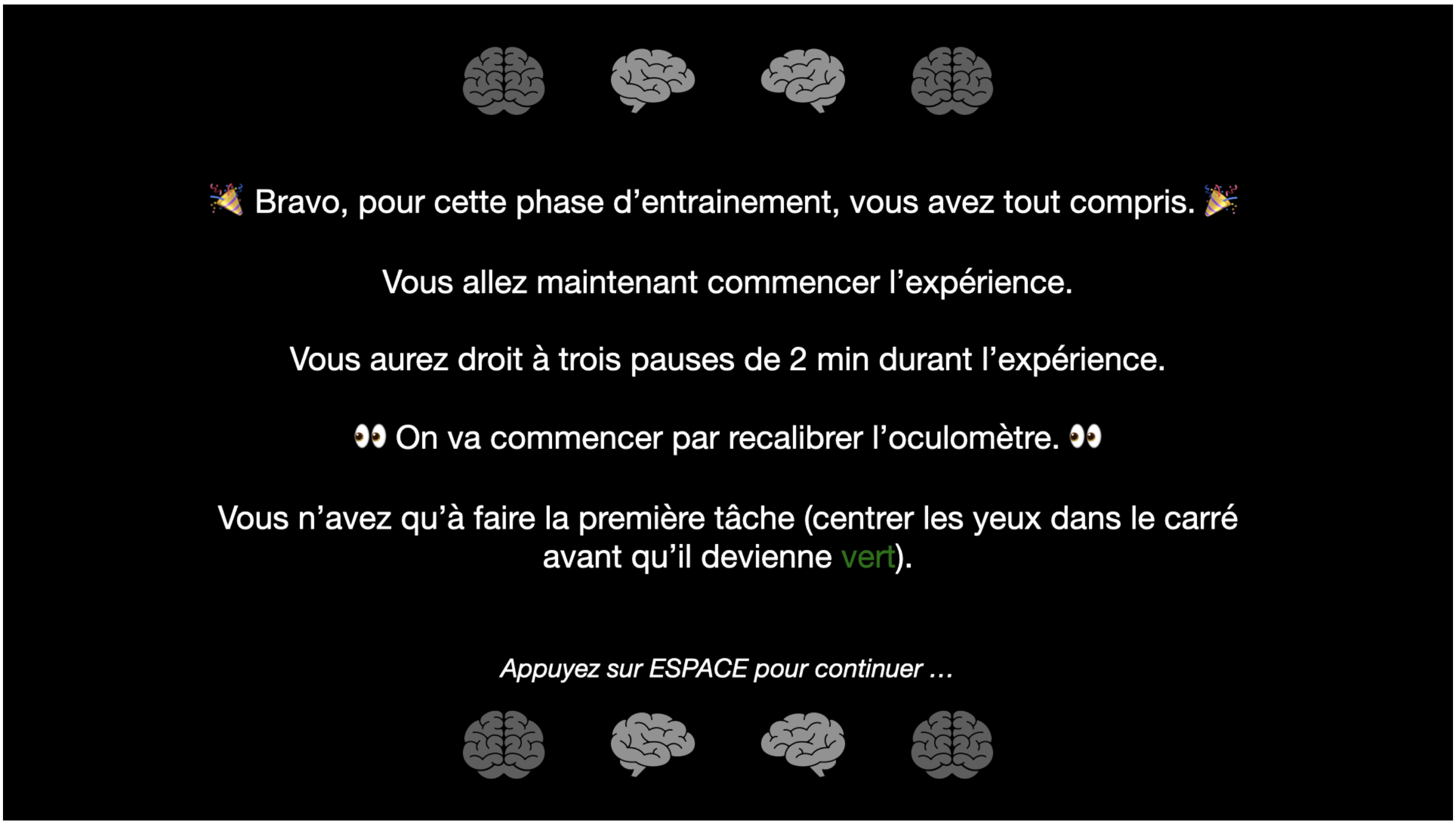


Figure D.11: Instruction n°7,c for the participant, *the first step of the calibration was necessary at the end of the training*

