UNIVERSITY OF COPENHAGEN

DEPARTMENT OF NORDIC STUDIES AND LINGUISTICS

Bayesian Model of Multimodal Perception: Individual Differences in Perception of the Ventriloquist Illusion after-effect

Miłosz Holeksa (tmk107)
Weronika Kopytko (mgr983)

Computational Cognitive Science 2

MSc in IT & Cognition
Spring 2024

Division of work

The authors attest that they contributed equally to preparing the project, accordingly to information presented in the table below.

Section	**Author**
Abstract	Miłosz
1.1 Predictive Processing (PP) theory	Weronika
1.2 Bayesian Modeling	Weronika
2. Literature review	Weronika
3.1 Data	Miłosz
3.2 PyMC library for Bayesian inference	Weronika
3.3 Baseline Model	Miłosz
3.4 Multimodal Model	Miłosz
4. Results	Miłosz & Weronika
5. Discussion	Miłosz & Weronika

Abstract

The Predictive Processing (PP) theory posits that the human brain continually generates predictions about the environment, informed by prior beliefs, to guide behavior and perception. Bayesian modeling serves as a mathematical framework to show how these predictions are updated based on sensory input. This study investigates Bayesian modeling in the context of multimodal perception, with a focus on the integration of audio-visual stimuli and phenomenon known as the Ventriloquist Illusion. Utilizing publicly available data from a previous study (Kayser et al, 2020), we develop a Bayesian model to examine individual differences in perception and susceptibility to the Ventriloquist Illusion. Our analysis centers on the aftereffect in auditory trials following exposure to audio-visual stimuli, modeling it as a function of the perceptual shift towards visual stimuli. Through this investigation, we aim to contribute to a deeper understanding of the intersection between Bayesian principles and PP theory in shaping human perception.

1 Introduction

1.1 Predictive Processing (PP) theory

Predictive Processing (PP) or *predictive coding* theory of mind is one of the most debated theories in cognitive science (Clark, 2013; Hohwy, 2020). It posits that a given system (such as the human brain) is constantly generating predictions about its environment, based on an internal model of the world. Any discrepancy between the prediction and sensory input gives rise to *prediction error*, that can be used to update the models' predictions. While facing new sensory input which yields big prediction error, a person can either update their beliefs or take action to change the world in order for it to match the mental model. The main task of the brain is to minimize the prediction error, such that the model's performance would be improved. The predictions guide our behavior, interactions with the world and shape our subjective experiences, as they are based on our set of beliefs about the world. Beliefs are represented as prior knowledge. The view of the brain as a predictive machine follows closely the concept of Bayesian Inference, which is a statistical method which leverages *Bayes' Theorem* [1]for updating the probability of a given hypothesis. By calculating Bayesian probabilities, we can model the brain in accordance to PP theory.

1.2 Bayes' Theorem

Bayes' Theorem is a cornerstone of probability theory in statistical inference, as it forms the foundation of Bayesian modeling. It provides a mathematical framework, which allows to calculate the probability of any hypothesis (H) given the data (D):

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)} \tag{1}$$

P(D|H) in Bayes' theorem [1] is a *likelihood*, P(H) is a prior probability of the hypothesis and P(D) is a marginal probability of data. If we know the values of those probabilities, we can calculate P(H|D) - the posterior probability of the hypothesis given our data. By expressing the relationship between defined probabilities, Bayes' Theorem allows for incorporation of prior knowledge with observed data, to produce updated, more accurate probabilities.

Since we can use Bayes' theorem to represent a posterior probability of any given hypothesis, it makes a powerful tool in experimental research, providing a probabilistic mathematical models for testing various hypotheses.

