

https://github.com/PeterThramkrongart/CompMod_1/blob/master/CompMod1.md?fbclid=IwAR39SmGg9GpPhgo3d_qQBVFJSoY8DKLs4Hy2EVpm9UDxbsmPxQsqt_6TxDY

Assignment 1

Foraging experiment

Introduction

The conceptual research question, which this study aims to investigate, is how scanpaths during visual search and simple counting tasks differentiate due to both top-down and bottom-up factors. Specifically, we will focus on the differences in the saccade lengths and their distributions in the two conditions - count and search task.

The impact of this question is more conceptual. We are interested in uncovering how task goals can influence an otherwise quite bottom-up driven process.

Hypothesis

We hypothesize that the search or *foraging* condition will differ systematically from the count condition. In the foraging condition, we expect to observe a heavy-tailed distribution of saccade lengths and that the saccade amplitudes will be longer, compared to the count condition. This would imply that bottom-up features of the stimulus are not the only factor influencing the eye gaze movements, but that the top-down factors such as the type of task can impact the pattern of eye movements as well.

Methods

Participants

There were 6 participants in total in the foraging experiment, all of them 4th-semester Cognitive Science students at Aarhus University. One goal during the data collection was to introduce confounding factors that would intervene with the eye tracker's ability to detect the participants' pupils. Hence, most participants were wearing contact lenses or glasses and/or a lot of mascara on their eyelashes. This handicap was added so that we can observe what 'real', messy data looks like and have the opportunity to learn not only how, but also why, it is necessary to control for these confounds in future experiments. Furthermore, during one of the trials the research assistant accidentally exited the PsychoPy script, resulting in loss of data.

Task

The experiment was comprised of two tasks:

(1) Search condition: in this condition, participants viewed images of natural scenes for a set amount of time, with the goal of finding a transparent star, hidden in the picture. If the participant located the star, they were instructed to fixate on it.

(2) Count condition: in this condition, participants viewed images of natural scenes for a set amount of time, in order to count elements depicted in the image (e.g. counting the number of birds or sheep). The element which the participant was supposed to count was not explicitly stated, but rather inferred from features of the image itself.

Eye-Tracker Setup

The participants' gazes were measured with the EyeLink1000 at COBE Lab. The machine has a monocular measure, which means that we need to measure the participant's dominant eye, which can be tested by the Miles test. After setting the sample rate at 500 Hz and illuminator power at 75%, calibration of the eye ensued, in order to make sure that the eye tracker could detect the pupil from different angles as accurately as possible.

Preprocessing Eye-Tracking

Using the EyeLink1000 software, the visual eye-tracking data was translated and grouped into saccades and fixations of differing lengths. Consequently, the scan paths were determined (*Figure 1*). This data was then exported from the eye-tracker. Afterwards the data was cleaned by deleting a few nonsensical outliers, which were deemed artifacts from the eye-tracker.

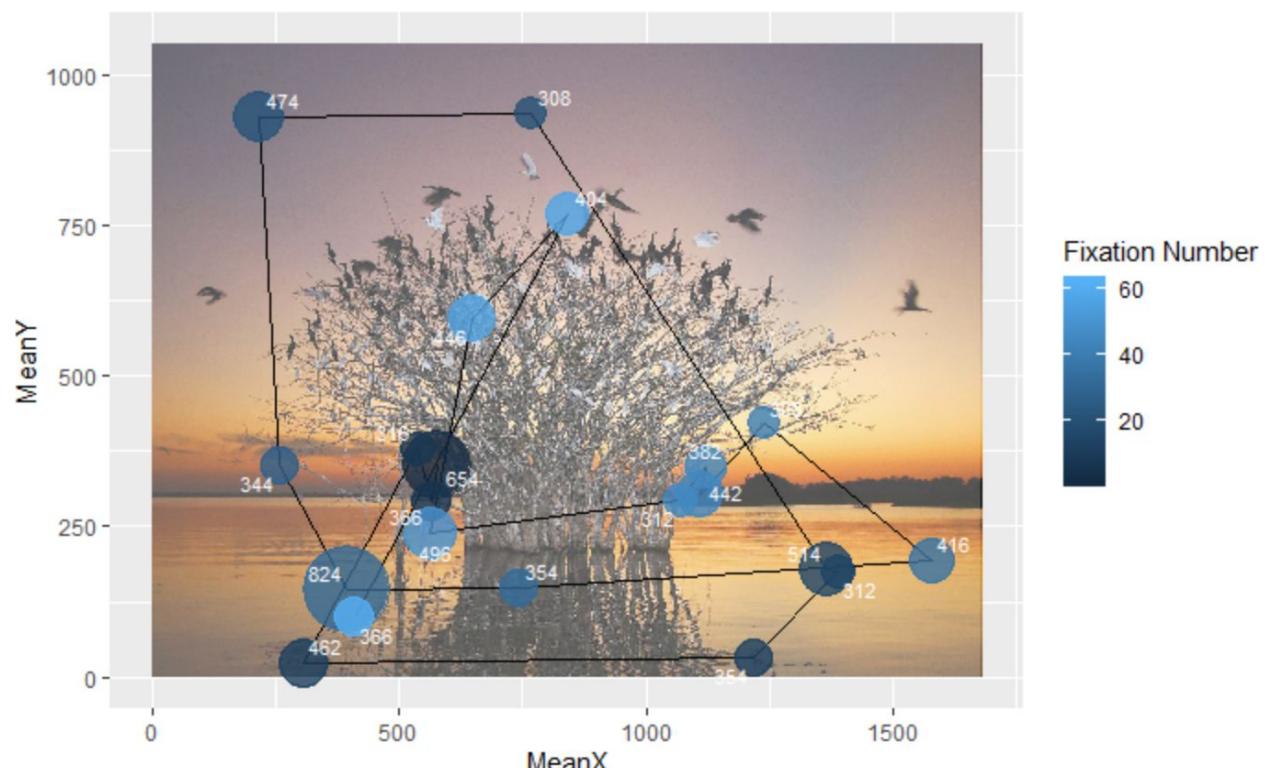


Figure 1, scanpath example with saccades and fixations

Statistical Modeling

We fitted a linear mixed-effects model (LMM) to the data, where saccade amplitude is predicted by the fixed effect condition (count/search), while taking the random effects of participants and stimuli into account. Further, these random effects of ID's and stimuli may vary according to condition.

$$\text{Saccade Amplitude} \sim 1 + \text{Condition} + (\text{Condition|ID}) + (\text{Condition|Stimulus})$$

After a visual inspection of the model residuals via a distribution plot, we find the data to be non-normally distributed. In particular, the distribution has long positive tails and is skewed towards zero, since there are naturally no negative values of saccade amplitudes. In an attempt to combat this and maintain our assumption of normally distributed data points, we decide to transform the saccade amplitude to log scale.

However, after comparing residuals of log scale and gaussian scale model with a simulation-based approach (DHARMA) (see *figure 2* and *figure 3*), we discovered two things. Firstly, the transformation did not remedy the non-normalcy of the data, secondly, both models have low Goodness of Fit (GOF). Despite this, we continue the analysis using the log-transformed data, because this is the method we often utilize, when we have non-normal positive, heavy-tailed distributions.

In order to transform the coefficients of the model (log-odds) back to the correct scale, informing us about the angle sizes of the visual field and thus amplitude of the saccades, we exponentiate the beta values.

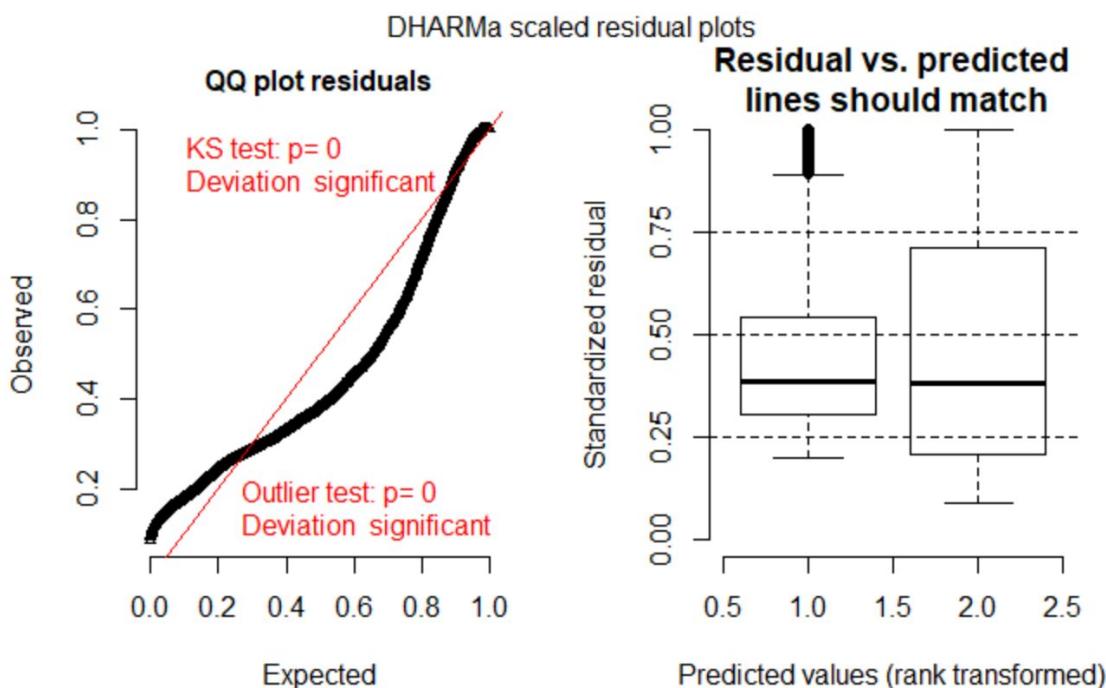


Figure 2 - residual plots from gaussian model

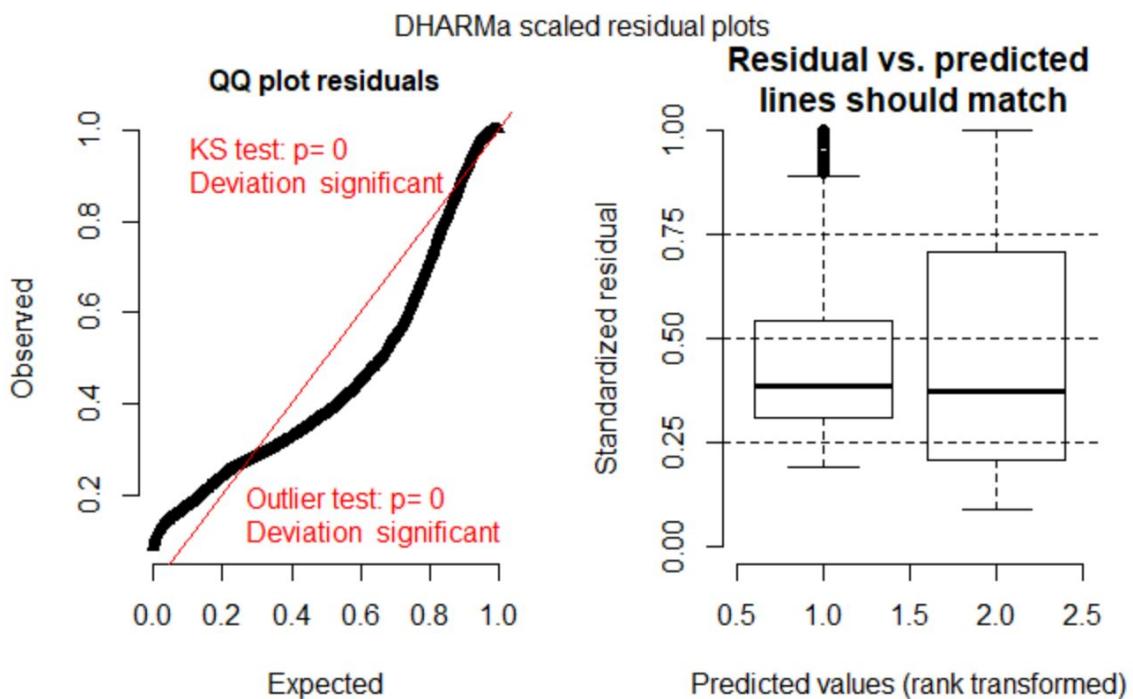


Figure 3 - residual plots from log-transformed model

Results

The results of the mixed effects analysis is depicted in *table 1* (on the log scale). According to the model the average amplitude in the foraging condition is larger by 1.8 degrees compared to the count condition.

Condition	Beta-value	SE	p-value
Search	0.53761	0.07907	<.001

Table 1 - results from model analysis

We also calculate the marginal R^2 of the model, which is 0.00529. This signifies that the model can account for 0.529 % of the variance.

In terms of whether or not the tasks will yield a data distribution with heavy tails, there seems to be a visual difference between the two conditions. In *figure 4*, we see a density distribution over the saccade amplitudes. We observe heavy-tails in both conditions, but the distribution in the foraging or search condition is more distinct. However, a formal analysis, such as the Hill Estimator, is needed to confidently draw a conclusion about the significance of this difference.

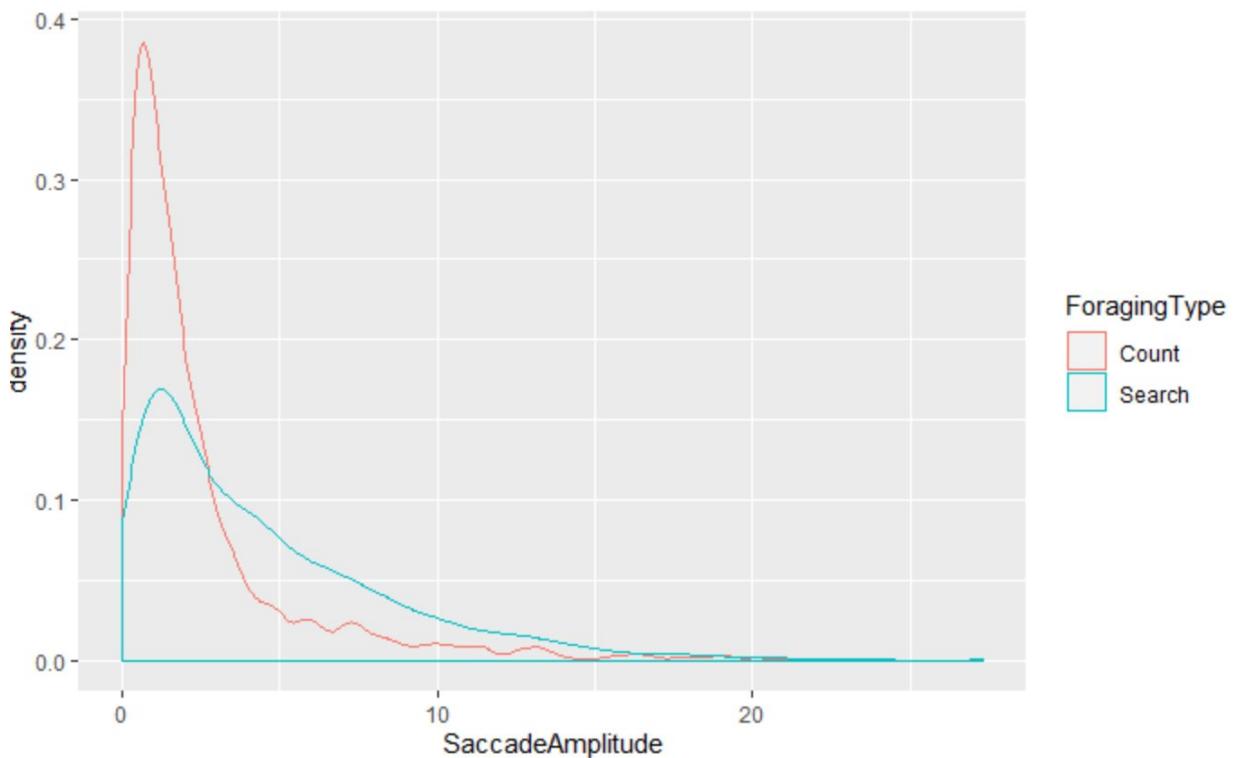


Figure 4 - density distribution of saccade amplitudes

Discussion

The results of the model analysis revealed that there is a significant difference between the duration of the saccades in foraging type tasks and counting tasks. As hypothesized, we found a larger average amplitude during foraging compared to the count condition, meaning that the foraging task elicits a response in the participants, which influences their scanpaths between fixations to become longer. This result is congruent with our hypothesis and implies that the task at hand makes a difference to how participants move their gazes, specifically that the saccade amplitudes are lengthened during visual foraging. See *figure 5*, which depicts the linear model superimposed on the distribution of saccade amplitudes between conditions. The search condition distribution has noticeably heavier tails compared to the count condition.

The results imply that condition and therefore top-down processes of the brain influence how visual tasks perform in modulating the saccade amplitudes. Nonetheless, there are several problems in the experiment, which need to be addressed. First off, having only 6 participants always imposes a great constraint on the generalizability of the results. Second, the simulation-based approach of the model's residual evaluation suggests that overall model performance is quite low. Therefore, it is good practice to be rather guarded when concluding on the results.