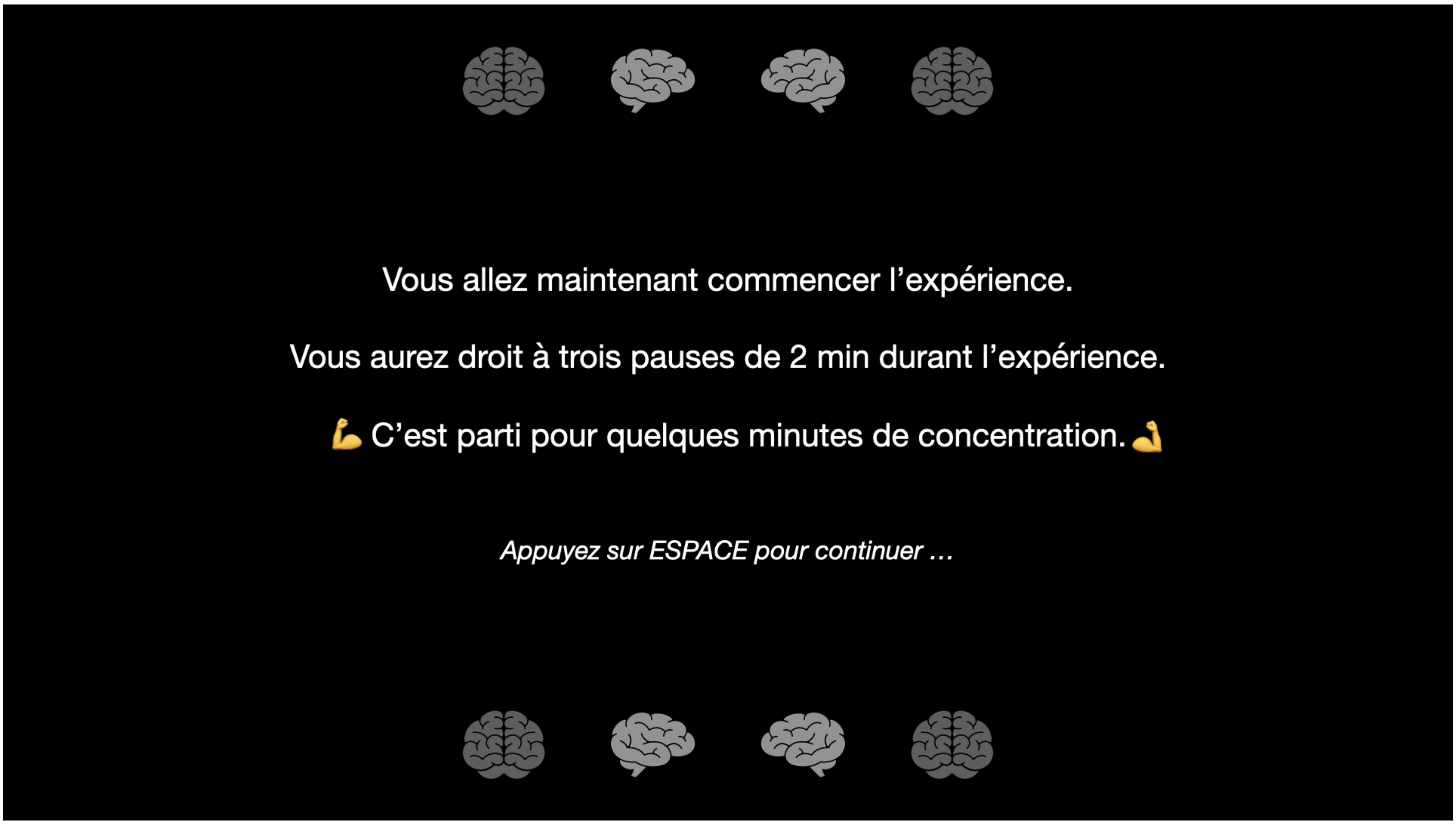


Figure D.12: Instruction n°8 for the participant

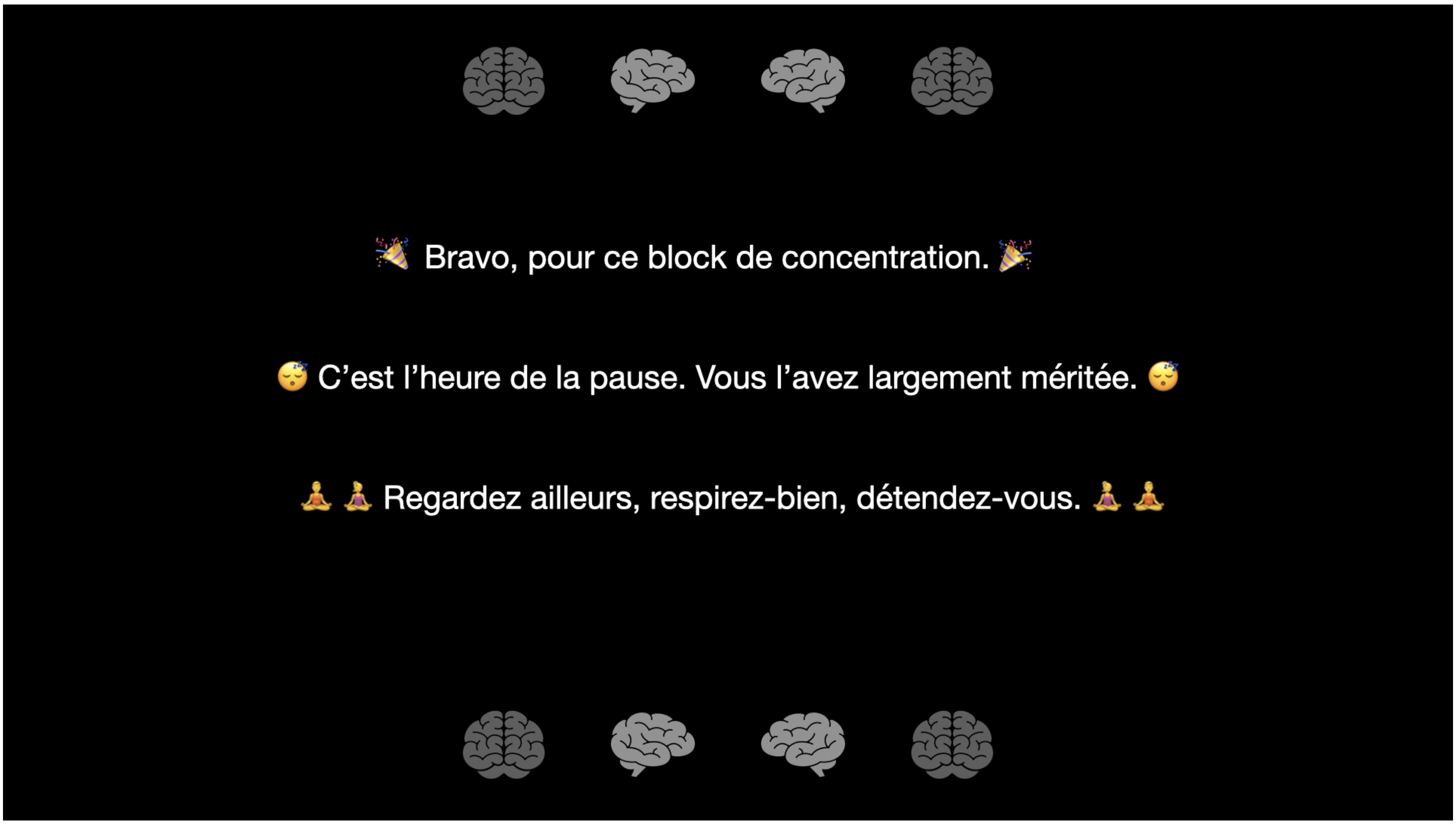


Figure D.13: Instruction n°9 for the participant, *during this slide, participants were blocked and had to wait 2 min*