1.3 Bayesian modeling for multimodal perception

The Bayesian approach is particularly valuable in cognitive science, where many phenomena are inherently uncertain and influenced by numerous factors. Researchers can create models allowing for nuanced understanding of how humans process information, make decisions and reason about the world around them. In perception, Bayesian inference can explain how the brain combines noisy sensory inputs to form a coherent representations of the world. We can model how the brain combines priors corresponding to different modalities to form a unified, multimodal percept. One compelling example of multimodal perception is the integration of audio-visual data, well illustrated by commonly known Ventriloquist Illusion. This phenomenon occurs when the perceived location of the audio stimuli is biased towards the location of the visual stimuli. When the ventriloquist speaks, the audience perceives the sound as coming from the dummy's mouth rather than the original source (the ventriloquist). This effect can be explained through Bayesian modeling, where the brain uses visual cues, which are generally more reliable for spatial localization, to update auditory spatial estimates.

2 Literature review

The PP theory of mind has been explored in a large body of work. It originated with Helmholz (1860), as a theoretical view of 'perception-as-inference', but it has since come a long way, evolving into the computationally-based theory. Due to compelling research, perception seems to conform to Bayesian ways of combining sensory evidence with prior knowledge (Clark, 2015). Bayesian models are used to explore human perception in a wide range of its aspects. It has been shown, that Bayesian approach to human brain function can model sensory systems: auditory cortex (Kumaar et al, 2011; Baldeweg, 2006), visual cortex (Homann et al, 2017; Schellekens et al, 2016; Rao & Ballard, 1999) or olfactory system (Zelano et al, 2011) among others.

One of the main arguments against the PP theory is the alleged lack of neurophysiological evidence (i.e. how the brain is executing the theory

assumptions on a general neuronal level). However, increased number of both technological and theoretical advances, has caused this empirical gap to be filled in recent years, showing that the theory is neurally plausible which enhances its further development (Walsh et al, 2020; Barron et al, 2020).

A body of recent work has also focused on the relationship between PP and how human process data from different modalities (Sidhu & Vigliocco; Hohwy, 2020). In particular, the central role of integrating multisensory data has been explored. This is due to the role of *crossmodal correspondences* - the mutual dependence of processes underlying perception of data from different modalities. As an example, it has been shown that humans consider audio and visual data as coming from the same source even if their actual emission locations differ, if they believe that there is a causal link between the data (the effect commonly known through example of Ventriloquist Illusion) (Stawicki et al, 2019). This belief can be represented as (multimodal) prior knowledge about the co-occurrence of stimuli of different modalities in perceived scene.

A previous study by Kayser et al (2022), explored the Ventriloquist Illusion, concluding that audio-visual stimuli influences how we perceive audio stimuli presented afterwards (so-called *after-effect*). In this research, we analyze whether Bayesian model of multimodal perception can accurately model individual differences in people's beliefs (priors) and their receptivity to Ventriloquist Illusion and the after-effect. This introduces new insights to how participants perceived the audio-visual stimuli based on their individual differences. We formulate three hypotheses: (H1) Initial individual receptivity to after-effect varies across participants, (H2) the rate of change of the receptivity varies across participants and (H3) a Multimodal Model, which includes individual specific parameters, is more predictive than the Baseline Model, which only accounts for general effect across all participants.

3 Methodology

3.1 Data

In our analysis we reuse readily available data, collected in a study conducted by Kayeser et al (2023). The data is publicly available at github_link and corresponds to the eleventh experiment listed. The dataset contains stimuli and response information for 19 participants. The experiment measures the ventriloquist illusion and the aftereffect. In particular, the authors aim to

show, how a number of previous trials (one or two) influences the magnitude of these effects. The experiment consists of sequences of trials of different modalities. The AV trials measure the size of the Ventriloquist Illusion. In these trials participants were presented with both a visual and an auditory stimulus. The participants' task was to locate the source of the sound. The shift of the perceived location of the sound towards the location of the visual stimulus represents the magnitude of the ventriloquist illusion. Each of the AV (audio-visual) trials was followed either by an A (audio) or a V (visual) trial. The A trials were used by the authors to measure the size of the ventriloquist *after-effect* - the shift of perceived sound location primed by the Ventriloquist Illusion in the previous AV trials. The V trials consisted of visual stimuli only. The purpose of these stimuli was to balance the load between modalities and they were discarded both in the original study, as well as in our analysis.

We prepare the data by loading it in. Considering the specification of our analysis, we restructure the data so that it only consists of pairs of trials: AV - A. For our dataset, we drop any variables that are not necessary for the analysis (for more information about the original data check the github repository).