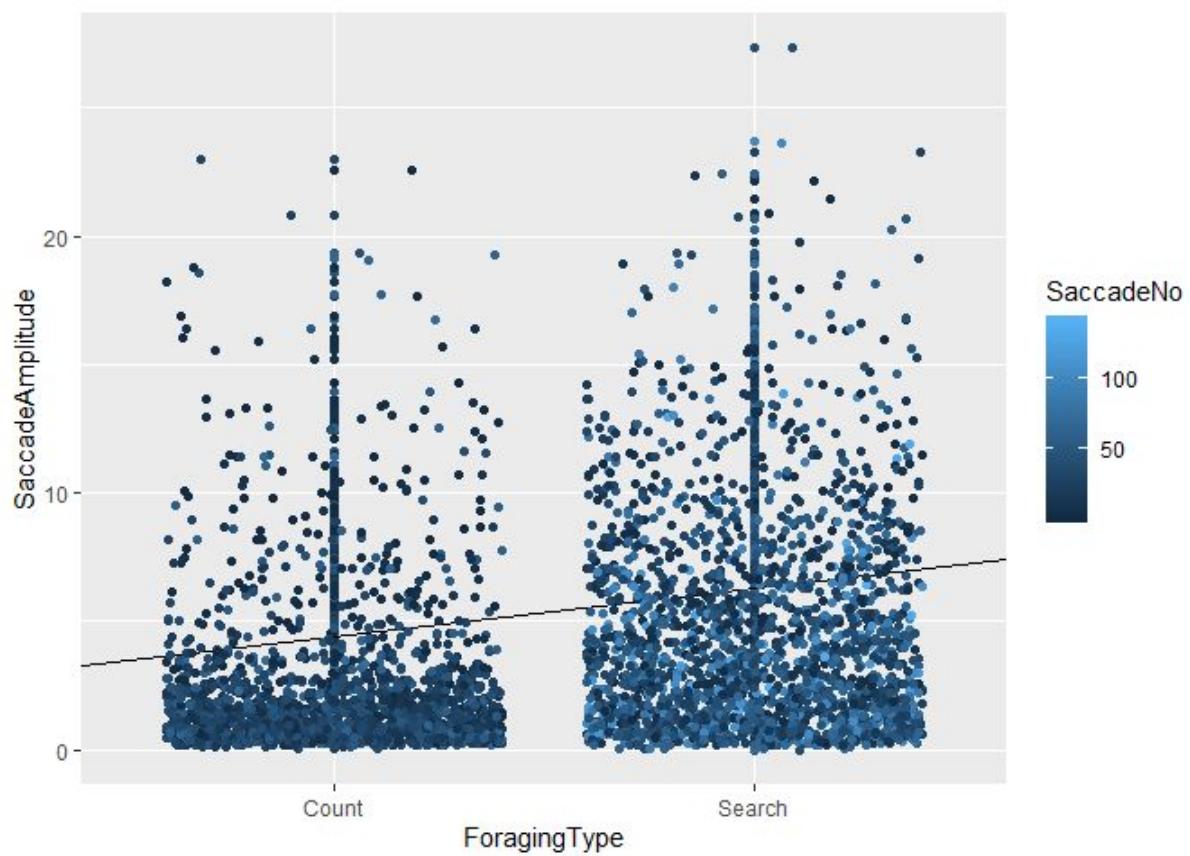


Figure 5 - saccade amplitude divided by condition with regression line

Social engagement experiment

Introduction

The conceptual research question of this study was to investigate how social cognition differs during social interaction compared to social observation. More specifically, we want to elucidate whether there is a difference in physiological arousal, determined by pupil size, depending on the social situation. We look at whether two types of ostensive cues used to simulate different states of social engagement, affect pupil size. In other words, does social engagement versus observation matter in how we process social information?

These research questions have implications for research on social cognition in everyday life and in clinical conditions. For instance, if we were able to distinguish social interaction from social observation, we would have a good argument to postulate that there might, in fact, be two separate models of social cognition. This could be employed in detecting impairments related to mental or developmental disorders that show deficits in social interactions like people on the autism spectrum.

Hypothesis

We hypothesize that interactive, engaging stimuli would elicit more emotional arousal, resulting in a greater pupil dilation than stimuli affording a more observational attitude in the participant. In other words, the condition relying on two ostensive cues rather than 1 or 2 would cause the greatest physiological arousal compared to other conditions (see the *Methods* section). Furthermore, we predict that there might be an interaction effect between the cues. We hypothesize that two ostensive cues in concert will elicit greater physical arousal than the cues on their own.

Methods

Participants

There were 6 participants in total in the social engagement experiment, all of which were 4th-semester Cognitive Science students at Aarhus University. One of the goals during data collection was to introduce confounding factors that would intervene with the eye tracker's ability to detect the participants' pupils effectively. Hence, most participants were wearing contact lenses or glasses and/or a lot of mascara on their eyelashes. This handicap was introduced to prompt our work with 'real', messy data.

Task

The task directly compared interactive and observational social cognition in order to test the hypothesis. Participants watched short videos with either a male or female actor. The actor in the video placed an object on the table performing either one, two, or

no social cues. The different video clips therefore varied in their level of social engagement, enabling us to test the hypothesis.

The experiment has four conditions comprised of a two times two factorial design; **(1)** ostensive and direct: An actor facing the direction of the participant (direct), exhibits an interaction-initiating cue by raising their eyebrow (ostensive) before performing the object-related gesture.

(2) non-ostensive and direct: An actor is facing participant (direct), but does not exhibit ostensive cues (non-ostensive)

(3) ostensive and averted: An actor sits at an averted angle, facing someone outside the perspective of the camera (averted) and exhibits the eye-brow lift and nod (ostensive).

(4) non-ostensive and averted: An actor sits at an averted angle, facing someone outside the perspective of the camera (averted) and doesn't exhibit ostensive cues (non-ostensive).

Shifting between the conditions with ostensive cues vs no ostensive cues and altering the direction of the actor would shift the experience of participants between observational mode and an interactive mode. The most interactive condition is (1), while the most observant condition is (4). This enables us to ask the question whether social observation is any different from social interaction..

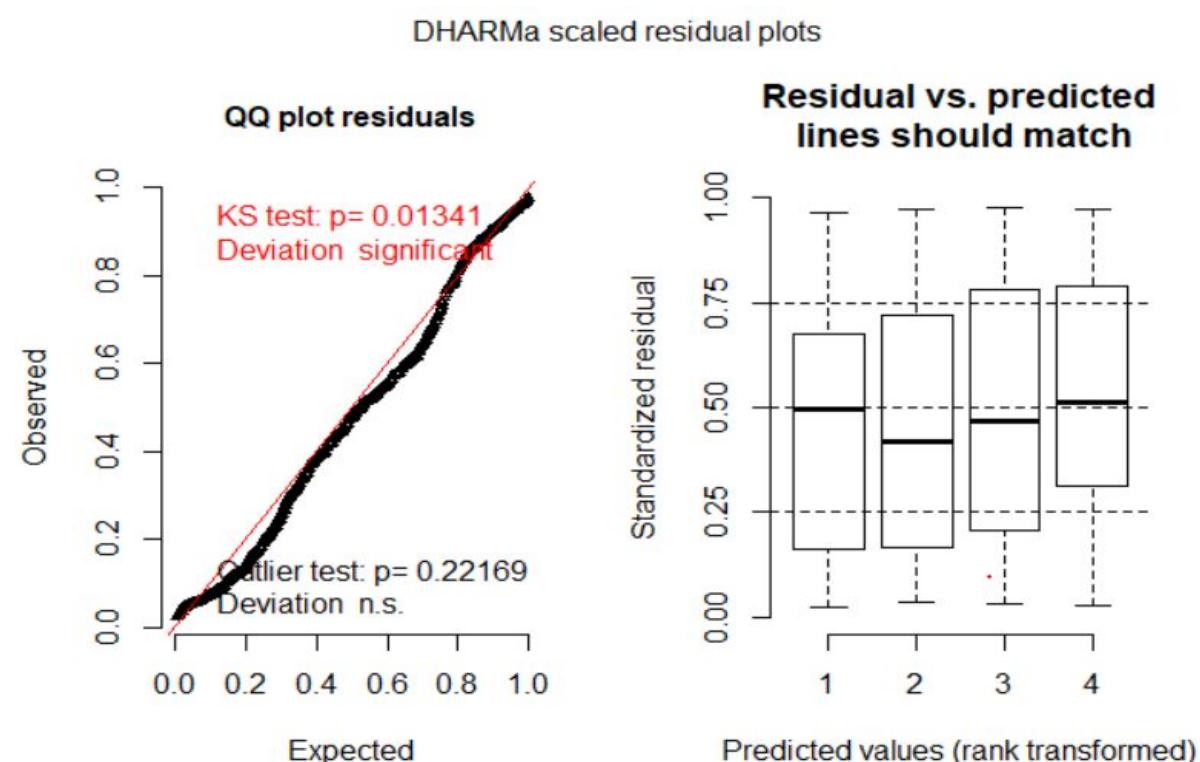


Figure 6 - residual plots from gaussian model Eye-Tracker Setup

The participants' gazes were measured with the EyeLink1000 at COBE Lab. The machine has a monocular measure, which means that we need to measure the participant's

dominant eye, which can be tested by the Miles test. After setting the sample rate at 500 Hz and illuminator power at 75%, calibration of the eye ensued, in order to make sure that the eye tracker could detect the pupil from different angles as accurately as possible

Preprocessing Eye-Tracking

Using the EyeLink1000 software, the visual eye-tracking data was translated and grouped into saccades and fixations of differing lengths. Consequently, the scan paths were determined. This data was then exported from the eye-tracker. Afterwards the data was cleaned by deleting a few nonsensical outliers, which were deemed artifacts from the eye-tracker.

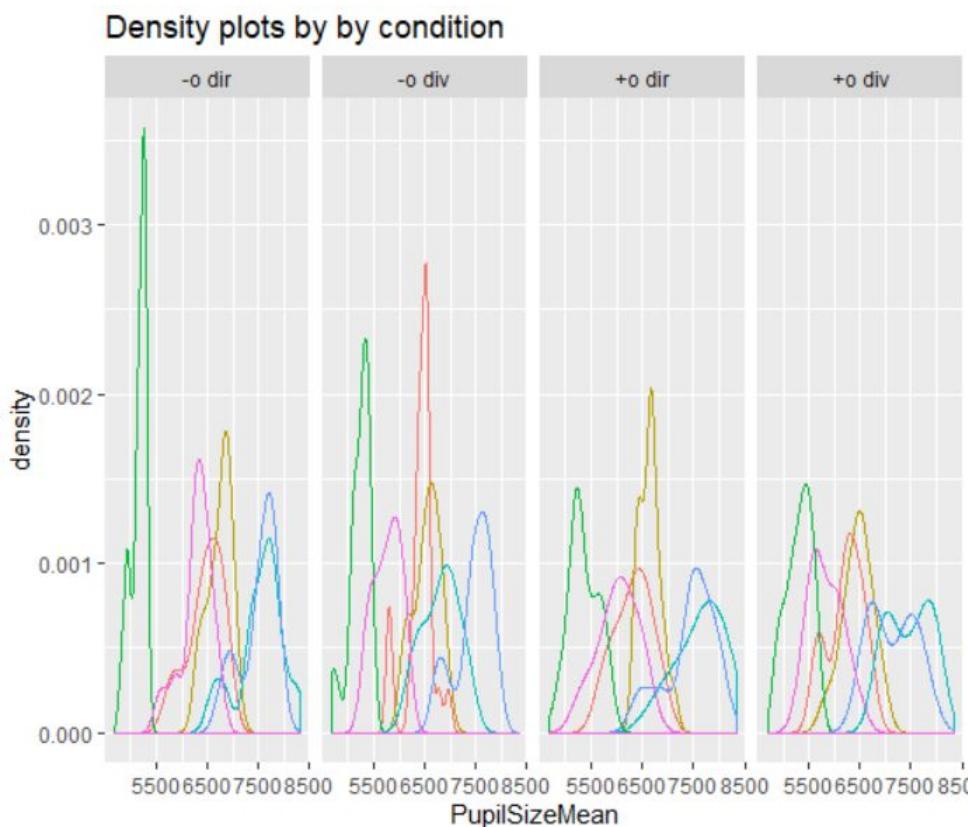


Figure 7 - density plot for each participant

Statistical Modeling

We used two-by-two factorial design (making up four conditions) with the main factors Direction and Ostension, each of them having two levels.

We fitted a linear mixed-effects model (LMM) to the data, where pupil size is predicted by the fixed effect Direction (direct/averted) and its interaction with the fixed effect Ostension (ostensive/non-ostensive), while taking the random effects of participants into account. Each participant is expected to have individual intercepts and slopes dependent on the interaction between Direction and Ostension.

$$\text{Pupil Size Mean} \sim 1 + \text{Direction} * \text{Ostension} + (1 + \text{Direction} * \text{Ostension} | \text{Participant ID})$$

We were able to assume normalcy of the data after looking at the model residuals using DHARMA (see *figure 6*) as well as the density plots in *figure 7*.

Results

The output of the model is reported in *table 2*. See *figure 8*, which depicts the distribution of the pupil size by condition in a box plot. Visually, it is already evident that the directed condition without ostension affords the largest pupil sizes, while the averted (div) condition without ostension seems to be correlated with smaller pupil sizes.

We do not report the R^2 because there are no significant effects to analyze,

Condition	Beta-value	SE	p-value
Direct	250.29	128.011	0.108
Ostension	45.65	138.480	0.755
Direct:Ostension	-92.69	131.254	0.511

Table 2 - results from model analysis

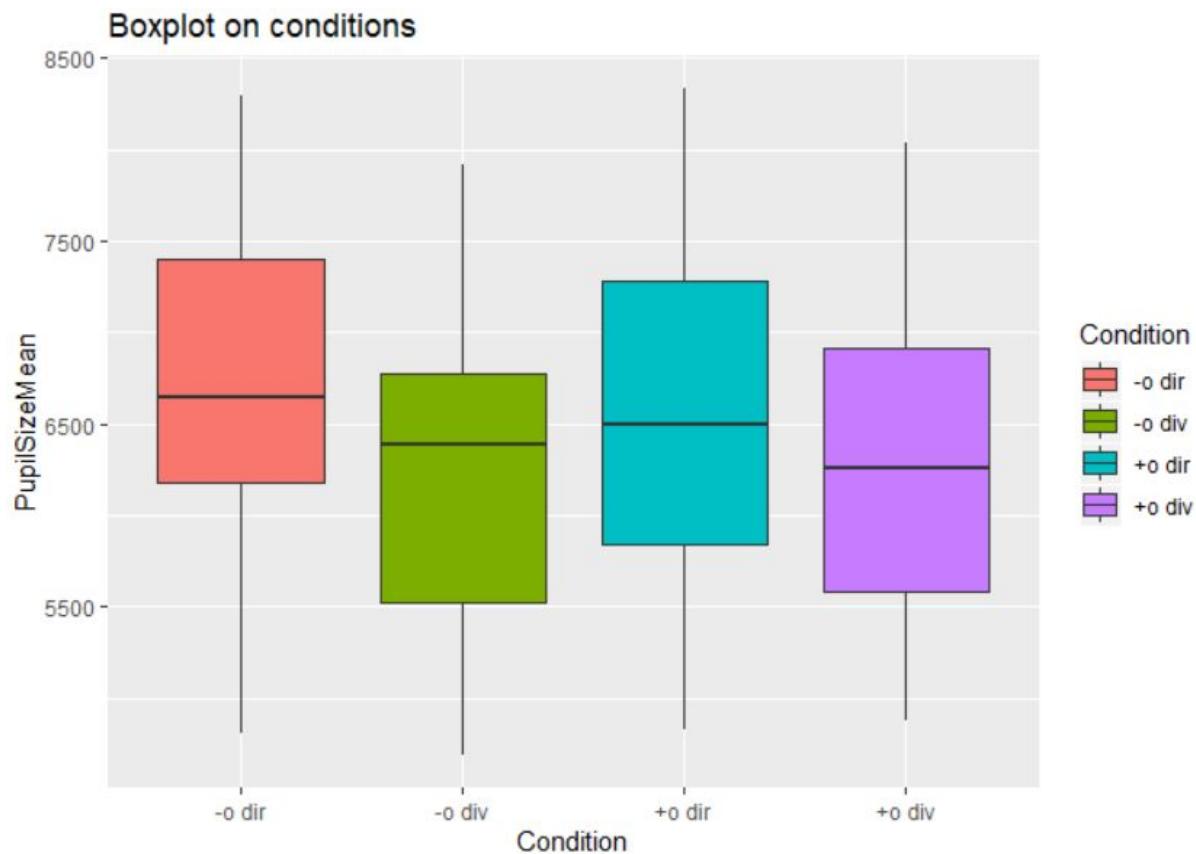


Figure 8 - boxplot over pupil size in the four conditions

Discussion

The results from our model imply no significant difference between the different social situations. According to the model, pupil size and therefore physical arousal increases, when the interlocutor is directed towards the participant compared to when they are facing in a different direction.

Further, pupil size is not significantly dependent on the interaction effect between direction and ostension. According to the results pupil size shrinks when the actor, who is directed towards the participant, raises their eyebrows in an ostensive gesture. On the other hand, when the actor is facing away from the participant, the ostensive cue causes the pupil size to increase, implying arousal. See *figure 9*, depicting the interaction effect, which, however, is non-significant.

Against the background of these results, we can conclude that the pupil size is not significantly different in the social observation and social interaction scenario. Our results do, therefore, not confirm our hypotheses. We also predicted to observe the most arousal when both social cues (Direction and Ostension) were present, but this was not the case. We found that the ostensive cue actually shrunk the pupil size when coupled with the direct condition, even though these results were, again, not significant.

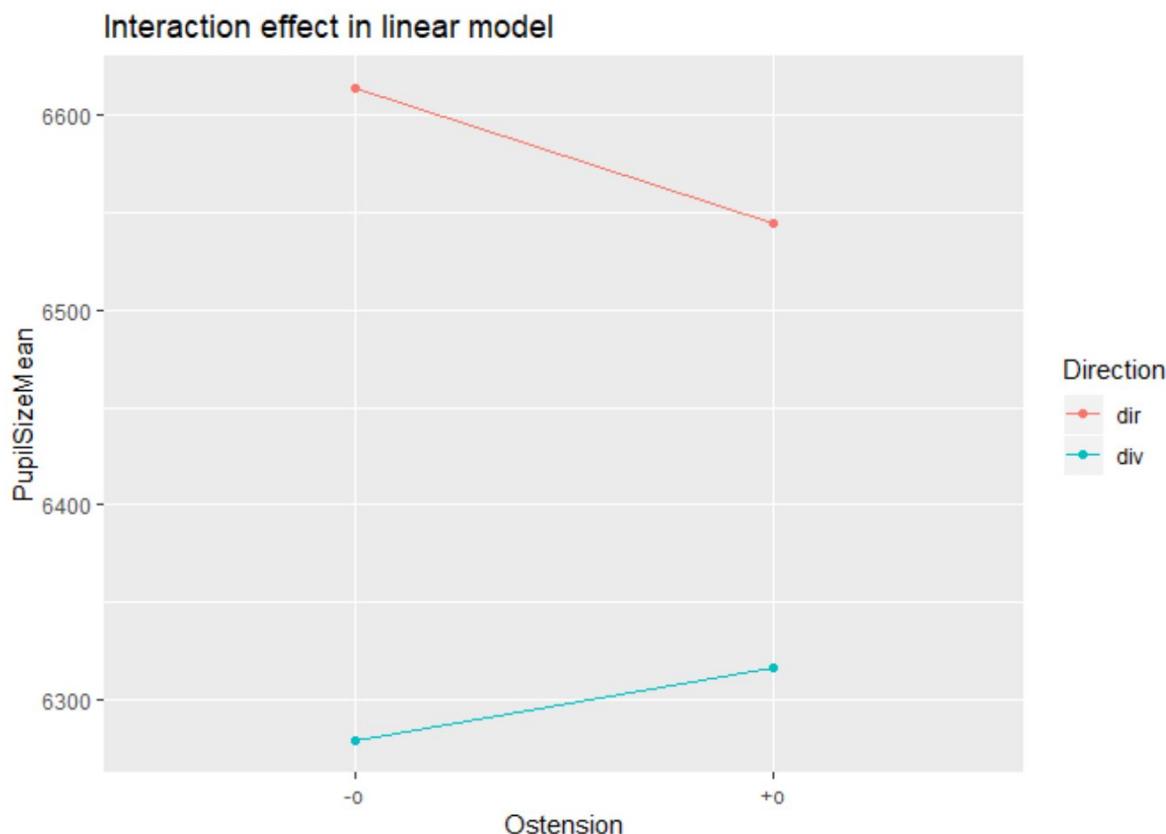


Figure 9 - plot of the interaction effect between ostension and direction

Package Reference

RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA
URL: <http://www.rstudio.com/>.