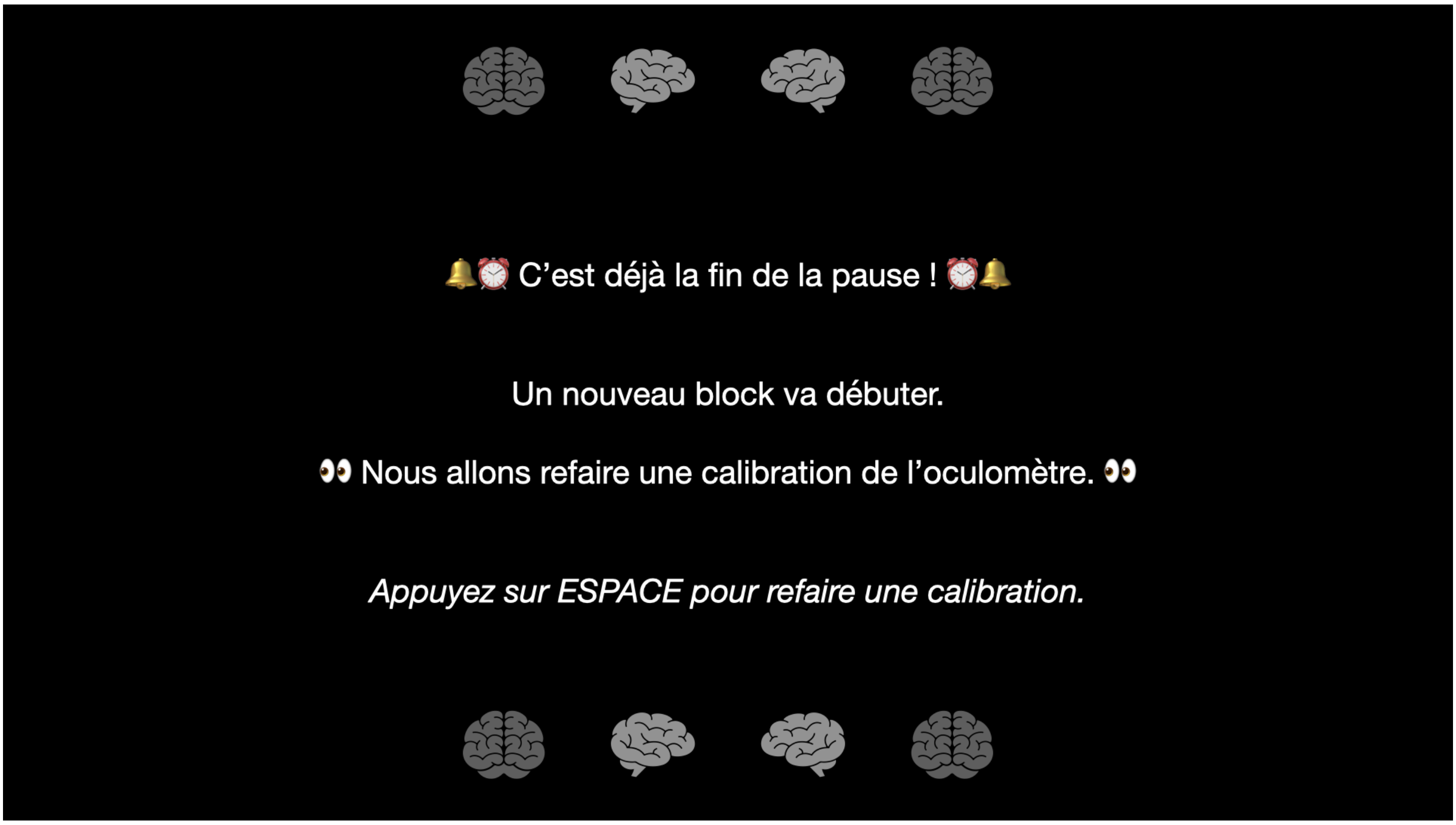


Figure D.14: Instruction n°12,b for the participant

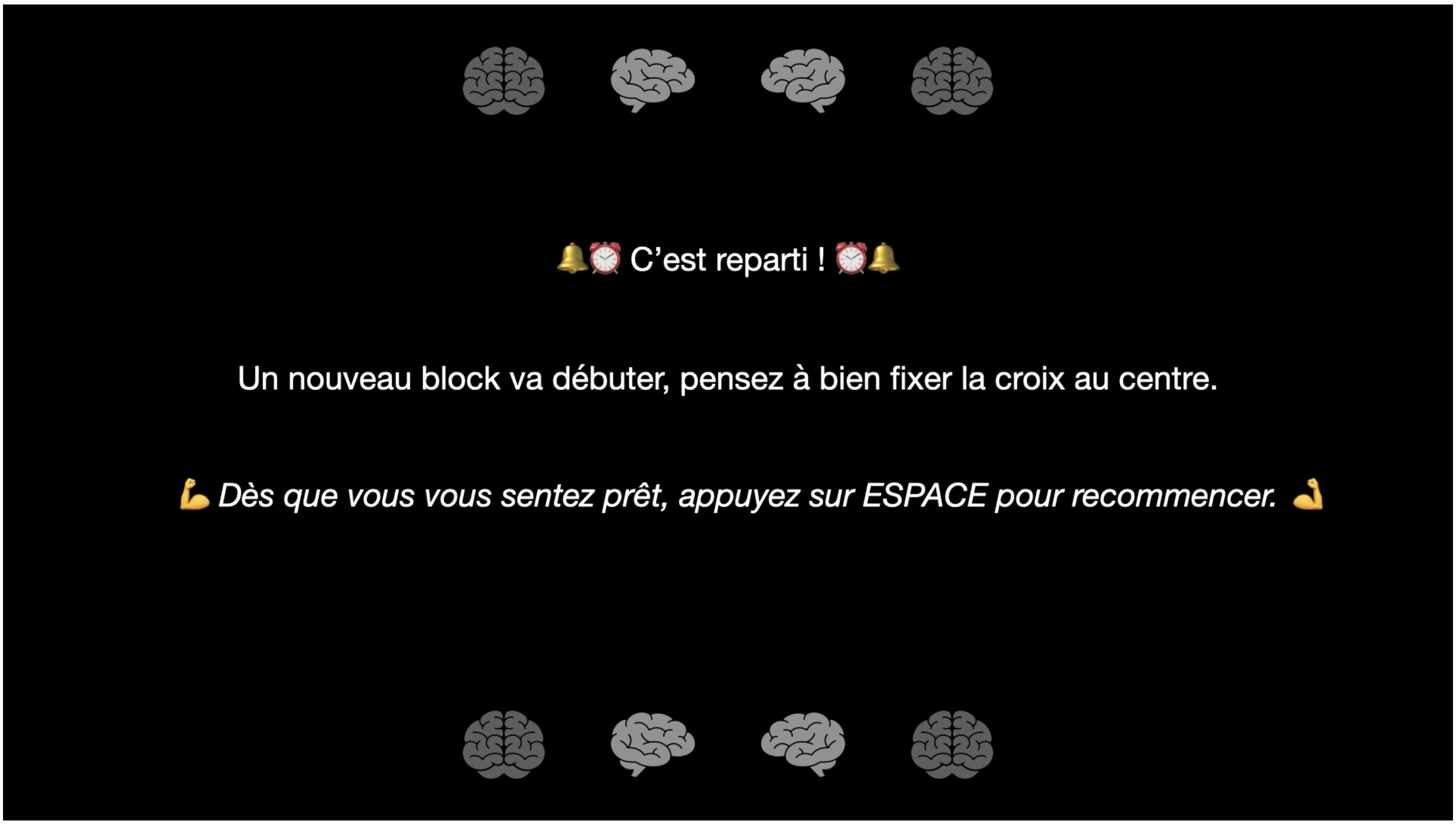


Figure D.15: Instruction n°10 for the participant

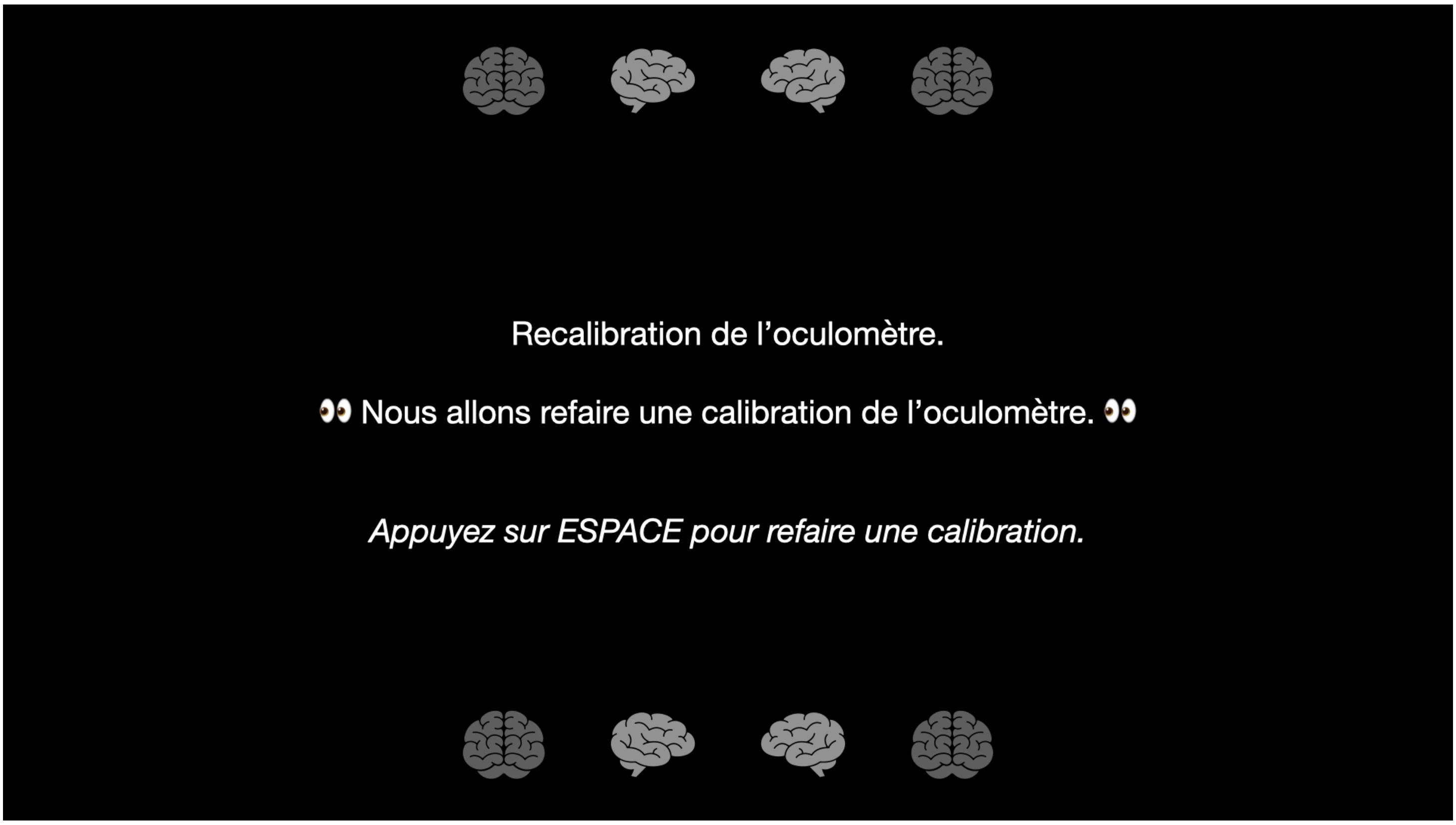


Figure D.16: Instruction n°11,a for the participant, *the participant didn’t knew it but this slide was announcing the reproduction task... Indeed, it is better to start with the calibration so they don’t have to be bother with it later*