3.2 PyMC library for Bayesian inference

We use Python's PyMC 5.10.2 library for our analysis, which is designed for Bayesian statistical modeling and probabilistic machine learning, allowing for designing complex probabilistic models that use sampling methods based on *Markov Chain Monte Carlo* (MCMC) algorithm. PyMC leverages automatic differentiation to effectively compute gradients, which is an essential aspect for MCMC methods. To define a Bayesian model, we specify the likelihood function for the observed data (*predicted variable*) and prior distributions for model parameters (*predictors*). Distributions for priors encapsulate our initial beliefs about the parameters, before observing the data. The likelihood function describes how the observed data is generated given the parameters. PyMC then uses this information to construct a joint probability distribution over all variables defined in the model. It leverages the MCMC sampling methods to approximate the posterior probability distributions of the parameters. We choose to use *No-U-Turn Sampler* (NUTS), an advanced algorithm that adapts the step size and direction of its step to efficiently explore the parameters space. NUTS is a variant of the *Hamiltonian*

Monte Carlo (HMC) method, which uses gradients of the log-posterior to guide the sampling process. The algorithm's strategy is to avoid inefficient backtracking (U-turns) during the sampling process, which would cause it to explore some parts of the distribution more than the others. HMC is built on principles from physics, simulating the movement of a particle in a potential energy field, which leverages gradients to inform the direction and distance of each step. This results in faster convergence and better approximation of the distribution then other, more traditional MCMC methods (such as Metropolis-Hastings algorithm).

We fit two Bayesian models in our analysis: Baseline Model and Multimodal Model, which let us analyze the Ventriloquist after-effect in terms of individual differences in participants' priors.

3.3 Baseline Model

The basic task for our model is to predict the size of the after-effect in A trials, which directly follow AV trials. We therefore start with preparing a baseline model. We define our model as following:

$$
\begin{aligned}
&Shift\_A \sim Normal(\mu, \sigma),\\
&\mu = \alpha + \beta * Shift\_AV,\\
&\alpha \sim Norma(0, 1),\\
&\beta \sim Normal(0, 1),\\
&\sigma \sim Exponential, (1)
\end{aligned}
$$

where the Shift_AV is the shift of the perceived sound location towards the visual stimuli (which can also take negative values if the shift was away from the visual stimulus and is measured in degrees). Conversely Shift_A represents the shift of the perceived location of the sound location in the A trials. It has a positive value if the direction of the shift was in line with the difference of stimuli location in the preceding AV trial and a negative value if the directions are opposite. Shift_A is normally distributed with mean μ and standard deviation σ. We assume a linear relationship between these variables. The mean μ of Shift_A depends on an intercept α and a slope β multiplied by the predictor variable $\text{Shift}_A V$. The intercept α and slope β follow a normal distribution with mean 0 and standard deviation 1. The

standard deviation σ follows an exponential distribution with rate parameter 1, ensuring it is positive.

3.4 Multimodal Model

In Multimodal Model we introduce a *multimodal prior*, which models each participant's receptivity to the after-effect primed by the illusion. We initialize it as a normal distribution, which makes no assumptions about the individual beliefs other than the mean and standard deviation. We then check how the distribution changes as we update the model in each trial for each participant, representing individual beliefs. The structure of the model builds upon the baseline model. However here we introduce individual-specific parameters, allowing for variation across different individuals. We therefore effectively build a separate regression model for each participant. We also introduce a variable *obs*, which represents the participant-specific index of stimuli (the trial number for each participant). The individual β_{oi} slope allows for modeling how the trial number influences the predicted Shift_A variable. Model is defined as following:

$$
\begin{aligned}
&Shift\_A \sim Normal(\mu, \sigma),\\
&\mu = \alpha_i + \beta_{o_i} * obs + \beta_{s_i} * Shift\_AV,\\
&\alpha_i \sim Norma(0, 1),\\
&\beta_{o_i} \sim Normal(0, 1),\\
&\beta_{si} \sim Normal(0, 1),\\
&\sigma \sim Exponential(1),
\end{aligned}
$$

where the mean μ of Shift$_A$ depends on individual-specific intercept α_i, and individual-specific slopes β_{oi} and β_{si} for the predictor variables obs and Shift$_A V$, respectively. The priors for α and β_{s_i} are analogous to those in the baseline model, however they are stratified by participants. The individual-specific slope β_{oi} for obs follows a normal distribution with mean 0 and standard deviation 1.