Hadley Wickham (2016). ggplot2: Elegant Graphics for Data Analysis.
Springer-Verlag New York. <https://CRAN.R-project.org/package=ggplot2>.

Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48.<doi:10.18637/jss.v067.i01>.

Kamil Slowikowski (2019). ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. R package version 0.8.1.:
<https://CRAN.R-project.org/package=ggrepel>

Florian Hartig (2020). DHARMA: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. R package version 0.2.7.:
<https://CRAN.R-project.org/package=DHARMA>

Assignment 2

Computational Modeling for Cognitive Science

Github:

[https://github.com/PeterThramkrongart/CompMod2/blob/master/assignment-2-last-document
.md](https://github.com/PeterThramkrongart/CompMod2/blob/master/assignment-2-last-document.md)

This assignment centres around evaluating rates from a binomial distribution, using the case of assessing the teachers' knowledge of Cognitive Science. In order to execute these assessments we utilize bayesian statistics. The key intuitions we gained from this assignment was how priors and sample sizes influence the posterior distribution in interactive ways.

PART ONE

1. Riccardo's estimated knowledge of CogSci using a uniform prior, Uniform(0,1)

Riccardo estimated knowledge of cognitive science is normally distributed with a MAP 0.5 (SD = 0.2) which is depicted in the plot below (figure 1) and was computed using a grid approximation.

Using a quadratic approximation the computed posterior distribution has the mean 0.5 (SD = 0.2). This is the same result we get using a grid approximation, because the MAP and the mean are the same value whenever we have a normal distribution.

According to our calculations, the probability that Riccardo knows more than chance, given a flat prior and a correct answering rate of 3:6, is 50 %. This is due to 50 % of the probability distribution being > 0.5 (see figure 1).

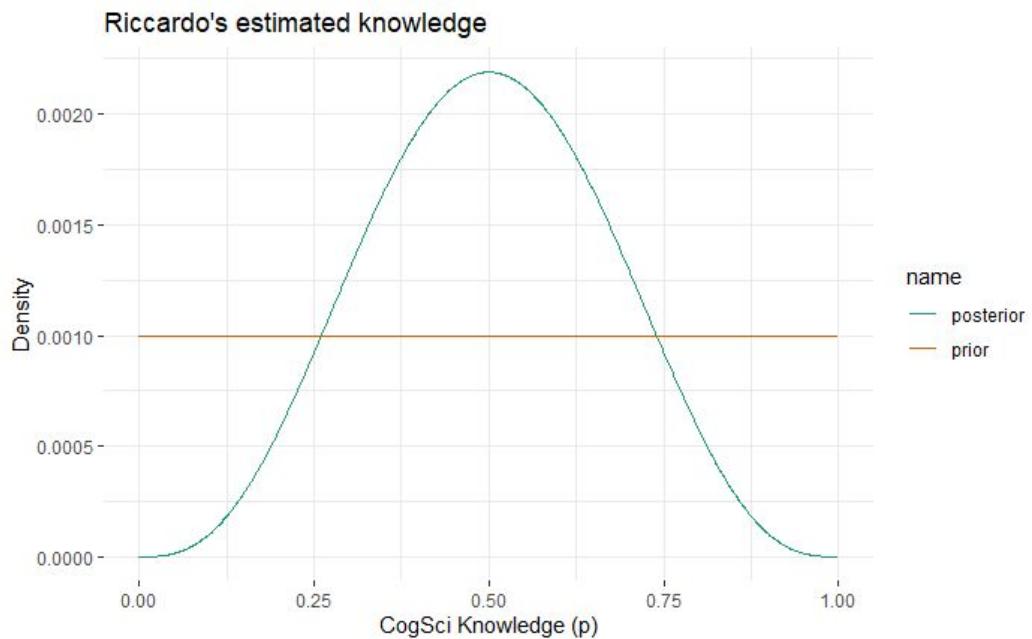


Figure 1 - Illustration of the Riccardo's estimated knowledge with the uniform prior

2. Estimation of all the teachers' knowledge of CogSci using grid approximation and the uniform prior, uniform(0,1)

Teacher	Lower HPDI of 80 %	Upper HPDI of 80 %	MAP	SD
Riccardo	0.26	0.70	0.50	0.17
Kristian	0.58	1	1	0.20
Josh	0.77	0.84	0.81	0.28
Mikkel	0.45	0.56	0.5	0.04

Table 1- Teachers' Cognitive Science knowledge estimated by grid approximation

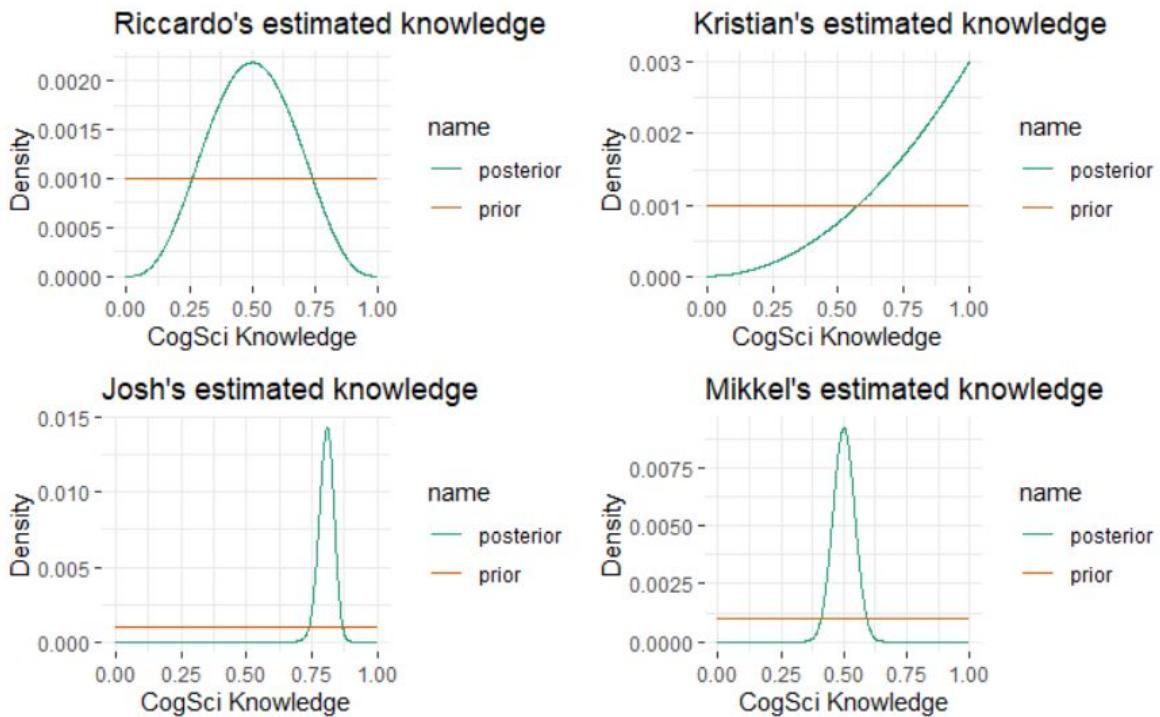


Figure 2 - Illustration of the uniform prior and calculated posterior of teachers' knowledge

Josh's narrow and relatively high HPDI (0.77, 0.84) and high MAP (0.81) reflects that he got many questions correct out of a large number. Compared to Kristian who, despite having the highest MAP value (1), has very broad HPDI (0.58, 1) due to very few answered questions. We can see that while MAP is primarily affected by how many questions were correctly answered, HPDI takes into account as well what is the overall number of questions and thereby gives us more information.

On the other hand, Mikkel and Riccardo have similar MAP estimates, but the spread is very different (see figure 2). Mikkel (MAP = 0.5, SD = 0.04) has answered more questions and therefore his estimates are more certain than Riccardo's (MAP = 0.5, SD = 0.17). Whether Riccardo or Mikkel is the most proficient in cognitive science is hard to say. The model is more certain that Mikkel knows *something* about cognitive science, but Riccardo has the opportunity to surprise more - both in a positive and negative direction.

When determining who is the smartest teacher, we can exclude Riccardo and Mikkel immediately as their performances are about chance level. To compare Joshua and Kristian, we contrasted how much the posterior probability of Kristian's knowledge lies below the posterior probability of Joshua's knowledge. Using this method we calculated that Joshua has a 51.35 % chance of being smarter than Kristian which almost equals to chance. Furthermore, there is too much uncertainty due to Kristian's low sample size to confidently say which teacher is the most knowledgeable.

3. Estimation of teachers' knowledge given an informative prior - Normal(0.8, 0.2)

Teacher	Lower HPDI of 80 %	Upper HPDI of 80 %	MAP	SD
Riccardo	0.48	0.80	0.65	0.13
Kristian	0.69	1	0.89	0.13
Josh	0.77	0.84	0.80	0.03
Mikkel	0.45	0.56	0.51	0.04

Table 2 - Teachers' knowledge estimated with informed priors

The plots (displayed in figure 3) illustrate the prior and posterior distribution of each teacher given the informative prior. These plots and table 2 show that the more informed prior has changed teachers' posteriors in a way so that they were shifted to the right. This implies that the assumption (prior) that our teachers are pretty smart increases the probability of their predicted knowledge being higher.

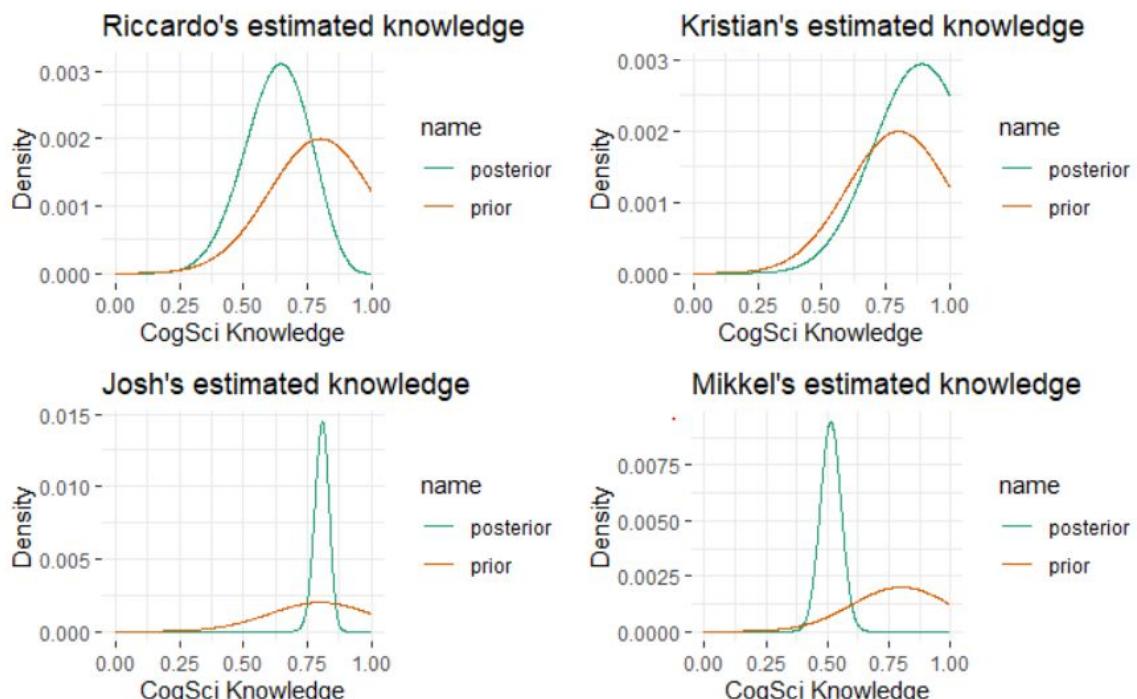


Figure 3 - Illustration of prior and posterior distribution of teachers' knowledge

4. Estimation of teachers' knowledge given a larger data set

If we multiply all teachers' answers with 100, we get new values for some teachers while other estimates remain close to constant.

For the new priors, if the teacher's initial sample size was small, multiplying their sample size with 100 makes the HPDI more narrow. This makes perfect sense since the model becomes more certain. Moreover, more weight will be given to the observed data, reflected in the MAP which is pulled further in the direction of the observed data compared to the prior. We don't see a big difference between uniform and informed priors since the priors have been "overruled" by the sheer amount of data.

This means that Kristian is suddenly more likely to be the most proficient teacher in cognitive science knowledge. It also implies that when the updating sample size becomes larger, the prior becomes less and less influential.

Teacher	Lower HPDI of 80 %	Upper HPDI of 80 %	MAP	SD
Riccardo	0.47	0.52	0.50	0.02
Kristian	0.99	1	1	0.005
Josh	0.81	0.81	0.81	0.003
Mikkel	0.49	0.50	0.50	0.004

Table 3 - Estimates of teachers' knowledge using the uniform prior

Teacher	Lower HPDI of 80 %	Upper HPDI of 80 %	MAP	SD
Riccardo	0.48	0.53	0.50	0.02
Kristian	0.99	1.0	1	0.005
Josh	0.80	0.81	0.80	0.003
Mikkel	0.49	0.50	0.51	0.004

Table 4 - Estimates of teachers' knowledge using the informed prior

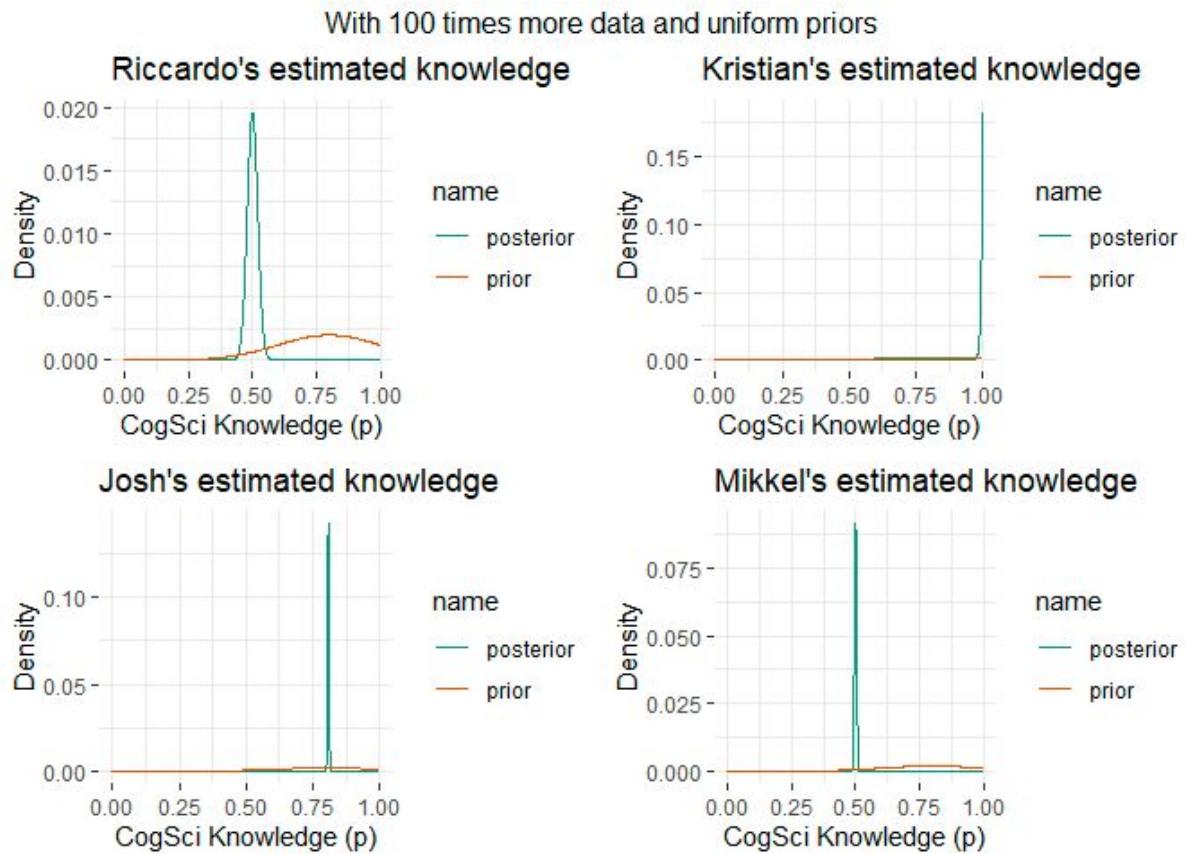


Figure 4 - Illustration of prior and posterior distribution of teachers' knowledge with uniform priors and 100 times more data

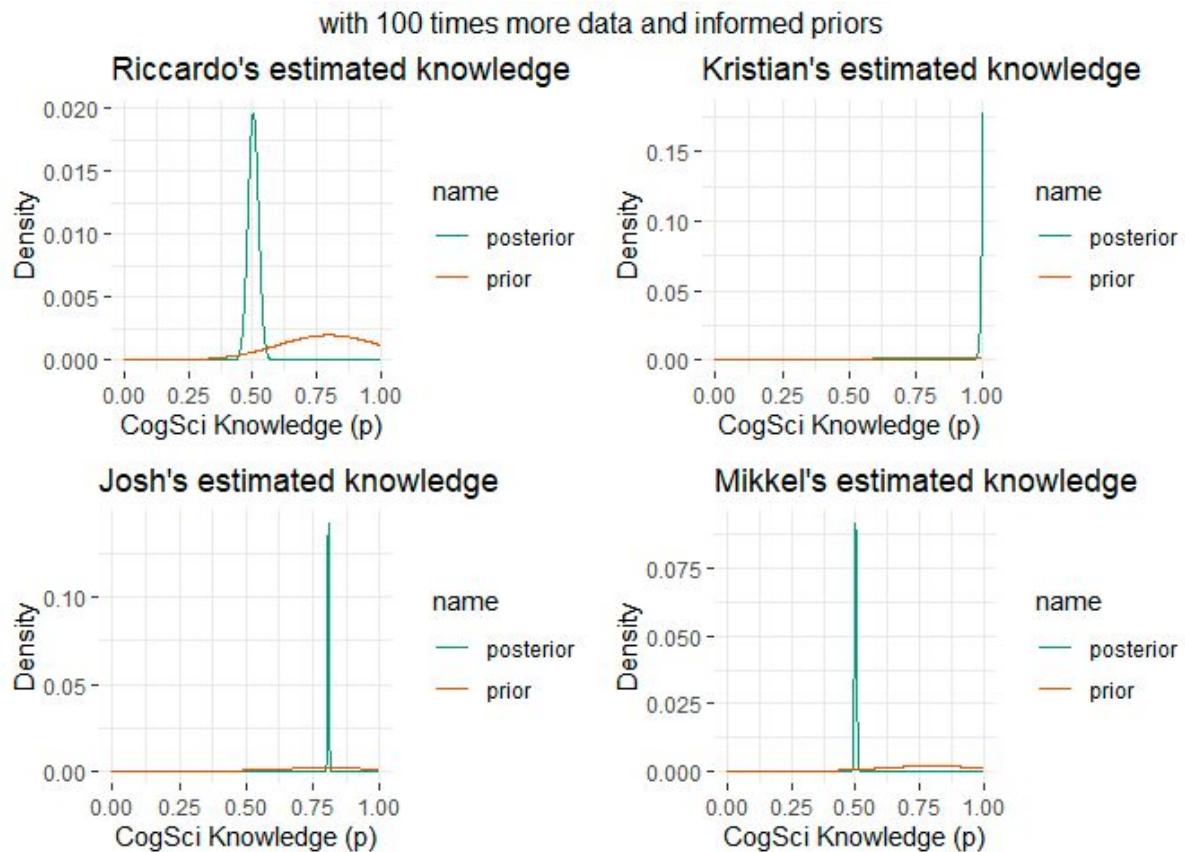


Figure 5 - Illustration of prior and posterior distribution of teachers' knowledge with informed priors and 100 times more data

5. Imagine you're a skeptic and think your teachers do not know anything about CogSci, given the content of their classes. How would you operationalize that belief?