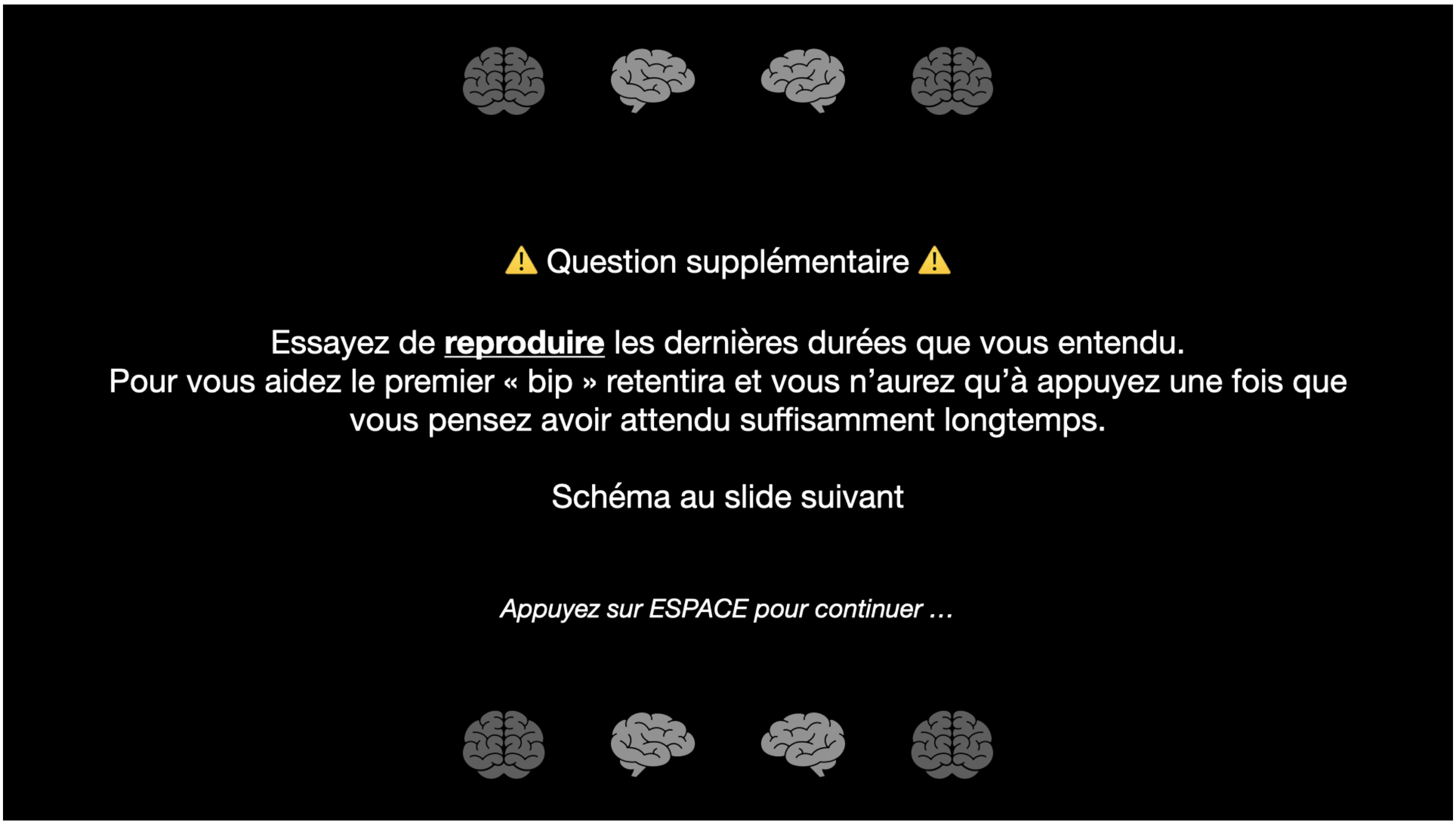


Figure D.17: Instruction n°11,b for the participant

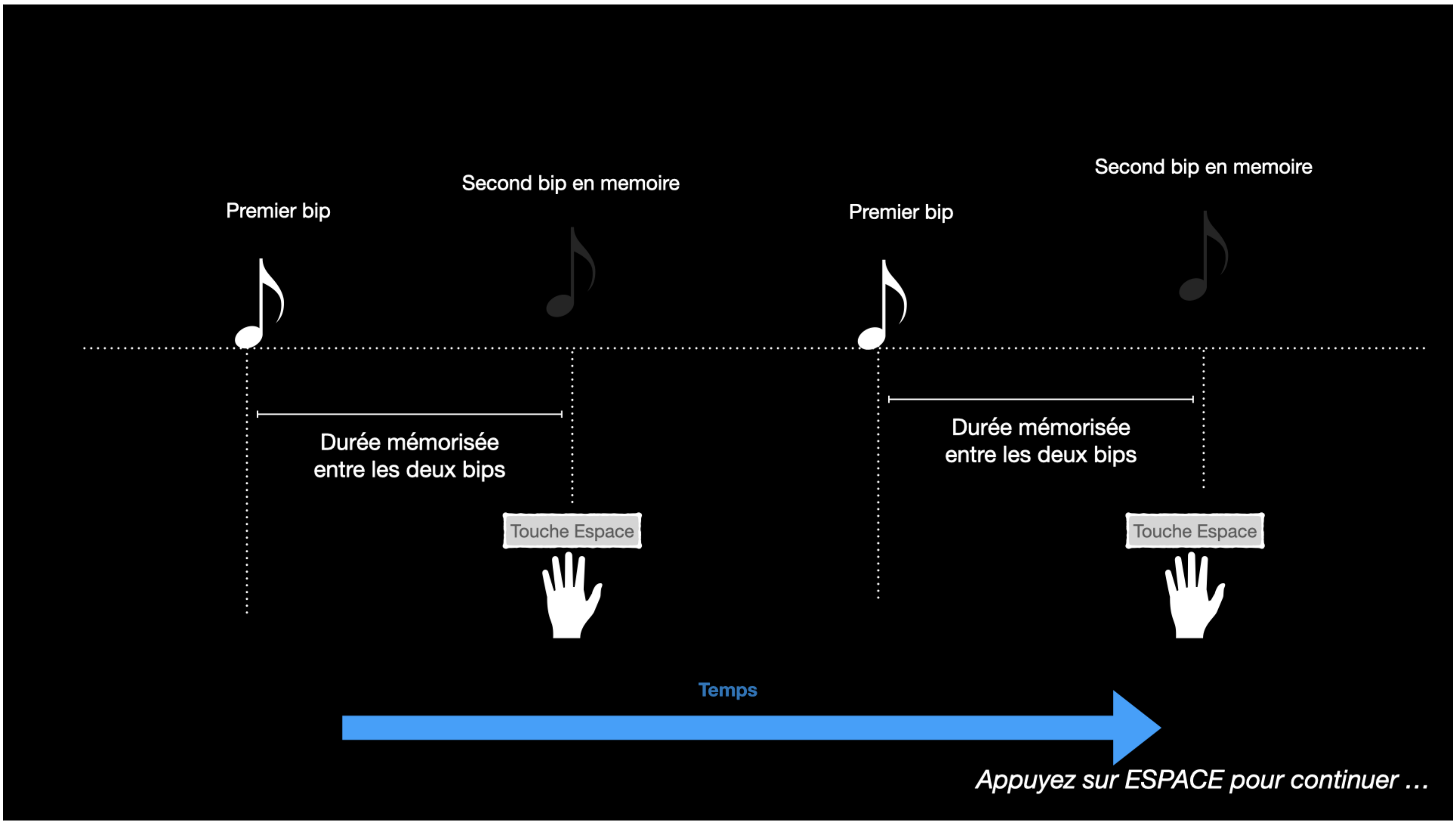


Figure D.18: Instruction n°12 for the participant

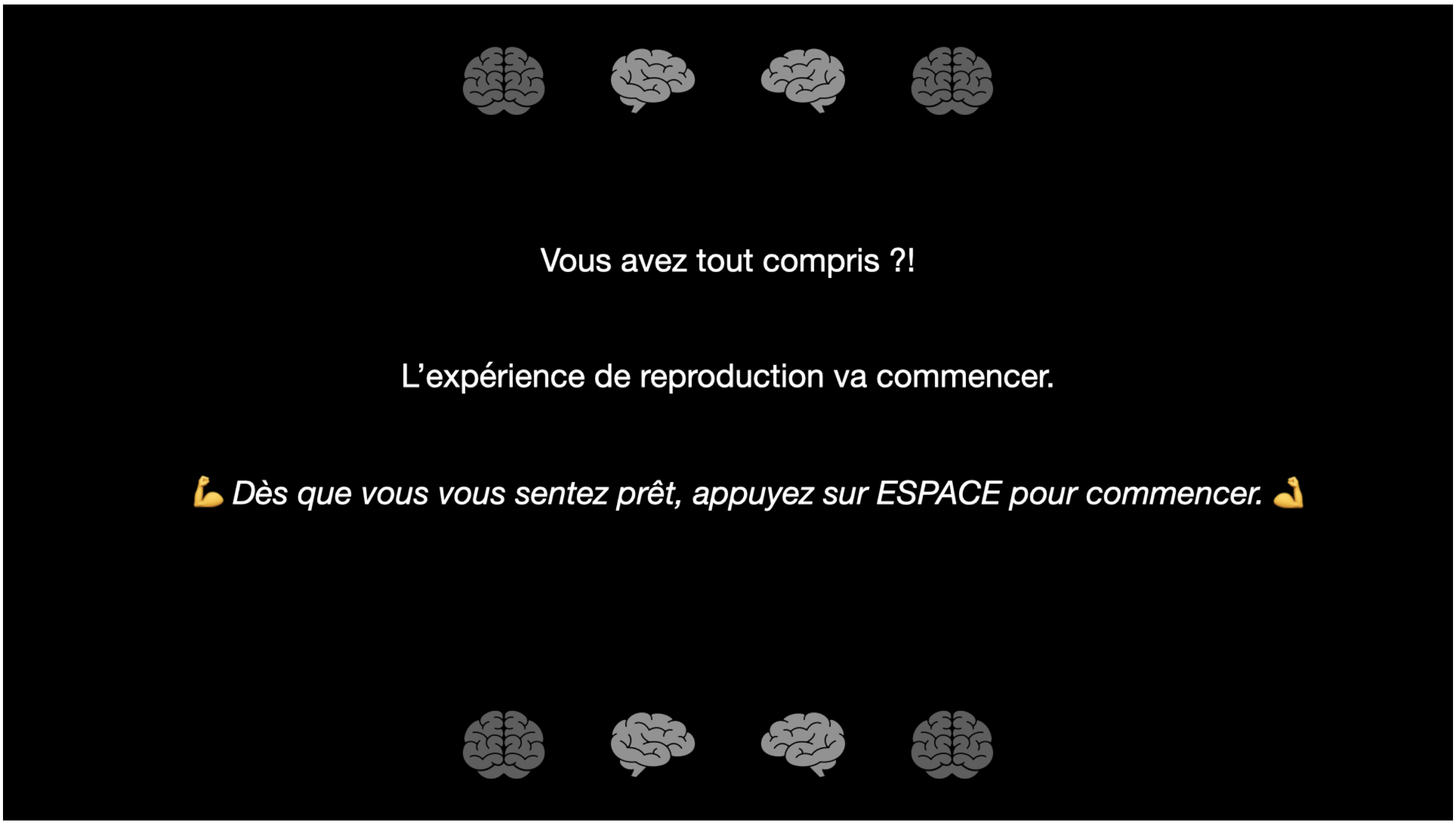


Figure D.19: Instruction n°13 for the participant

Figure D.20: Instruction n°14 for the participant

E

… as End

This is the END… at least for the report of my project.
However, Extra and the End both finish with an E.
Lets dive now in some extras.

E.1 : Internship CREST - Paris Brain Institute

Biological Determinants of Risk Aversion

During my gap year I also had the opportunity to work in a joint project between the CREST and the Paris Brain Institute (ex-Institut du Cerveau et de la Moelle Epinière). My supervisors at the CREST were Dr. Felix Tropf and Dr. Guillaume Hollard and my supervisor at the Paris Brain Institute was Baptiste Couvy-Duchesne, from the ARAMIS, th computational team of the Institute.

The main idea was to use the UK Biobank (UKB), a public data bank, in order to explore the morphometry of risky behavior and particularly in terms of neuroscience. A recent paper, Genetic underpinnings of risky behaviour related to altered neuroanatomy, from Aydogan et al., 2021, had already laid the foundations of this project.