Thanks to this model design we can interpret the α values as the participants initial receptivity to after-effect. This allows us to verify Hypothesis 1 by looking if this value varies significantly between participants. Similarly

we can interpret the β_{oi} values as the rate of change of the receptivity to after-effect. This allows us to verify Hypothesis 2 by looking if this value varies significantly between participants.

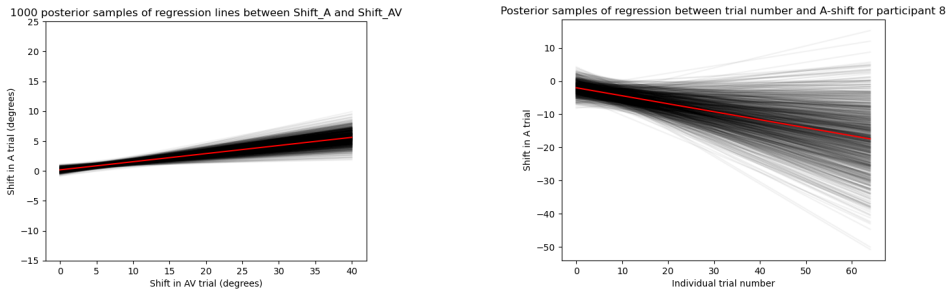
We performed inference for both our models using 4000 tune and 4000 draw samples, using four chains, three of which were used for diagnostic purposes.

As a final step of our analysis we compared these models using *Watanabe-Akaike Information Criterion* (WAIC). It is a widely used criterion for model comparison that estimates the out-of-sample prediction error of a Bayesian model. It balances model fit and complexity by penalizing overfitting, making it suitable for comparing models of different complexities based on their predictive accuracy.

4 Results

We tested three hypotheses with the Baseline Model [3.3] and the Multimodal Model [3.4]: (H1) Initial individual receptivity to after-effect varies across participants, (H2) the rate of change of the receptivity varies across participants and (H3) a Multimodal Model, which includes individual specific parameters, is more predictive than the Baseline Model, which only accounts for general effect across all participants.

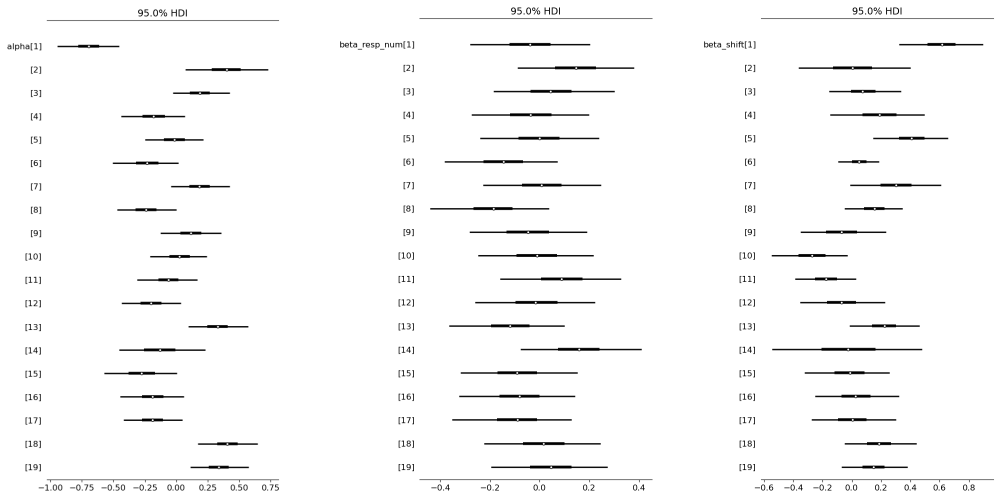
After fitting the baseline model to the data we received posterior distributions for α and β values. The mean value for α was 0.174 with a 95% HDI interval of (-0.395, 0.805). The mean value for β was 0.135 with a 95% HDI interval of (0.068, 0.205). In order to better represent the posterior distribution we plotted one thousand regression lines sampled from the posterior (see Figure 1a). The red line represents the mean and each gray line is a separate sampled regression.