If we were skeptical of our teachers' knowledge, we would need to adjust the prior accordingly. Assuming a normal distribution of teachers' knowledge, we would set a mean of the new prior to the lowest possible value: 0. Furthermore, we can reflect our skepticism in the standard deviation. The tighter it is, the more the model will be likely to pull the estimate down to zero, thus being skeptical. However, should we set it too strict, e.g., 0.001, we would probably run into divergence problems.

PART TWO

Last year we assessed the teachers' knowledge. This year we want to re-test them and assess whether our models are producing reliable predictions. Frequentist statistics use a machine-learning style to assess predictions (e.g. rmse on testing datasets). Bayesian stats makes things a bit more complicated.

1. Discussion of differences in the assessment of prediction performance in Bayesian vs. frequentist models

Assessment difference between **Bayesian** and **frequentist** models has its roots in the different ways the two interpret probabilities.

Frequentists consider probabilities as frequencies and they see data as a repeatable random sample (random variable) with a specific frequency/probability, which is defined as the relative frequency/probability of an event as the number of trials approaches infinity. In frequentists' view, events have a specific probability and the variation is in our sampling. Their model prediction performance is judged based on single-point estimates such as confidence intervals, hypothesis testing with p-values etc.

In contrast, **Bayesians** consider probabilities as degrees-of-belief, where a probability distribution is interpreted as quantifying our uncertainty about the world. In particular, this means that we can meaningfully talk about probability distributions of parameters, since even though the parameter is fixed, our knowledge of its true value may be limited. Bayesians assess model predictions appropriately measuring and accounting for the uncertainty in their models and parameters after looking at how a new piece of information has updated the prior and improved the certainty of the posterior.

2. Model prediction errors for the teachers given new data

Riccardo: In Riccardo's case, we can see from the histogram in figure 4 that the model almost never overpredicts Riccardo's knowledge. Instead it usually underpredicts his number of correct answers. We can see by the distribution that the model most often underpredicts by a value of 3.

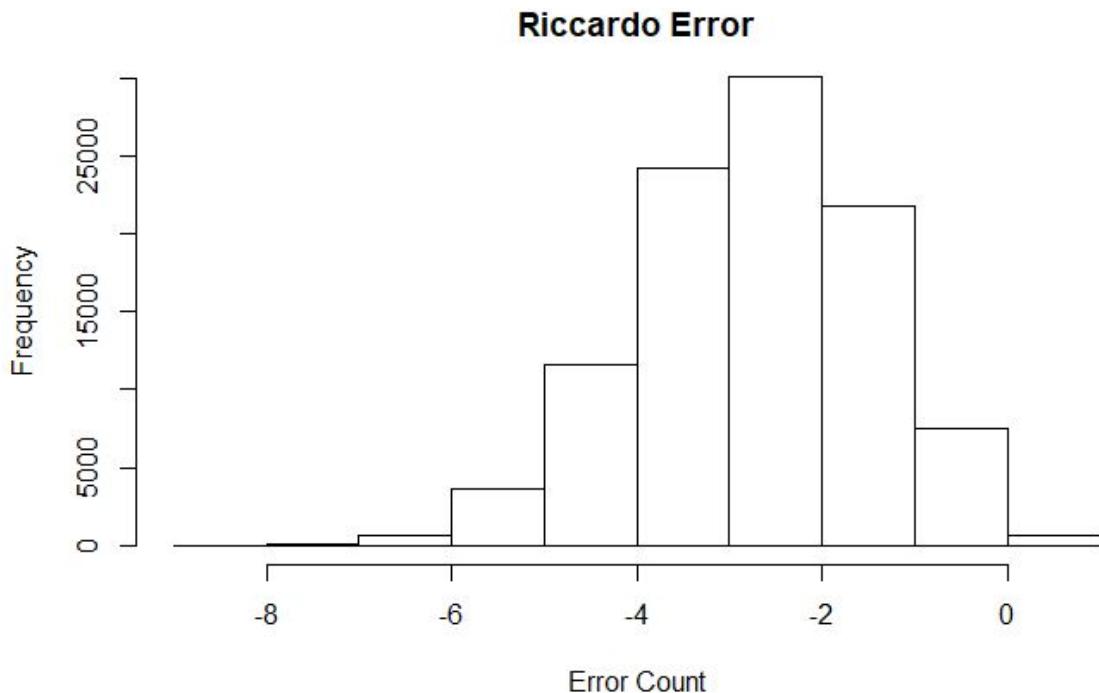
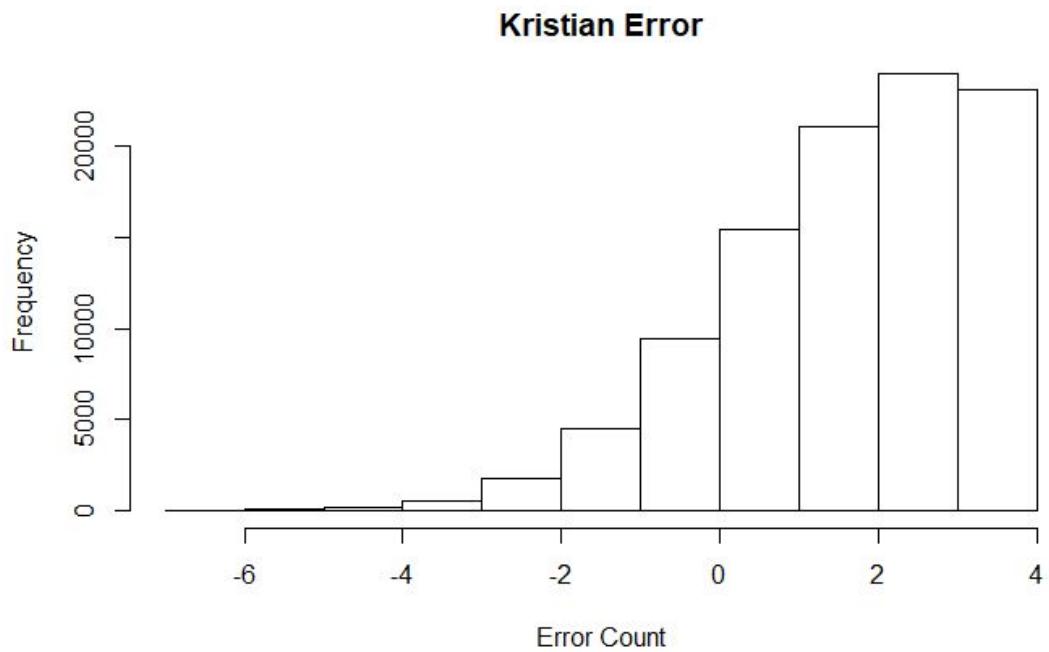


Figure 6 - The distribution of error from Riccardo's model given the observed data

Kristian: Looking at the histogram (figure 5) depicting the error of the model concerning Kristian's knowledge, we see that it both underpredicts and overpredicts according to the observed data. However, the model is most likely to overpredict, illustrated by the overweight of the distribution being above 0.



Mikkel: As for Mikkel, figure 7 illustrates how the model is more likely to underpredict, although the model overpredicts as well. However, since Mikkel has answered 66 questions, and his error centers around -2, the underprediction is arguably less prominent than in the model's other cases.

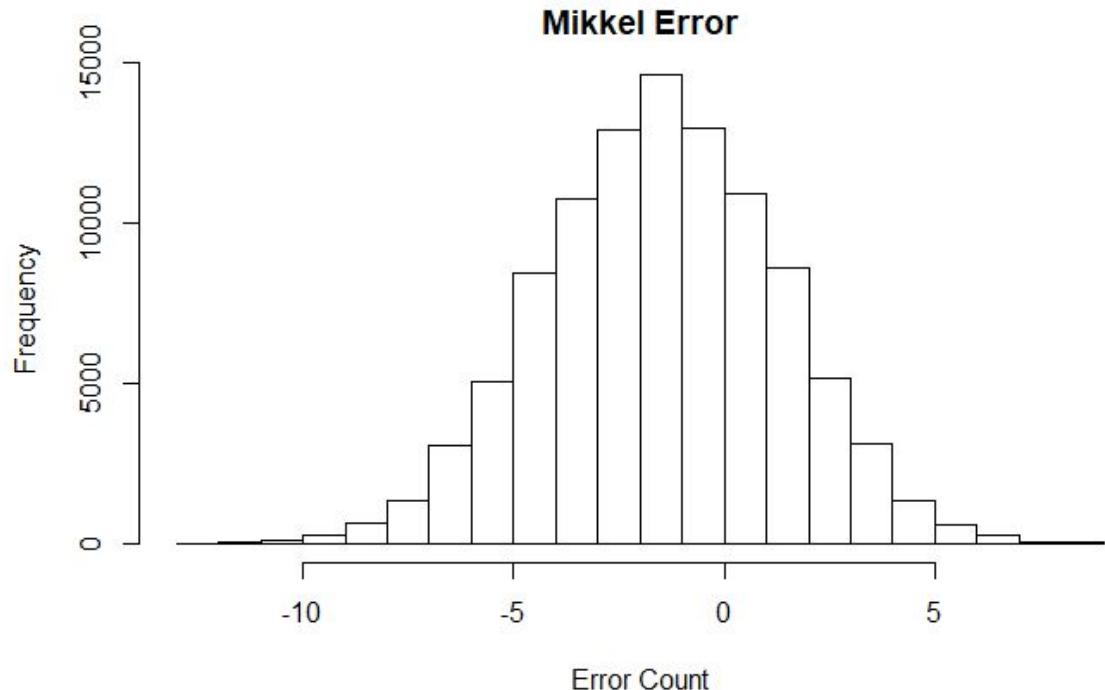


Figure 9 - The distribution of error from Mikkel's model given the observed data

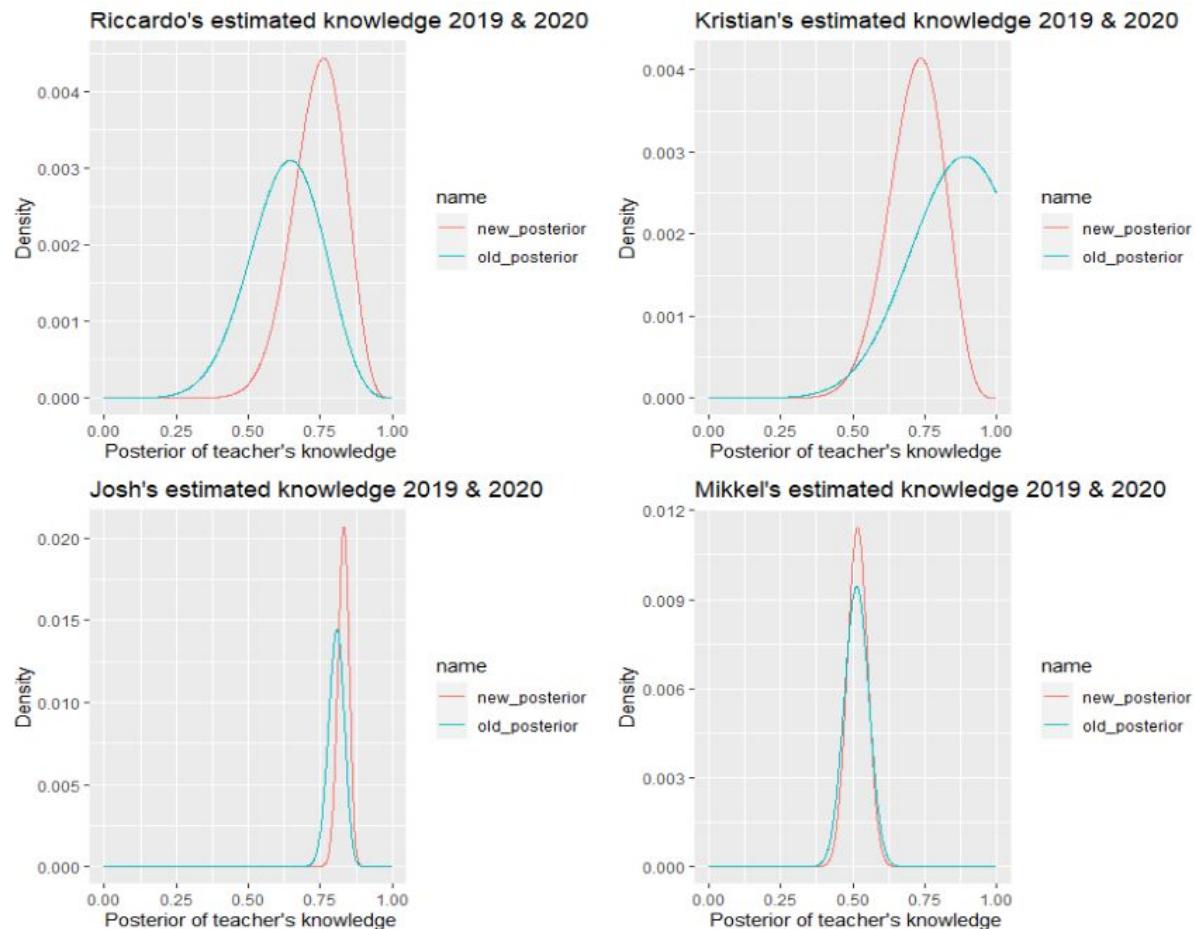


Figure 10 - Teachers' posterior given the old data compared to the updated posterior

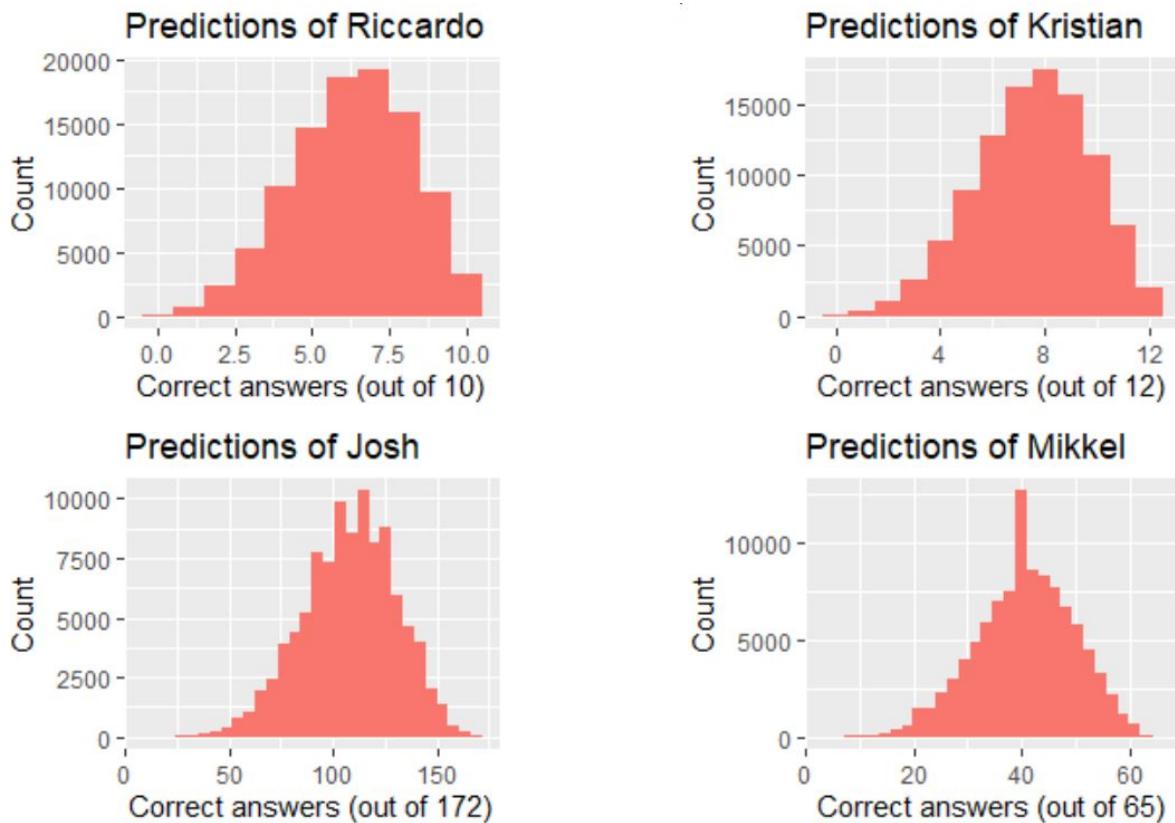


Figure 11 - Histograms displaying the predictions sampled from the posterior

Sources:

1. <https://cxl.com/blog/bayesian-frequentist-ab-testing/>
2. https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/readings/MIT18_05S14_Reading20.pdf
3. <https://www.probabilisticworld.com/frequentist-bayesian-approaches-inferential-statistics/>
4. <http://jakevdp.github.io/blog/2015/08/07/frequentism-and-bayesianism-5-model-selection/>
5. <https://www.probabilisticworld.com/intuitive-explanation-p-values/>

Assignment 3

Computational Modeling for Cognitive Science

GitHub Link:

https://github.com/PeterThramkrongart/CompMod-3/blob/master/Assignment3_Peter_Bella_Jakub_Bianka_Ruta.md

Introduction

Patients diagnosed with Schizophrenia are oftentimes found to possess altered control and distinction of self-other representations (e.g. hearing of voices or mind reading). Therefore this assignment was created with the aim of investigating this matter. In particular, we have examined altercentric intrusions (inability to inhibit social information) to see whether it can truly be associated with Schizophrenia and these aforementioned psychotic symptoms. In addition, we have also checked whether apathy (severe lack of motivation in taking care of oneself, e.g. getting out of bed or showering to show up for work) could be in relation with altercentric intrusions. In order to test our research questions we built multiple regression models. The data used in this assignment was simulated data. This is because the intention of this assignment is to explore some of the biases that can be introduced and the effects that can be unmasked when using multivariate modelling.