Therefore, the aim of the internship was to try to replicate the founding of this Paper. However, the preprocessing of the data was much harder that expect and we were not able to finally find the same data sample as the authors. Indeed, the authors of the article went through an extremely complex preprocessing. Therefore, their results can be questionable since they were sub-sampling a lot the data.

Moreover, the project was even more difficult since it needed a collaboration between Institute, a collaboration that tried to be established during my internship.

E.2 : Master 2 University Paris-Saclay | Cenrale Supelec;
Computational Neuroscience and Neuroengineering

During this master I had the opportunity to discover the realm of neuroscience more deeply. Therefore I had classes in the physiological bases of neuroscience and also the methods for measuring and stimulating neuronal activity. Moreover I also had the opportunity to model with dynamical systems computational neuroscience models and to understand Neural bases of perception. Finally i had two project that I presented.

E.1.a : Closed-loop neuroscience.

During this class I discussed the recent development of Brain - Machine Interface (BMI) that try to create a prosthetic arm that decode motor signal from the brain (efferent signal). On the other way around, this prosthetic arm try to give sensory feedback to the user in order to optimize the movement of the arm.

E.1.b : Classification of sleep stages.

During sleep there are 5 sleep stage each characterized by different physiological signature. For instance, we tend to dream during the Rapid Eye Movement state (REM state). In order to track sleeping disorder, a startup called Dreem follow the activity of a huge number of users thanks to the helmet that they created which follow the electrical activity of the brain, heart beats, movement of muscles on the faces, etc... The goal was to create a classifier that could automatically separate the different stages. For this supervised learning task I tried various ethods. Yet the one that gave me the f1-score (which was the metric of interest for this hackaton) was simply a random forest classifier.

Mathias Vigouroux
Scolarité Extérieur :
Double diplôme ENSAE - ENS Ulm

Note de synthèse - Anglais

Bayesian Modeling of the Unfolding of Temporal Predictions.
Under the supervisor of Dr. Sophie K. Herbst, Neurospin, Unicog, CEA-INSERM Unit..