(a) Baseline Model (b) Multimodal Model (ID = 8)

Figure 1: α and β_{oi} values for the Baseline Model (a) and Multimodal Model for the 8th participant (b).

In the second step we performed inference on the Multimodal Model. In this case we received separate α, β_{si} and β_{oi} values for each participant. Due to the size of these results, we plot them together and the exact values are available in Appendix 1. The alpha and both beta values we obtained are shown in Figure 2a, Figure 2b and Figure 2c (please note that the data on the scale are standardized, so they do not reflect the original scale, but still show the magnitude of the effect. In Appendix 1 you can find plots for all participants with the original data size).



(a) α values (b) β_{oi} values (c) β_{si} values

Figure 2: α (a), β_{oi} (b) and β_{si} values for the Multimodal Model. The thinner lines represent the spread of the data for each participant. The thicker lines represent 95% HDI intervals.

Following a typical approach in Bayesian statistics we verify our hypothesis by comparing the posterior distributions rather than performing test, as is often the case in frequentist statistics. More precisely we compare the 95% HDI intervals to see if values vary significantly.

The results show that there were significant differences between participants in terms of the α value, representing Initial individual receptivity to after-effect. Similarly we can also observe non overlapping HDI intervals for the β_{oi} value, which represents the rate of change of the receptivity of the aftereffect.

As a final step of our analysis we compared these models using WAIC analysis, obtaining the results shown in Table 1.

model	**rank**	**elpd_loo**	**p_loo**	**elpd_diff**	**weight**
multimodal	0	−1712.88	59.90	0.0	0.59
baseline	1	−1721.80	3.85	8.92	0.40

Table 1: Comparison of Multimodal and Baseline models using WAIC

These results show that the Multimodal Model has a lower rank, which

means it has a lower (better) WAIC value. We also get a slightly higher elpd_loo (expected log pointwise predictive density), which represents the model's predictive power as higher for the Multimodal Model. The p_loo, representing model complexity, is unsurprisingly much higher for the Multimodal Model. The weight value is higher for the Multimodal model, showing that it has a higher chance of being the best model given the data. In general these results indicate that the Multimodal Model has a better predictive power, even when accounting for the difference in complexity.

5 Discussion & Conclusion

The results above allow us to confirm all three of our hypotheses. Participants differ in initial individual receptivity as well as the rate of change of the receptivity in time (H1 and H2). Additionally, the Multimodal Model, which includes individual specific parameters, is more predictive than the Baseline Model (H3). Bayesian modeling seems to be a useful tool in representing individual differences in perception of multimodal stimuli. In particular, these results show how cross-modal correspondences (the mutual relationships between different modalities) can be modelled in a Bayesian framework. In a broader perspective they show how human perception can be successfully modelled using relatively few parameters.

However, these results also have their limitations. We only introduced linear terms for the regressions and these relationships could be more complex and of non-linear nature. On top of that, our parameters α and β_{oi}, which represent participant's receptivity to the illusion, could in fact be biased by some additional factors. That is, there could be some other variables influencing the shapes of the regression, which we did not account for. It would be interesting to extend the Multimodal Model by adding more parameters and seeing how that affects the model's performance.

An interesting aspect of the study is the model's ability to assign different receptivity values to different participants. This aligns with the PP theory of mind, showing individual differences in predictive mental models realized by human brain and leaves space for further exploration - particularly about the underlying causal sources of such differences between participants. Our study also leaves other interesting ideas for further development. In particular, it could be interesting to explore how the Bayesian framework could be applied to modelling the relations between other modalities. Perhaps an even further

integration could be possible, by preparing larger and more complex models, capable of integrating more than two modalities at a time. As a potential application, these could for example be implemented in human-like robots, whose task would be to mimic human perception.