Hypotheses

In this paper we aim to investigate two distinct hypotheses:

H1: Schizophrenia is related to altercentric intrusions

H2: Altercentric intrusion is related to Schizophrenic symptoms (i.e. delusions of voice hearing & mind reading)

PART ONE

1.1. Altercentric intrusion in schizophrenia and controls

Model and results

In order to check the hypothesis that altercentric intrusion is higher in schizophrenia patients compared to controls, we built a univariate regression model with altercentric intrusion as the outcome and diagnosis as a predictor.

Model 1: Altercentric Intrusion diagnosis model

Altercentric Intrusion ~ 0 + Diagnosis

Since we are utilizing Bayesian modeling, in addition to designing the model, we assign priors for all of the parameters in the model. These priors define the initial information state of the model, before seeing the results. The prior for sigma representing the expected average error of our model is deduced from a standard deviation of altercentric intrusion. The beta prior was constructed to accommodate the entire distribution of altercentric intrusion, but still having a regularization function to prevent from overfitting, (see the Figure 14 for the prior check in the appendix).

Priors:

$$\beta \sim \text{Normal}(4, 1)$$

$$\sigma \sim \text{Normal}(1, 2)$$

The results of the model indicate a credible difference in altercentric intrusion in the two groups, supporting our hypothesis ($b = 0.36$, $\text{CIs} = 0.16, 0.56$, $\text{ER} = 665.67$). The difference in spread of data is depicted in Figure 1.

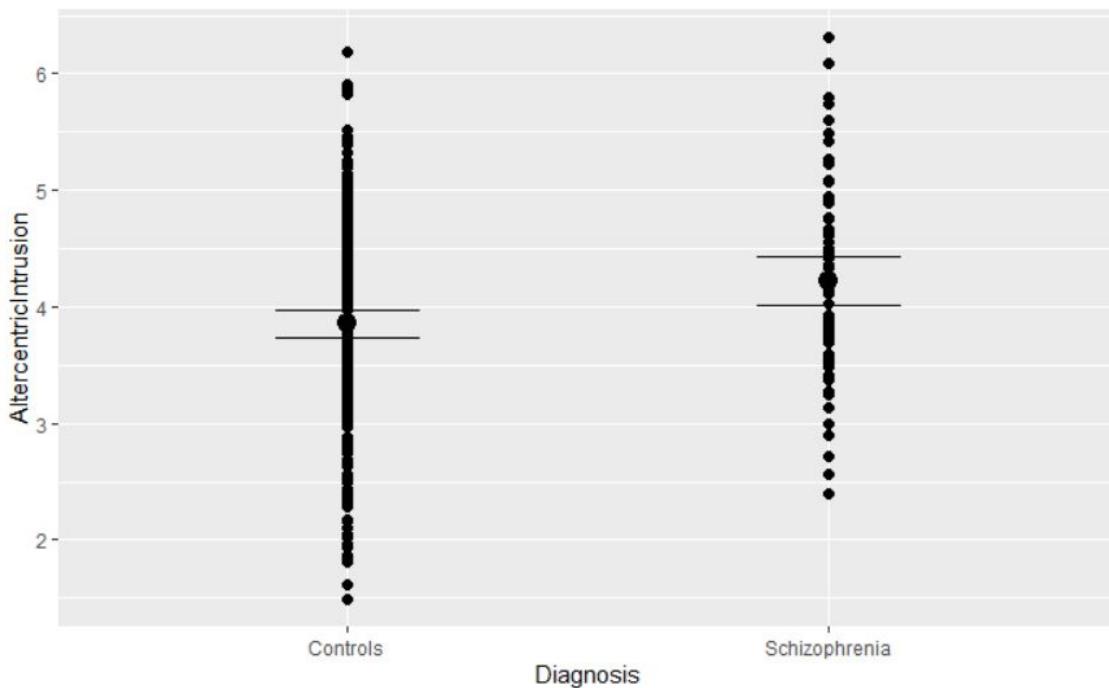


Figure 1: Illustration of altercentric intrusion in both Schizophrenia and Controls

Controls showed an on average altercentric intrusion effect of 3.86 (CIs = 3.75, 3.98) and schizophrenia of 4.22 (CIs = 4.02, 4.43). The model had no divergences, a Rhat of 1, and an effective sample size above 2000 for both Bulk and Tail.

The results of the analysis indicates that it is credible that altercentric intrusion is more prevalent among people with schizophrenia than among controls. This is evidence in favor of our first hypothesis, H1. Looking at the posterior predictive check (see Figure 15 in the appendix), we can see that your model given the predictor is fairly accurate at predicting new data, hence we can conclude that model makes sense. The update check plot (see in the Figure 16 in the appendix) shows the improvement from posterior, initial knowledge, to the posterior after fitting the model to data. As the posterior shrinks, it becomes more certain and learns more from the data.

1.2. Altercentric intrusion relation to separate symptoms in schizophrenia

Models and the results

In order to figure out which symptoms are related to altercentric intrusion in schizophrenia patients, we built univariate regression models and assigned priors for all of the parameters in the models. As a part of the preprocessing, we scaled the data so that the symptom severities would be comparable. Furthermore, we only used the data from the diagnosed individuals due to the nature of the question.

In the analysis, we model altercentric intrusions as the outcome and the separate symptoms, namely, voice hearing, mind reading, and apathy, as predictors. Moreover, to unmask any shared variance, we include two multivariate models.

Due to the fact that all variables were generated on the same scale, we are able to use identical priors across all models. The prior was designed following both conceptual knowledge and trial-and-error learning with the aim of model's regularization. The intercept prior was set at zero as we scaled the data. In terms of sigma, we deduced the value from a distribution of altercentric intrusion. Lastly, the beta prior was centered at zero due.

We have run the prior and posterior predictive checks, also, updating checks with hypothesis testing for all models, for obtained plots see the appendix Figures 17-29. For all of the models the posterior is more certain than the priors, therefore we can assume that the models have learned something from the data.

Priors:

$$\alpha \sim \text{Normal}(0, 1)$$

$$\beta \sim \text{Normal}(0, .3)$$

$$\sigma \sim \text{Normal}(1, 2)$$

Model 2: Voice Hearing model

$$\text{Altercentric Intrusion} \sim 1 + \text{Voice Hearing}$$

The outcome of model 2 indicates a negligible positive effect of voice hearing on altercentric intrusion in support of our hypothesis ($b = 0.07$, $\text{CIs} = -0.15, 0.3$, $\text{ER} = 2.28$). The model had no divergences, a Rhat of 1 and Effective sample sizes above 2000 for both Bulk and Tail. The regression line is depicted in figure 2.

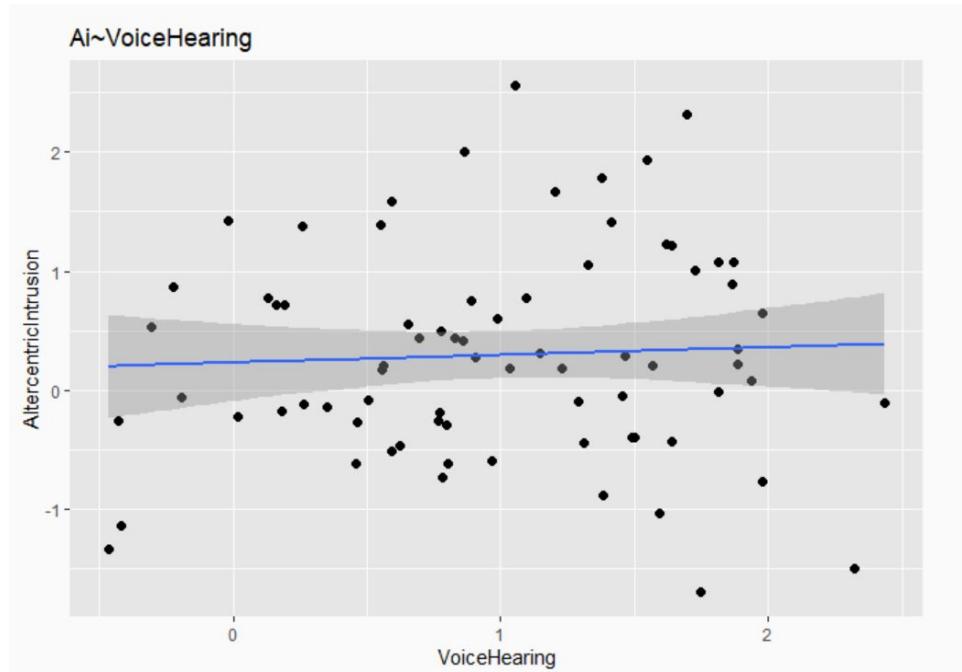


Figure 2: Plotting a regression line against the data

Model 3: Mind Reading model

$$\text{Altercentric Intrusion} \sim 1 + \text{Mind Reading}$$

Model 3 indicates a negligible positive effect of mind reading on altercentric intrusion, supporting our hypothesis ($b = 0.07$, CIs = $-0.11, 0.25$, ER = 2.87). The model had no divergences, a Rhat of 1 and Effective sample sizes above 2000 for both Bulk and Tail. The regression line is depicted in figure 3.

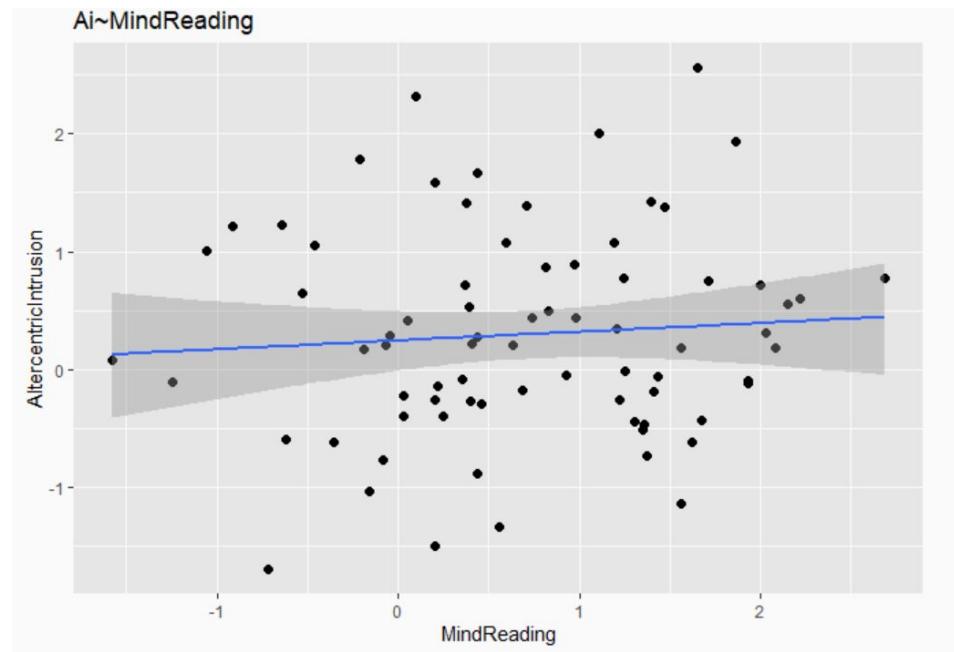


Figure 3: Plotting a regression line against the data

Model 4: Apathy model

$$\text{Altercentric Intrusion} \sim 1 + \text{Apathy}$$

Model 4 indicates a credible negative effect of apathy on altercentric intrusion, which seems to be against our hypothesis ($b = -0.23$, $\text{CIs} = -0.50, 0.04$, $\text{ER} = 3.66$), which predicts no effect of apathy on AI. The model had no divergences, a Rhat of 1 and Effective sample sizes above 2000 for both Bulk and Tail. The regression line is depicted in figure 4, showing a negative correlation between apathy and AI.

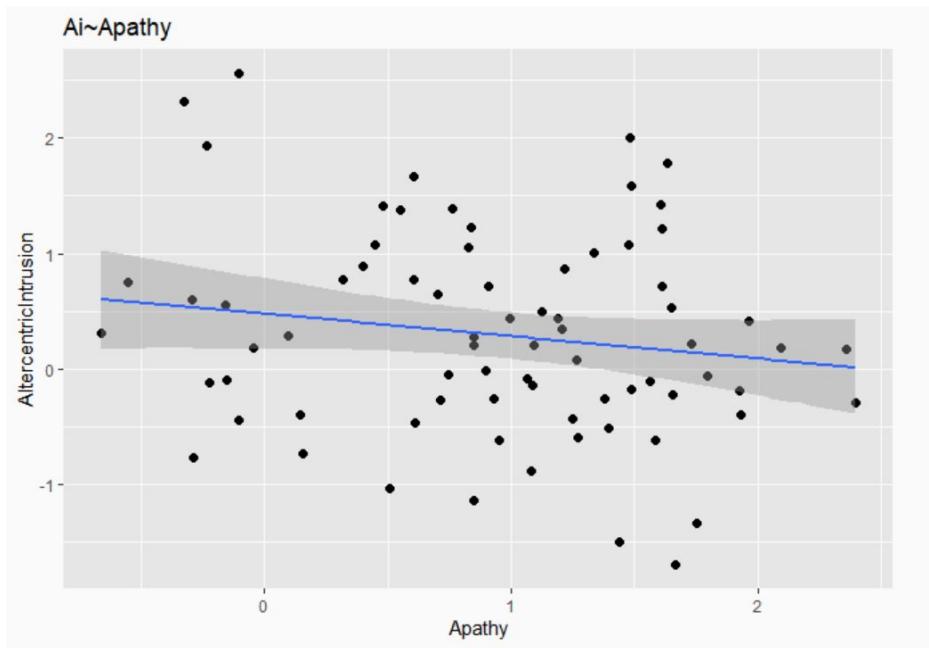


Figure 4: Plotting a regression line against the data

Model 5: Mind reading and Voice hearing model

$$\text{Altercentric Intrusion} \sim 1 + \text{Voice Hearing} + \text{Mind Reading}$$

Model 5 is a multivariate model, which examines how much of the variance either voice hearing or mind reading can explain on its own, when the other symptom is taken into account. The outcome indicates a slight effect of voice hearing on altercentric intrusion ($b = 0.11$, $\text{CIs} = -0.19, 0.39$, $\text{ER} = 3.42$), as well as a slighter effect of mind reading on AI ($b = 0.10$, $\text{CIs} = -0.13, 0.32$, $\text{ER} = 4.46$). See figure 5. The model had no divergences, a Rhat of 1 and Effective sample sizes above 2000 for both Bulk and Tail. Granted, while these results are pointing in the direction of our hypothesis, they are neither particularly satisfying nor convincing.

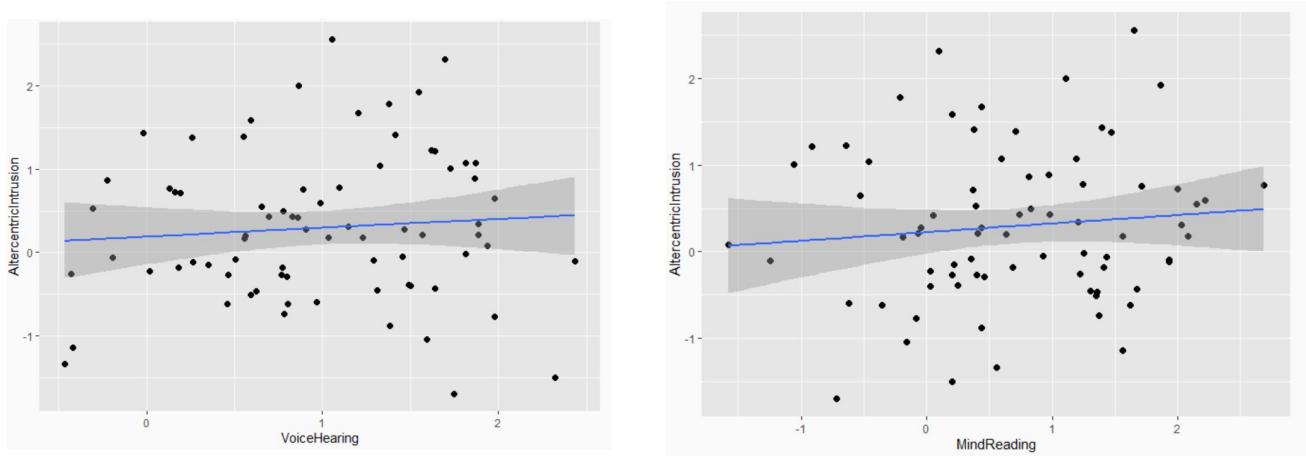


Figure 5: Plotting a regression line against the data

In order to see how much each of the predictors is contributing to explaining the altercentric intrusion when they are modeled together, we built a multivariate regression model with all of the symptoms as predictors and altercentric intrusion as an outcome. The data still consists of solely positively diagnosed individuals.

Model 6: Mind reading, Voice hearing and Apathy model

$$\text{Altercentric Intrusion} \sim 1 + \text{Voice Hearing} + \text{Mind Reading} + \text{Apathy}$$

The final model in this section, model 6, indicates a barely credible positive effect of voice hearing and mind reading, which supports our hypotheses, and a slightly more credible negative effect of apathy, which rejects our hypothesis once more: mind reading ($b = 0.04$, CIs = $-0.17, 0.24$, ER = 1.58), voice hearing ($b = 0.05$, CIs = $-0.21, 0.3$, ER = 1.77), apathy ($b = -0.17$, CIs = $-0.42, 0.07$, ER = 7.25). The model had no divergences, a Rhat of 1 and Effective sample sizes above 2000 for both Bulk and Tail. See figure 6 for regression lines from individual predictors.