My internship was a research internship in a cognitive neuroscience laboratory, the Unicog laboratory, directed by Dr. Stanislas Dehaene, located on the CEA campus of Saclay. The team in which I was part of, the Cognition and Brain Dynamic team is directed by Dr. Virginie van Wassenhove, study our perception of time in through different angles. The research question that we developed with my supervisor, Dr. Sophie K. Herbst, was to developed a model that best mimic our learning of small temporal intervals.
The brain predicts the timing of forthcoming events in order to guide behavior, for instance when once tries to catch a ball in the air. Indeed, integrating a temporal structure in our internal representation of the environment, which is built from sensory inputs, in other words making temporal predictions, optimizes actions and benefits perception. This phenomenon is referred to as temporal preparation. Yet, it still remains unclear how temporal predictions are formed from the temporal statistics of sensory inputs. The Bayesian framework is widely used to model the way we build inferences from sensory inputs, and is gaining popularity in the timing literature. According to this framework, a Bayesian observer learns the statistics of the sensory environment by updating an internal probability distribution. While it is usually hypothesized that human participants similarly represent a probability distribution of the external inputs, it is unclear whether humans are learning all the statistical parameters relevant to build the distribution. In the simplest scenario in which temporal predictions can be derived from a Gaussian distribution, which is fully described by just two parameters : the mean and the standard deviation, it still remains unclear if humans learn both the mean and the standard deviation of a distribution. Indeed, the hypothesis until now was that because both the mean and the standard deviation had an impact on human behavior then these two parameters were both learned by the participant and also by the Bayesian Observer. Thus, the models used until now, particularly in timing, were always learning both the mean and the standard deviation. The validity of this hypothesis can be assessed in the framework of implicit temporal learning by comparing models that learn different types of parameters. To our knowledge, comparing Bayesian observers that learn the mean and the standard deviation to one that learns the mean only was not done in the literature. Yet, our study showed that a Bayesian Observer that only learns the Mean better explains participants' behavior compared to an Observer that learns both the mean and the standard deviation.Then, the effect of standard deviation can also be questioned in the realm of explicit timing and memory of duration. Asking participants to reproduce a series of stimuli, separated by durations drawn from a defined distribution, and comparing the standard deviation of the series they reproduced to the one they were shown, was not done in the literature to our knowledge. However, our study showed that the standard deviation of the produced foreperiods was not directly correlated to the standard deviation of the presented foreperiod. Rather, the standard deviation of the presented foreperiod had an impact on the mean of the produced foreperiods, and then the mean of the produced foreperiods had an impact on the standard deviation of the produced foreperiods.
Therefore, this study challenges the common hypothesis that humans learn perfectly distributions of temporal series of stimuli, even though all the parameters, i.e. both the mean and the standard deviation, have shown robust effect on the behavior.

Note de synthèse-français

Modélisation Bayesienne de l’Apprentissage des Prédictions Temporelles.

Sous la supervision du Dr. Sophie K. Herbst, Neurospin, Unicog, CEA-INSERM Unit..

Mon stage était un stage de recherche dans un laboratoire de neurosciences cognitives, le laboratoire Unicog, dirigé par le Dr Stanislas Dehaene, situé sur le campus de Saclay du CEA. L'équipe dont je faisais partie, l'équipe Cognition et Dynamique Cérébrale, dirigée par le DrVirginie van Wassenhove, étudie tout particulièrement notre perception du temps sous différents angles. Dans le cadre de ce stage, la question de recherche que nous avons développée avec ma superviseure, la Dr Sophie K. Herbst, était de mettre au point un modèle qui imite au mieux notre apprentissage de petits intervalles temporels.

En effet, afin de guider notre comportement, notre cerveau cherche à prédire de courtes durées, par exemple l’arrivée d’un ballon entre nos mains. En effet, intégrer une structure temporelle dans notre représentation interne de l’environnement, construite grâce à nos divers capteurs sensoriels, optimise nos actions et bénéficie à notre perception. Dans la littérature, ce phénomène est désigné sous le nom de préparation temporelle. Pourtant, la manière dont les prédictions temporelles sont formées à partir des statistiques temporelles de nos capteurs sensoriels n'est toujours pas claire. Le cadre Bayésien est largement utilisé pour modéliser la façon dont nous construisons des inférences à partir de nos sens, et gagne en popularité dans la recherche sur la perception temporelle. Selon ce cadre, un observateur Bayésien apprend les statistiques de l'environnement en mettant à jour une distribution de probabilité interne. Bien que l'on suppose généralement que les participants humains construisent de la même manière les probabilités des évènements externes, il n'est pas certain que les humains apprennent tous les paramètres statistiques pertinents pour construire cette distribution. En effet, même dans le scénario le plus simple dans lequel les prédictions temporelles peuvent être dérivées d'une distribution Gaussienne, entièrement décrite par seulement deux paramètres : la moyenne et l'écart-type, il n'est toujours pas clair si les humains arrivent réellement à apprendre ces deux paramètres. En effet, l'hypothèse jusqu'à présent était que, puisque la moyenne et l'écart-type avaient un impact sur le comportement humain, ces deux paramètres étaient à la fois appris par le participant et donc par l'observateur Bayésien qui les modélise. Ainsi, les modèles utilisés jusqu'à présent apprenaient toujours à la fois la moyenne et l'écart-type. La validité de cette hypothèse peut être évaluée dans le cadre de l'apprentissage temporel implicite en comparant des modèles qui apprennent différents types de paramètres. À notre connaissance, la comparaison entre des observateurs Bayésiens qui apprennent la moyenne et l'écart-type et un observateur qui n'apprend que la moyenne n'a pas été faite dans la littérature. Pourtant, notre étude a montré qu'un observateur Bayésien qui n'apprend que la moyenne explique mieux le comportement des participants qu'un observateur qui apprend à la fois la moyenne et l’écart-type. Ensuite, l'effet de l'écart-type peut aussi être questionné dans le domaine de l'apprentissage temporel explicite et de la mémoire de la durée. Demander aux participants de reproduire une série de stimuli, séparés par des durées tirées d'une distribution définie, et comparer l'écart-type de la série qu'ils ont reproduite à celle qui leur a été montrée, n'a pas été fait dans la littérature à notre connaissance. Cependant, notre étude a montré que l'écart-type de la série temporelle produites par les participants n'était pas directement corrélé à l'écart-type de la série temporelle présentée aux participants. Au contraire, l'écart-type de la série temporelle présentée impactait tout d’abord la moyenne de la série temporelle produite, puis la moyenne produite avait un impact sur l'écart-type de la série temporelle produite.

Par conséquent, notre étude remet en question l'hypothèse commune selon laquelle les humains apprennent parfaitement les distributions de séries temporelles de stimuli, même si tous les paramètres, c'est-à-dire à la fois la moyenne et l'écart-type, ont montré un effet robuste sur le comportement.