Bibliography

1. Baldeweg, T. (2006). Repetition effects to sounds: Evidence for predictive coding in the auditory system. Trends in Cognitive Sciences, 10(3), 93–94. https://doi.org/10.1016/j.tics.2006.01.010

2. Barron, H. C., Auksztulewicz, R., & Friston, K. (2020). Prediction and memory: A predictive coding account. Progress in Neurobiology, 192, 101821. https://doi.org/10.1016/j.pneurobio.2020.101821

3. Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical Microcircuits for Predictive Coding. Neuron, 76(4), 695–711. https://doi.org/10.1016/j.neuron.2012.10.038

4. Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences, 36(3), 181–204. https://doi.org/10.1017/S0140525X12000477

5. Hohwy, J. (2020). New directions in predictive processing. Mind & Language, 35(2), 209–223. https://doi.org/10.1111/mila.12281

6. Homann, J., Koay, S. A., Glidden, A. M., Tank, D. W., & Berry, M. J. (2017). Predictive Coding of Novel versus Familiar Stimuli in the Primary Visual Cortex (p. 197608). bioRxiv. https://doi.org/10.1101/197608

7. Kayser, C., Park, H., & Heuer, H. (2022). Cumulative multisensory discrepancies shape the ventriloquism aftereffect but not the ventriloquism bias. bioRxiv (Cold Spring Harbor Laboratory). https://doi.org/10.1101/2022.09.06.506717

8. Kumar, S., Sedley, W., Nourski, K. V., Kawasaki, H., Oya, H., Patterson, R. D., Howard, M. A., III, Friston, K. J., & Griffiths, T. D. (2011). Pre-

dictive Coding and Pitch Processing in the Auditory Cortex. Journal of Cognitive Neuroscience, 23(10), 3084–3094. https://doi.org/10.1162/jocn_a_00021

9. Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. Nature Neuroscience, 2(1), 79–87. https://doi.org/10.1038/4580

10. Schellekens, W., van Wezel, R. J. A., Petridou, N., Ramsey, N. F., & Raemaekers, M. (2016). Predictive coding for motion stimuli in human early visual cortex. Brain Structure and Function, 221(2), 879–890. https://doi.org/10.1007/s00429-014-0942-2

11. Sidhu, D.M., Vigliocco, G. (2023). I don't see what you're saying: The maluma/takete effect does not depend on the visual appearance of phonemes as they are articulated. Psychon Bull Rev 30, 1521–1529 . https://doi.org/10.3758/s13423-022-02224-8

12. Spence, C. (2022). Exploring Group Differences in the Crossmodal Correspondences. Multisensory Research, 35(6), 495-536. https://doi.org/10.1163/22134808-bja10079

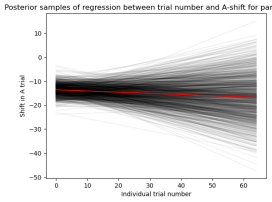
13. Stawicki, M., Majdak, P., & Başkent, D. (2019). Ventriloquist Illusion Produced With Virtual Acoustic Spatial Cues and Asynchronous Audiovisual Stimuli in Both Young and Older Individuals. Multisensory Research, 32(8), 745-770. https://doi.org/10.1163/22134808-20191430

14. Walsh, K. S., McGovern, D. P., Clark, A., & O'Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. Annals of the New York Academy of Sciences, 1464(1), 242–268. https://doi.org/10.1111/nyas.14321

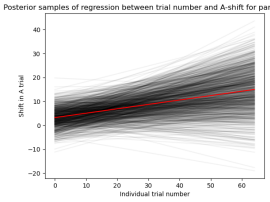
15. Zelano, C., Mohanty, A., & Gottfried, J. A. (2011). Olfactory Predictive Codes and Stimulus Templates in Piriform Cortex. Neuron, 72(1), 178–187. https://doi.org/10.1016/j.neuron.2011.08.010