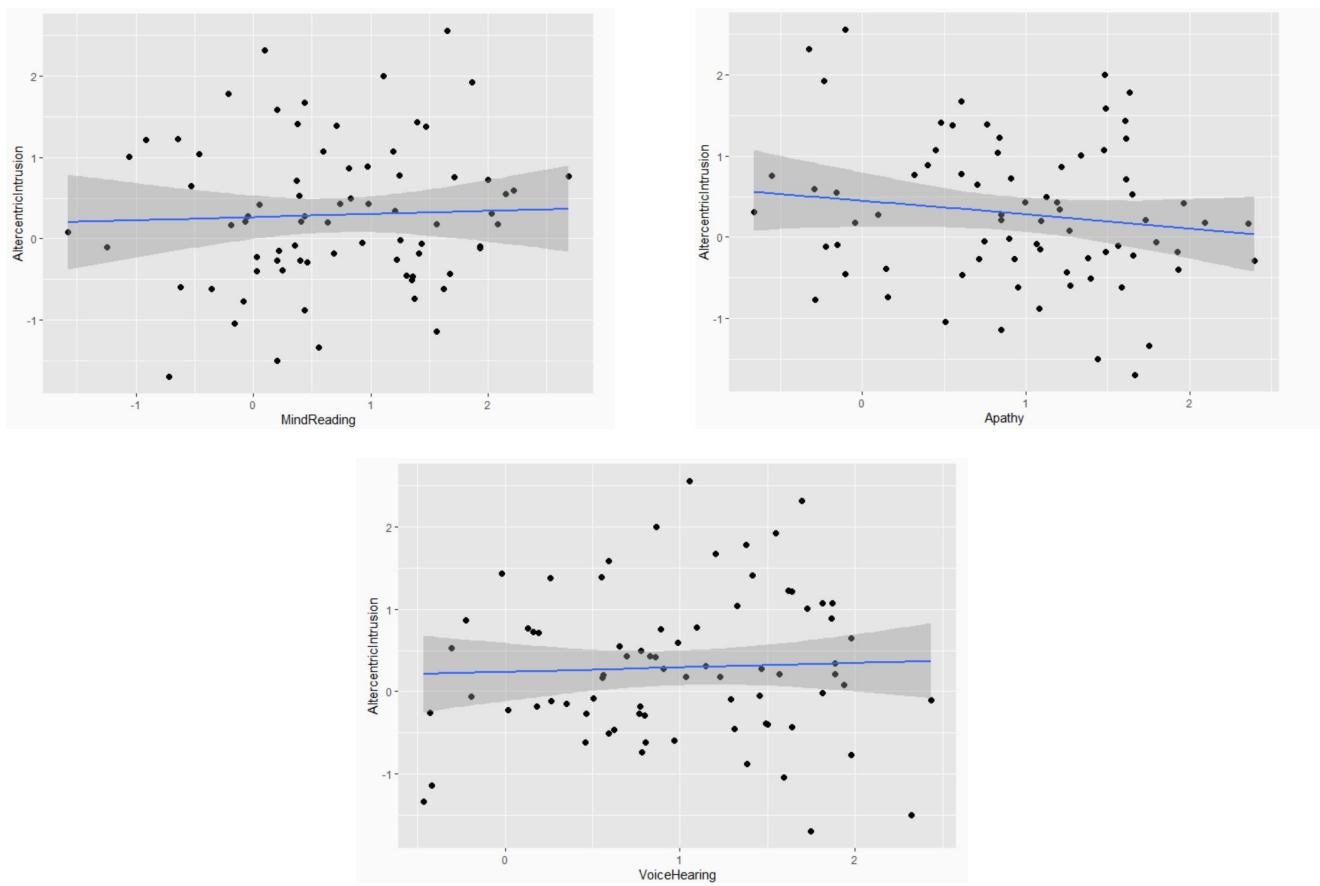


Figure 6 :Plotting a regression line against the data

We do a model comparison of our best models- which we argue to be model 6 and model 3 based on the significance of effects- in order to check, which model performs better at predicting altercentric intrusion. So as to achieve this we utilize a leave-one-out cross-validation. The results are shown in table 1 below. The model comparison indicates that predicting AI from apathy is the optimal course of action, if we want to minimize estimated out of sample error (stocking weight of 1). We see that adding other symptoms to the apathy model does not improve the generalizability of the model (stocking weight of 0).

Model	elpd difference	SE difference	Weight
M4, Apathy	0	0	1
M3, Mind Reading	-1.1	1.3	0
M2, Voice Hearing	-1.3	1.4	0
M6, VH + MR + A	-1.6	0.4	0
M5, VH + MR	-1.6	1.2	0

Table 1: Leave-one-out cross validation (schizophrenics only)

The immediate interpretation of these results is that in patients with schizophrenia, apathy seems to predict altercentric intrusion the best out of the three symptoms analyzed.

Discussion

According to our analysis and leave-one-out comparison, apathy is clearly the best predictor of altercentric intrusion out of the three symptoms, and the model does not improve significantly by adding the two other symptoms. The results of the given models thus go against our theoretical knowledge and, therefore, our hypotheses that mind reading and voice hearing are related to altercentric intrusion. Instead, we see that apathy is a more credible predictor, which is against our hypothesis. In the light of this evidence, we need to rethink our theoretical assumptions.

However, instead of readily rejecting our hypothesis, which is grounded in theory, we might attempt to analyze the data more thoroughly first, in order to exclude the possibility of statistical biases and confounds. In the next section we attempt to unravel biases in the analysis.

PART TWO

2.1. Investigating Model Biases

In order to investigate the data and the implicit inferences we make, when designing the analysis, we illustrate our best conjecture for the correlations between the symptoms and the diagnosis of schizophrenia in a directed acyclic graph (DAG) (see figure 7).

According to the DAG, which is congruent with our hypothesis, only voice hearing and mind reading are causally associated with altercentric intrusion, while all three symptoms together elicit a diagnosis.

In the small world the DAG presents, where the schizophrenia diagnosis is given by an additive combination of these three symptoms, people with the diagnosis are likely to either have medium to high values in voice hearing and mind reading, or high values in apathy. If apathy and voice hearing are not significantly correlated, it is improbable to score high in both categories, but probable to show one of two profiles. You could even count three profiles, if you wanted to dissociate VH and MR, but it is not necessary to explain the significant negative correlation between apathy and altercentric intrusion.

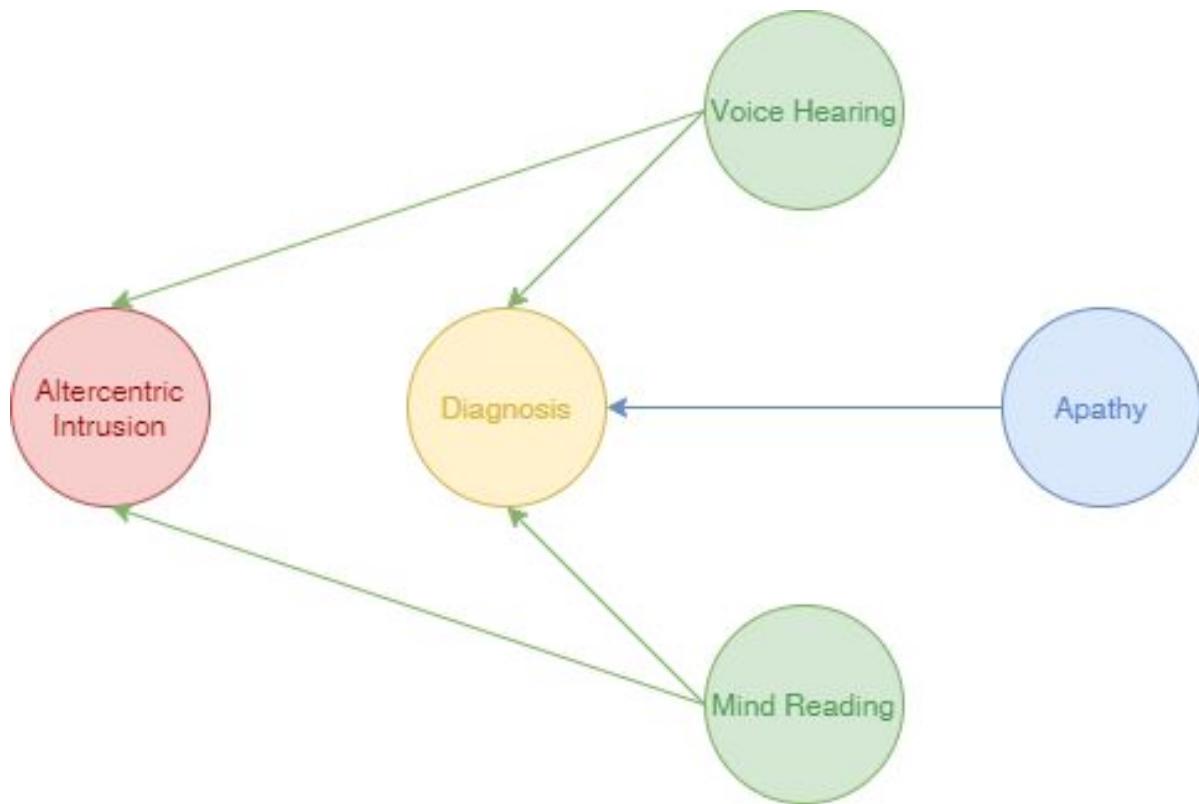


Figure 7: DAG illustrating our hypothesized causal relationships

Therefore the reason for apathy being a better predictor of altercentric intrusion than mind reading or voice hearing, despite the fact that altercentric intrusion is not causally related to apathy, is that the analysis deals with a diagnosed group of people, which are thus selected non-randomly from the whole dataset. If we examine how the diagnosis is administered, by looking at the data simulation, we see that people are assigned the diagnosis by being in the top 25 % according to the sum severity of three symptoms: apathy, mind reading, and voice hearing.

In a scenario where the three symptoms are independent (or, perhaps, weakly correlated), and where we only look at people with schizophrenia, apathy actually *should* be highly negatively correlated with altercentric intrusion for the very reason that they are *not* causally related in the DAG. Individuals who only experience weak voice hearing and mind reading, but are still deemed Schizophrenic, according to our simulator, *must* have relatively high scores in apathy. There are only those three symptoms. Since we only analyze diagnosed individuals, many of these individuals with high apathy scores will in contrast have weak scores for AI, thus creating the spurious correlation.

In our case the predictor variables are not independent, which makes inferring about biases more complicated. All predictors are correlated with each other, but equally and only weakly. Therefore we believe the bias is still present, and we can thus expect to see a negative correlation between apathy and altercentric intrusion within the Schizophrenia diagnostic group. We believe this bias originates from diagnosis being an addition of

multiple variables. This means that uncorrelated or positively correlated variables like voice hearing and apathy become negatively correlated within a patient group, thus making anything dependent on e.g. voice hearing, spuriously correlated with apathy as well.

If we are interested in examining the symptomatic correlations to altercentric intrusion, we suggest using both parts of the dataset: diagnosed and undiagnosed. Another approach could be to include more symptoms, so that the diagnosis isn't only caused by either the altercentric intrusion complex or apathy.

2.2. Rethinking the Analysis

When redesigning our analysis, we want to use some of the ideas and inferences from the section above. Since we have argued that selecting only schizophrenic patients creates a spurious correlation between apathy and altercentric intrusion, it seems intuitive to retry some of the original models using the whole dataset. If we do not see the same negative relationship between apathy and altercentric intrusion in the whole data set as in diagnosed individuals, we can be pretty certain that our results are influenced by some sort of bias.

The results of the models 2-6 as described in pseudocode under 1.2., when rerun on the entire data set of schizophrenic patients and controls alike, are shown in table 2 below. In order to attain these results, we run uni and multivariate regressions, reusing the priors from 1.2., with altercentric intrusion as the dependent variable. Predictor variables are shown in *Table 2*.

	b	CIs	ER
Model 2			
Voice Hearing	0.19	0.08, 0.30	Inf
Model 3			
Mind Reading	0.19	0.07, 0.30	1999
Model 4			
Apathy	0.08	-0.03, 0.20	12.75
Model 5			
Voice Hearing	0.16	0.06, 0.27	570.43
Mind Reading	0.16	0.06, 0.27	799
Model 6			

Voice Hearing	0.16	0.05, 0.27	306. 69
Mind Reading	0.16	0.05, 0.27	799
Apathy	0.01	-0.09, 0.12	1.57

Table 2: output of six separate bayesian analyses

Just by surveying the outcome of the analyses, it is evident that the large effect of apathy on altercentric intrusion has virtually disappeared, when we add controls to the analysis. On the other hand the credibility and evidence ratio as well as the estimates of voice hearing and mind reading being related to AI has increased.

In order to conclude which model performs the best, when predicting altercentric intrusion in individuals, we compare models using the LOO method, *see Table 3 below*.

Model	elpd difference	SE difference	Weight
M5, VH + MR	0	0	0.806
M6, VH + MR + A	-1	0.3	0
M2, Voice Hearing	-3.5	2.9	0.080
M3, Mind Reading	-3.6	3	0.114
M4, Apathy	-8.2	4.3	0

Table 3: Leave-one-out cross validation (all data)

The results of the analysis are indicated in *Table 3* above. According to our LOO cross-validation analysis, the model predicting altercentric intrusion from the combined symptom severity of mind reading and voice hearing is superior in performance with a stocking weight of 0.806. Model 5 (VH + HR) is thus the model which best minimizes error in predicting AI. See figures 11 and 12 for visualization of relation of the two predictors to altercentric intrusion according to model 5.

When comparing the univariate regressions (Model 2 and 3) to the multivariate regression (Model 5), the results seem to indicate a slight overlap in the variance that mind reading and voice hearing explains. However, the symptoms are still able to explain separate parts of the variance, lending credibility to our assumption in the DAG, that both symptoms influence AI separately.

The results from the rerun of the analysis on the entire dataset is in favour of our original hypothesis, namely that altercentric intrusion is related to a combination of factors, particularly mind reading and voice hearing. The new analysis paints an entirely different

picture of our dataset than our first attempt. It illustrates the importance of knowing your data, analysis, and assumptions, in order to weed out conceptual confounds.

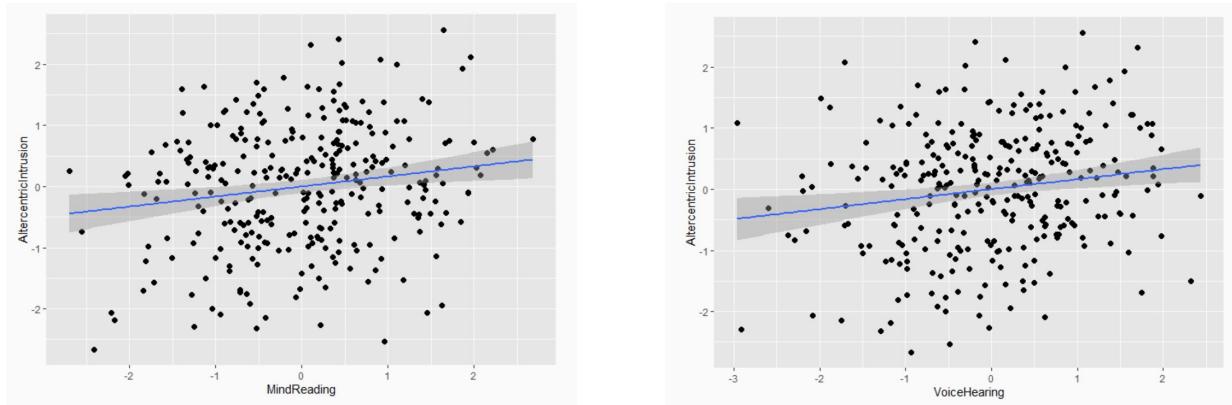


Figure 8: Plotting a linear regression line against the data, Model 5

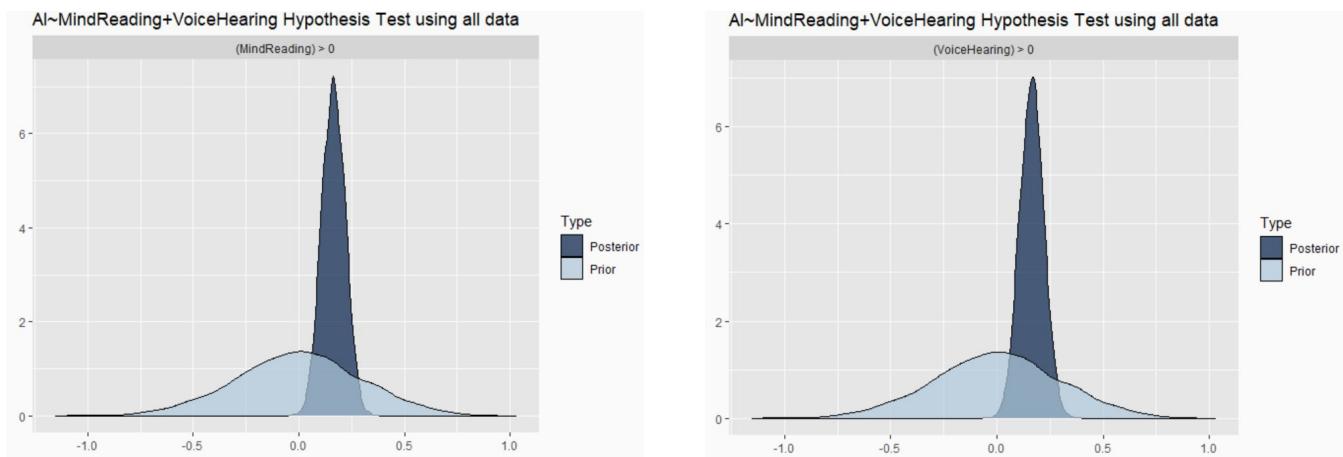


Figure 9: Visualization of our preferred model, Model 5

PART THREE

3.1 Simulation of data

When exploring how the data was generated, we can see that diagnosis is defined as a sum of all three symptoms with a threshold of 75 %. In other words, only the 25 % of all “people” with the highest sum of symptoms are classified as schizophrenic. Moreover, a weak correlation of 0.2 was set as symptoms usually co-occur with one another. In the simulation, altercentric intrusion was created to be 20 % correlated with both mind reading and voice hearing. This enables us to test whether the inclusion or exclusion of controls in our data is the cause of spurious correlations.

When we compare the correlations between variables of schizophrenia, we find that the inclusion of the control group leads to relations between variables similar to those defined in the variance-covariance matrix for generating the simulated data (see Figure 12).

On the contrary, the exclusion of the control group logically creates negative relations between variables (see Figure 10) because if one variable is high enough and causes a participant to be classified as schizophrenic, then the other symptoms are more likely to be lower. These results are in line with the findings in part 1.2 (see Figures 14 and 16) and, also, lead to apathy being the best predictor of AI, which is against our theoretical assumptions. Therefore, it is a great lesson of how knowing and grounding our assumptions in theory is an important part of making causal inferences without introducing statistical confounds and biases.

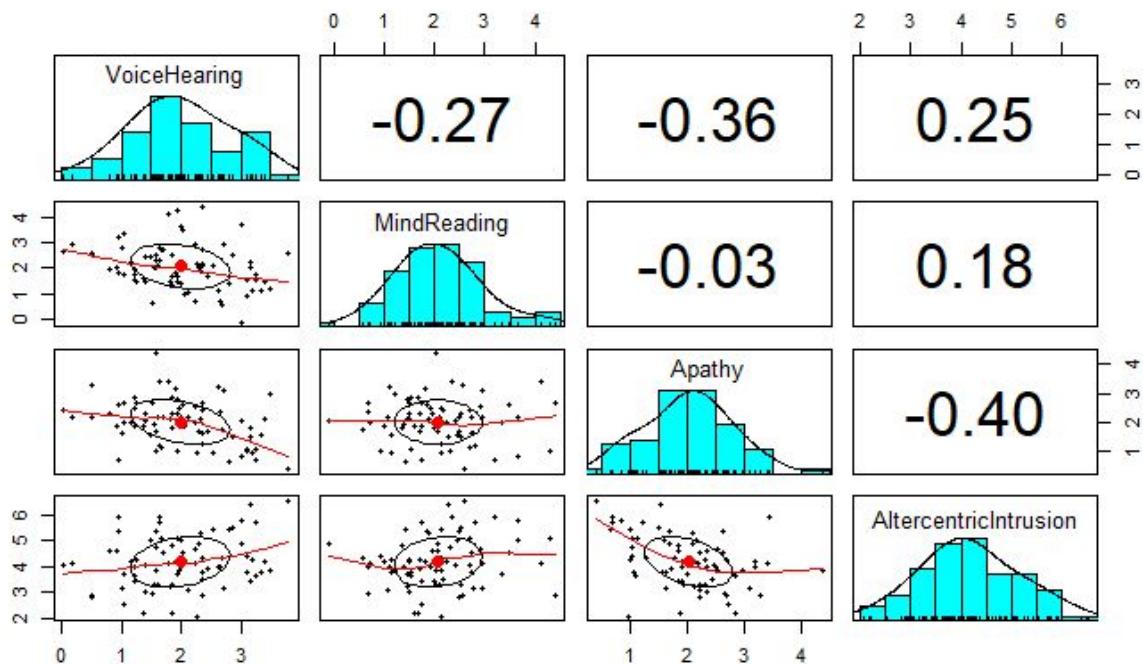


Figure 10: variance-covariance matrix of relations between variables in a simulated schizophrenia data

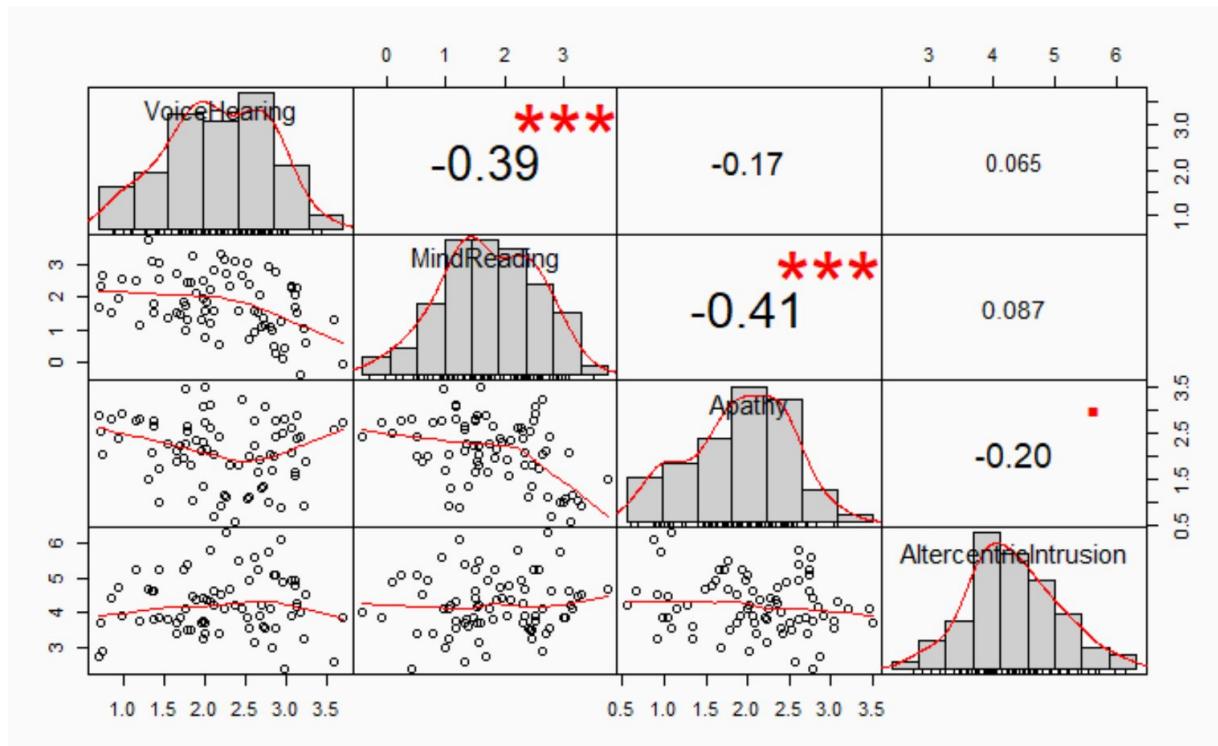


Figure 11: variance-covariance matrix of relations between variables in tested schizophrenia data

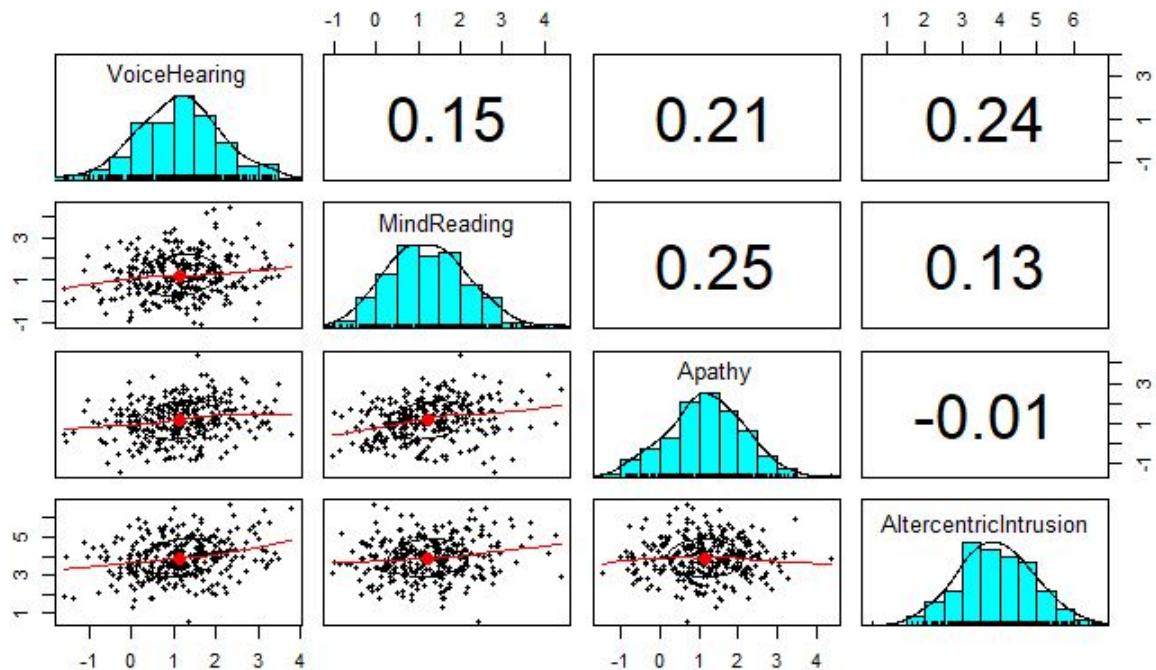


Figure 12: variance-covariance matrix of relations between variables in simulated controls data

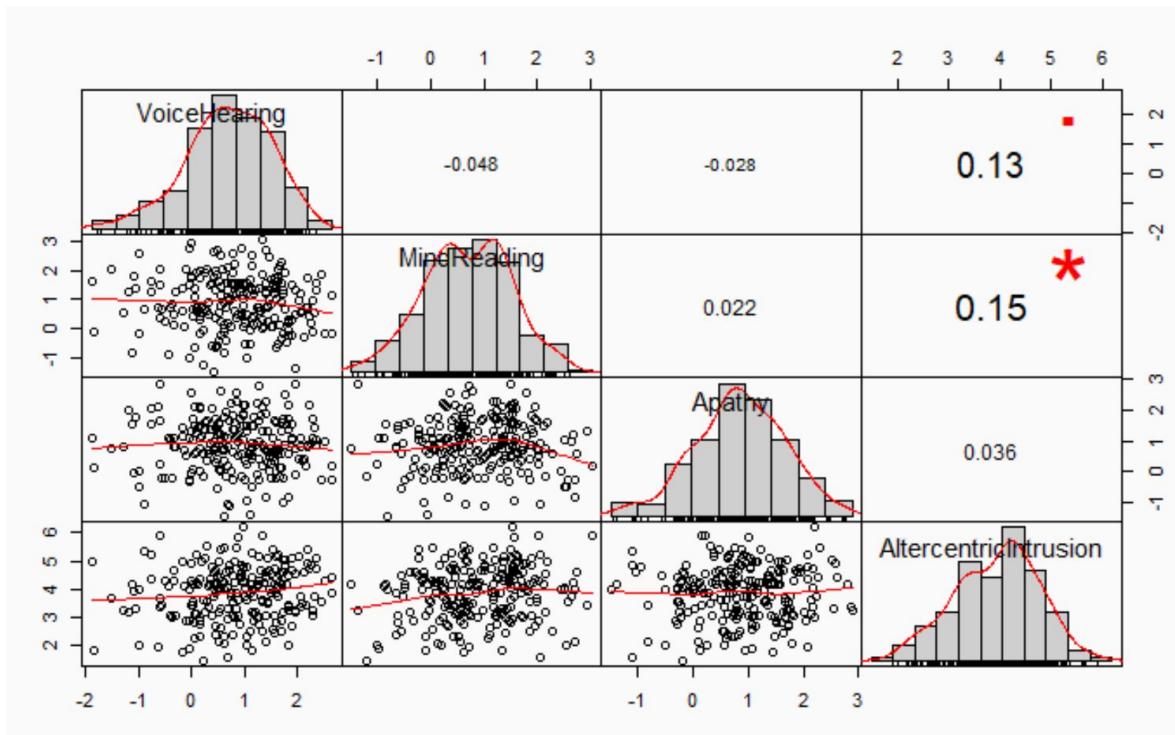


Figure 13: variance-covariance matrix of relations between variables in tested controls data

Appendix

Part One

AI~Diagnosis PriorCheck

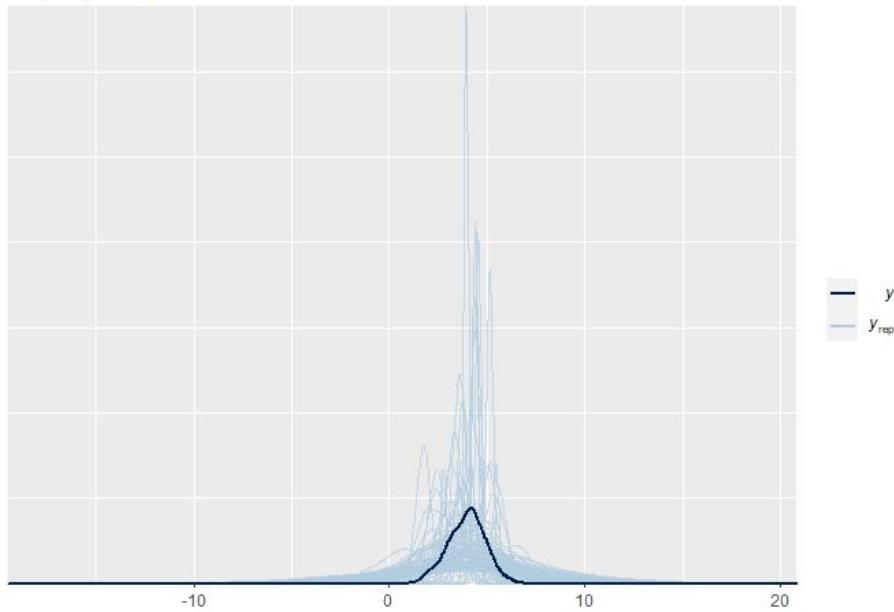


Figure 14: prior predictive check

AI~Diagnosis Posterior predictive check

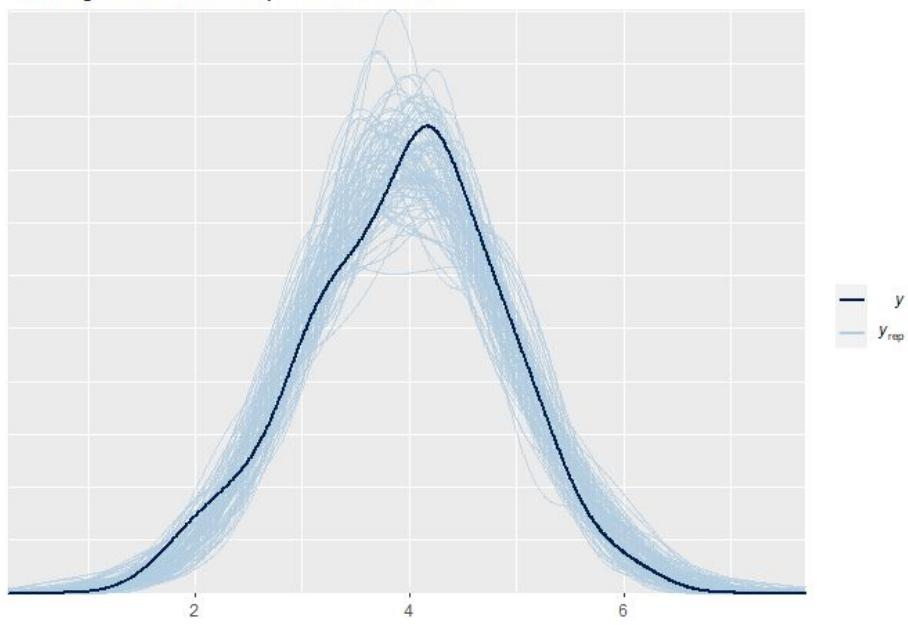


Figure 15: posterior predictive check

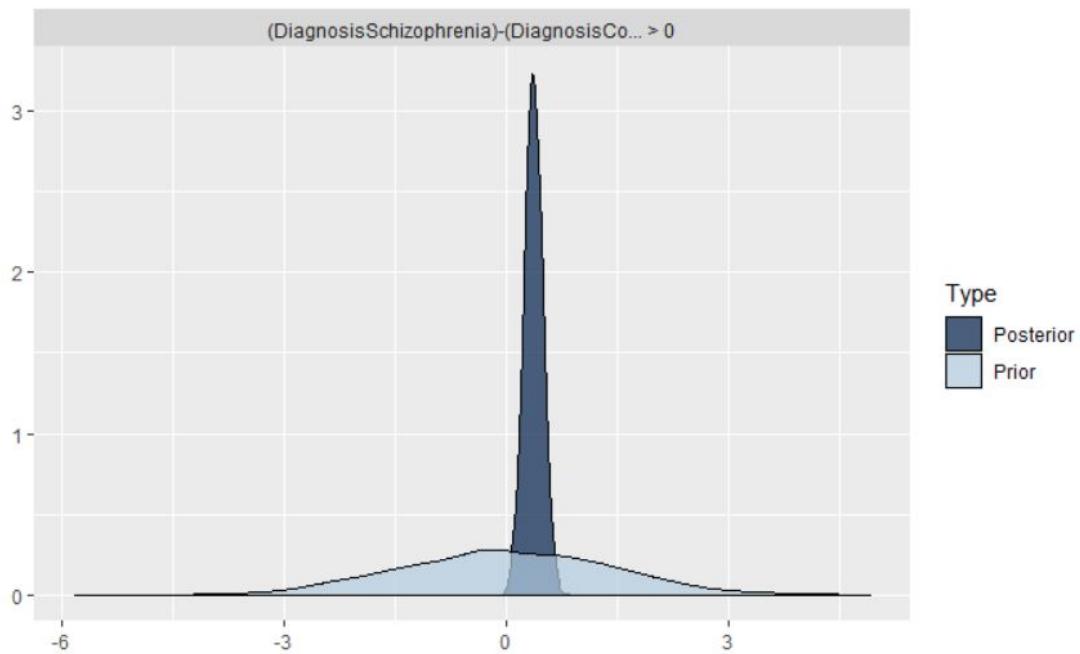
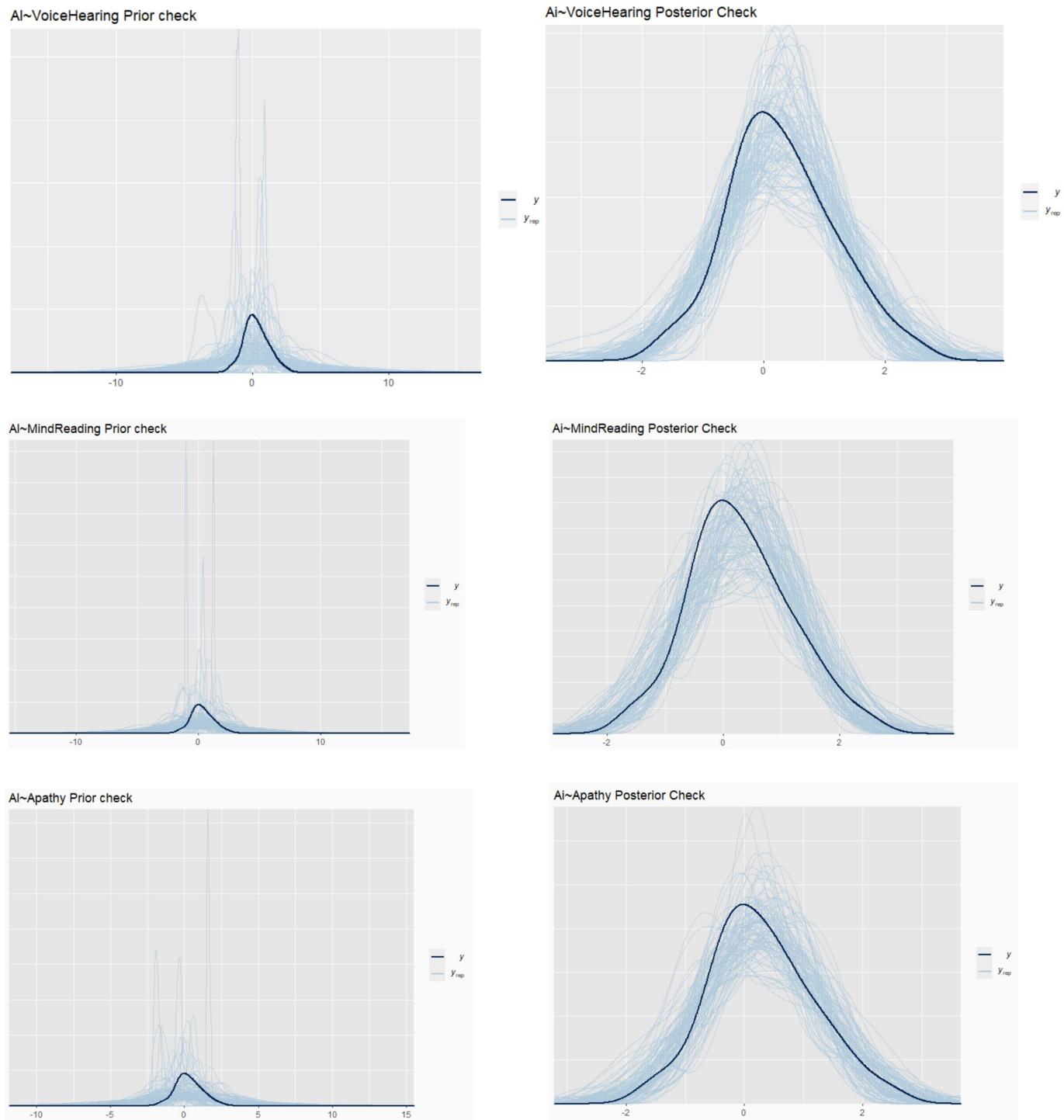