Appendices

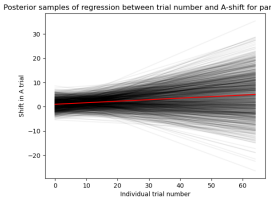
Apendix 1



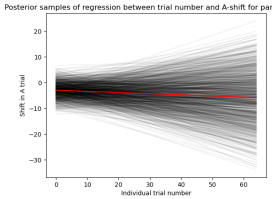
(1)



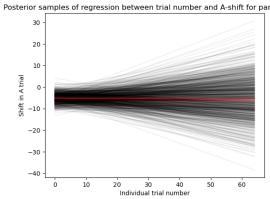
(2)



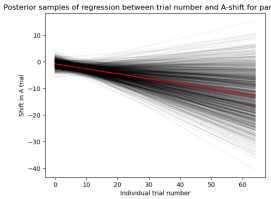
(3)



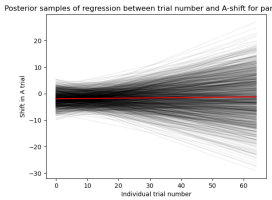
(4)



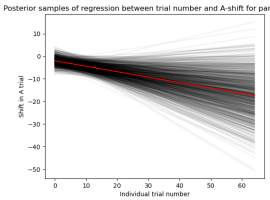
(5)



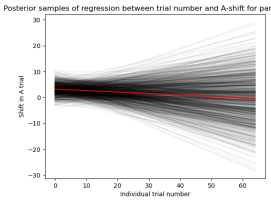
(6)



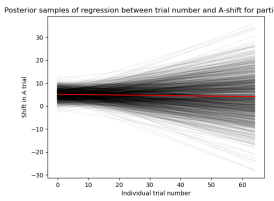
(7)



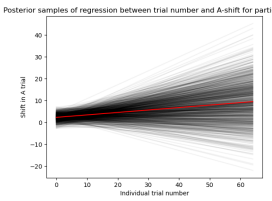
(8)



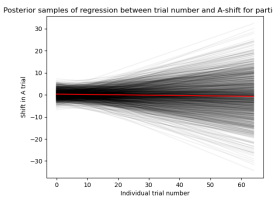
(9)



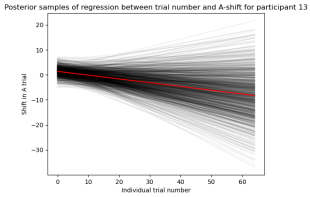
(10)



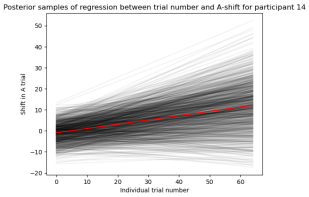
(11)



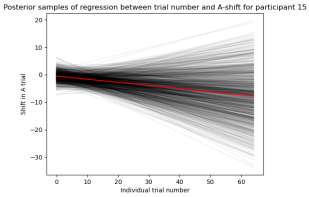
(12)



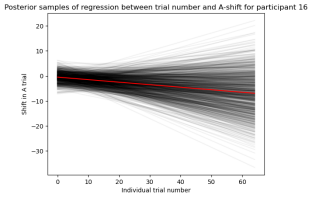
(13)



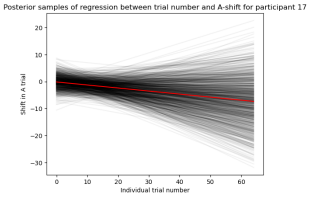
(14)



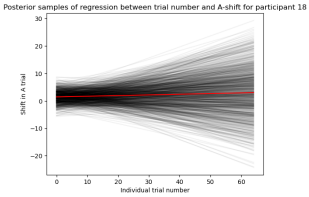
(15)



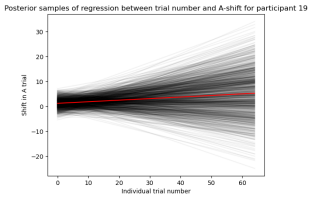
(16)



(17)



(18)



(19)

Figure 3: Regression lines in Multimodal Model for for all 19 participants. The thinner lines represent the spread of the data for each participant. The thicker lines represent 95% HDI intervals.