Figure 16: update check plot + hypothesis testing

*Figure 17: prior and posterior predictive checks*

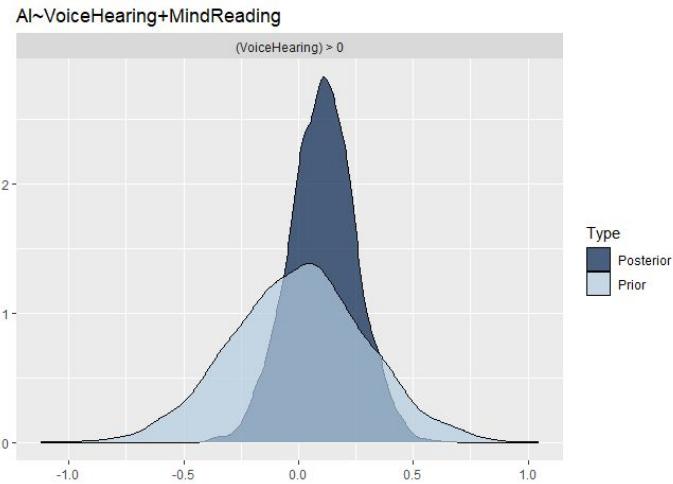


Figure 18

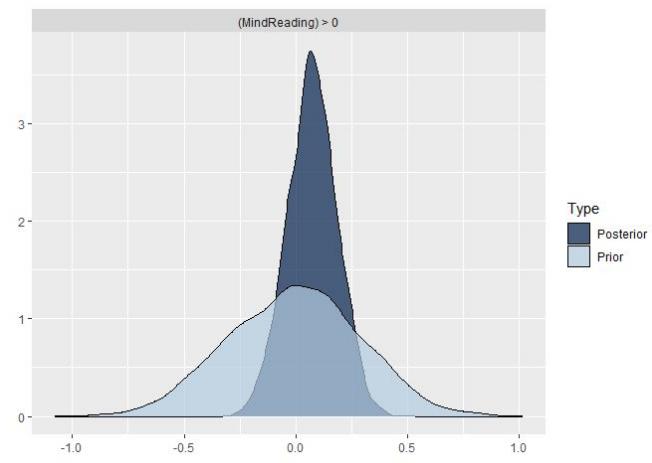


Figure 19

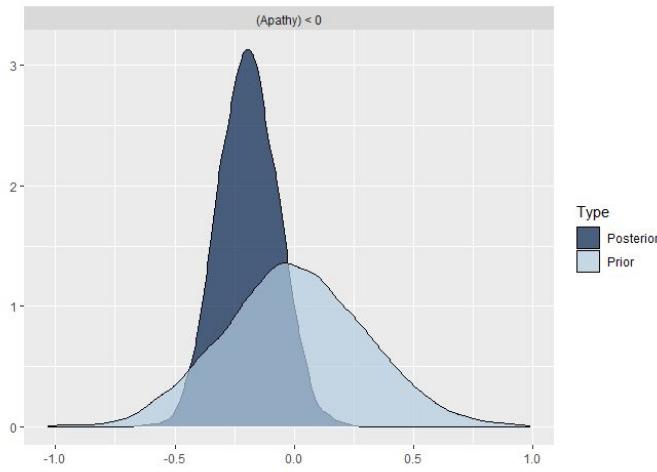


Figure 20

Figures 18-20: update check plots + hypothesis testing

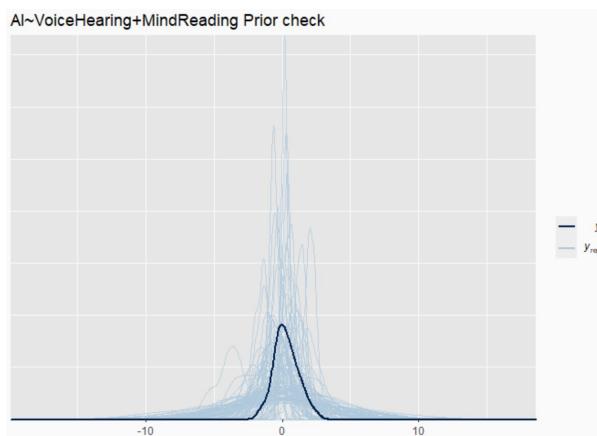


Figure 21 : prior predictive check

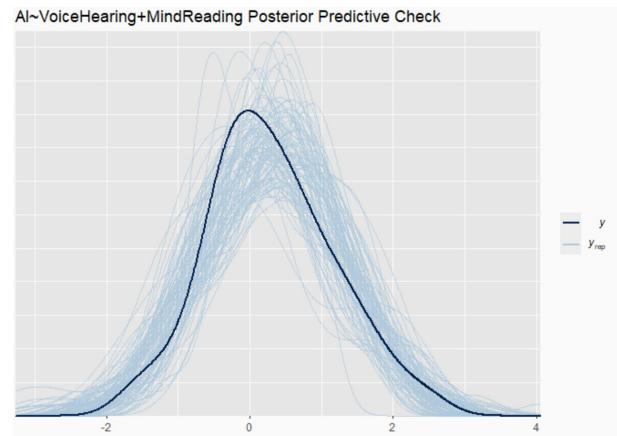


Figure 22: posterior predictive check

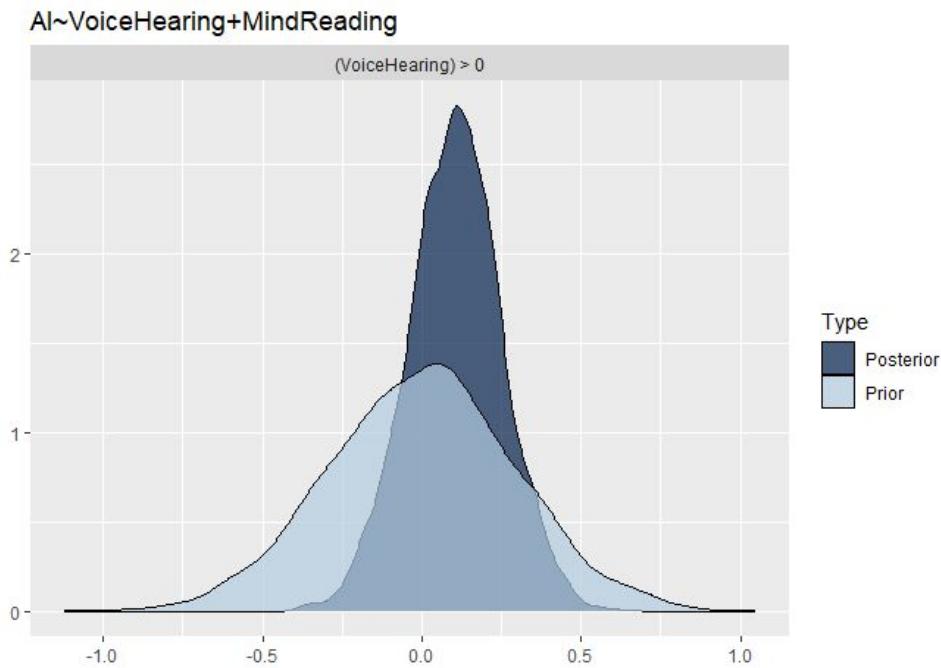


Figure 23 : update check plot + hypothesis testing

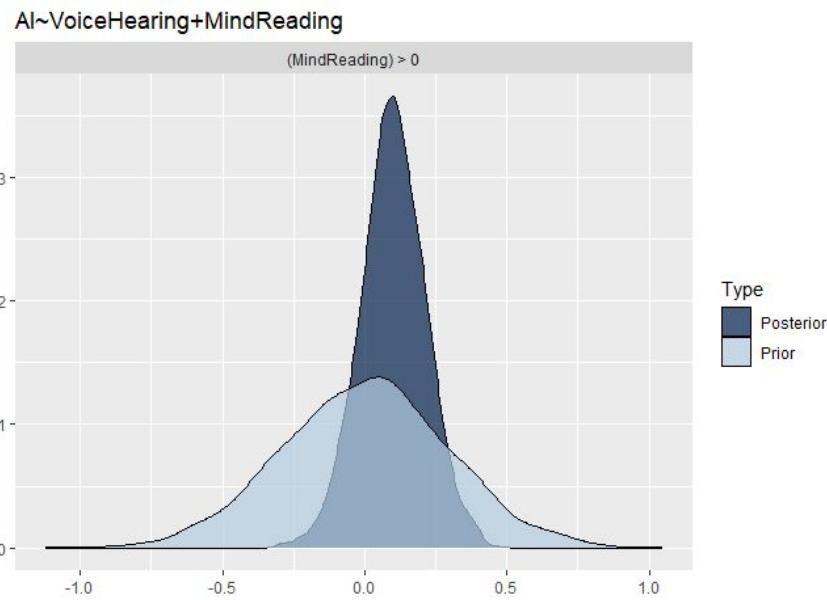


Figure 24 : update check plot + hypothesis testing

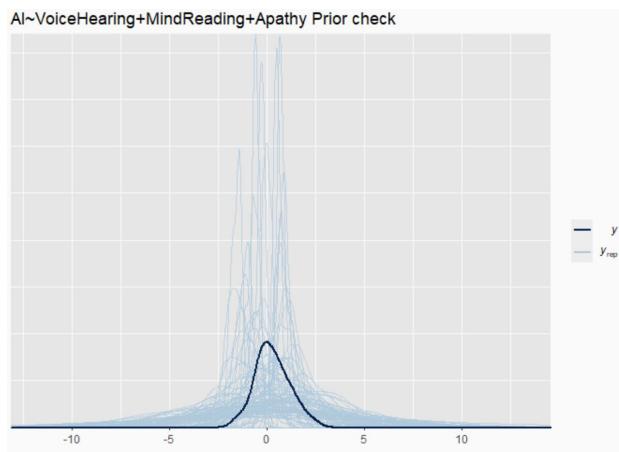


Figure 25 : prior predictive check

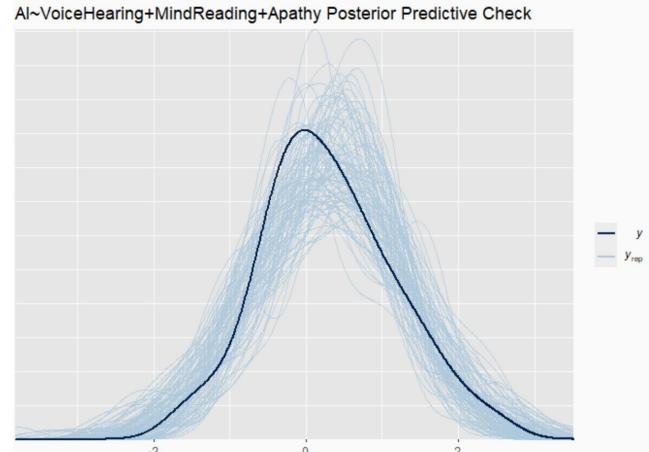


Figure 26 : Posterior predictive check

AI~VoiceHearing+MindReading+Apathy

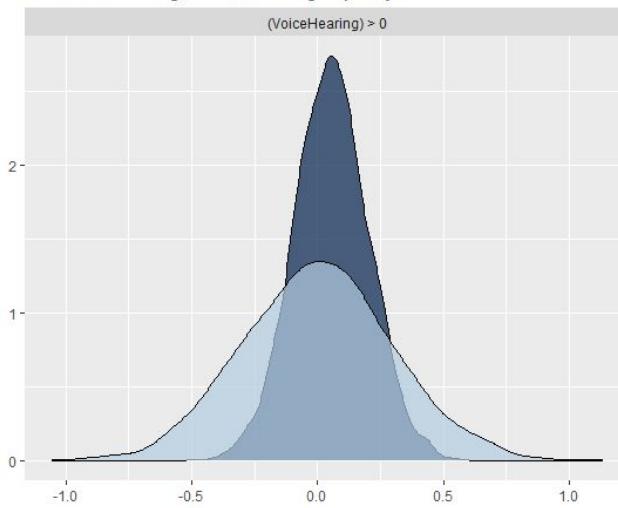


Figure 27

AI~VoiceHearing+MindReading+Apathy

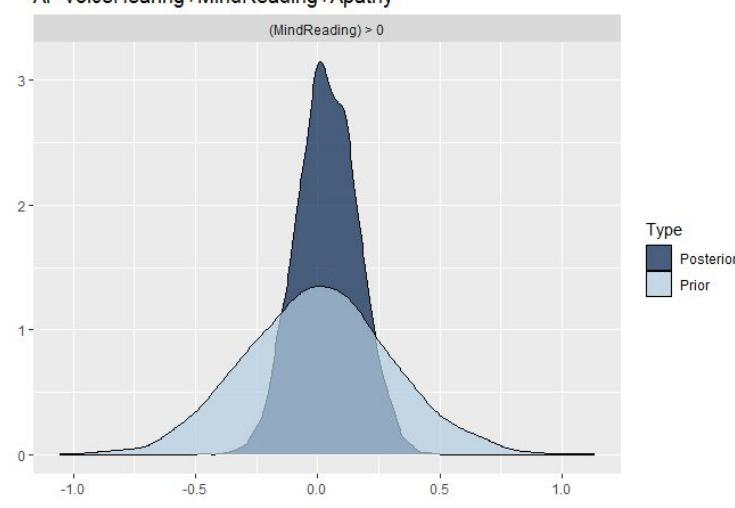


Figure 28

AI~VoiceHearing+MindReading+Apathy

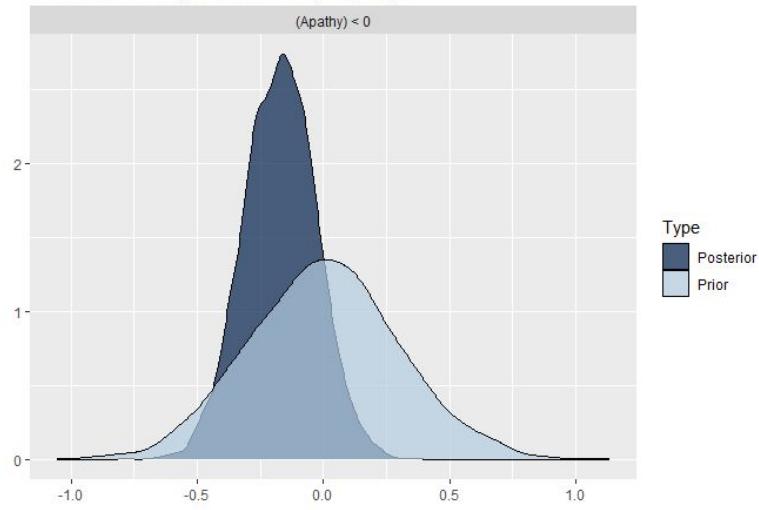


Figure 29

Figures 27-29 updating check plots + hypothesis testing

Assignment 4

Applying meta-analytic priors

Github link: <https://github.com/PeterThramkrongart/CompMod4/blob/master/Assignment4.md>

o

Contents

Introduction.....	1
Hypotheses.....	2
Modeling with conservative priors.....	2
Designing conservative priors	2
Modeling with meta-analytic priors.....	3
Meta-analysis	3
Designing meta-analytic priors.....	4
Comparative discussion of methods and results	5
Predictive checks.....	5
Results and estimates.....	5
Hypotheses testing.....	7
Comparing methodological approaches	8
Conclusion	9
Appendix.....	10

Introduction

In Bayesian modeling, choosing the right priors is, arguably, of the highest importance. There are multiple approaches to designing priors. Besides the most common one, a conservative prior choice, one can also employ meta-analytic priors to “stand on the shoulders of giants”. Although it aptly reflects the cumulative nature of science, it is of vast importance that one remains critical. To shine a light on this matter, we compare a conservative set of priors to meta-analytic ones in the analysis of pitch variability in autistic and control subjects.

Both in the fields of social and psychological studies, one recurring concern is the generalizability of a sample to the entire targeted population. A possible remedy to this problem is the use of

meta-analysis, as it compiles multiple heterogeneous samples into a single, more robust estimate. Therefore, it is more likely applicable to the entire population, including the sample of one's study. On the other hand, the quality of the meta-analysis is derived from the quality of the experiments included. Following the same rationale, compromised studies necessarily result in a compromised meta-analysis. However, there are other pitfalls along the way. When using meta-analytic priors, one must be aware of the possible cultural, national, or linguistic differences, as the effects might not be cross-culturally consistent. In theory, if the meta-analysis is rigorous, this approach is an adequate starting point for further studies, utilizing meta-analytic priors.

To see whether this theory applies to practice, we decided to put it to a test. We estimated pitch variability of autists in two studies – one American and one Danish. It is hypothesized that autistic people have a largely increased pitch and pitch variability, but we also want to test whether this difference was dependent on language. We employ Bayesian linear regression with varying effects and two priors. Either a conservative choice of our own or one based on the meta-analysis of 41 studies.

Hypotheses

The hypothesis we aim to evaluate in this paper is:

H1- Children with Autism Spectrum Disorder will have different pitch variability than their Typically Developing peers, however the effect may vary across languages.

Modeling with conservative priors

As we have data from two studies of different languages, we need to appropriately account for it. Therefore, we predict pitch variability from an interaction of diagnosis with language and varying effects for every participant. To know the relevance of the interaction, we further compared the interaction model with a baseline model. The formulas go as follows:

Baseline model: Pitch variability ~ 1 + Diagnosis + (1 | Participant ID))

Interaction model: Pitch variability ~ 0 + Language + Diagnosis:Language + (1 | Participant ID))

Designing conservative priors

In the process of designing the priors, we remained rather skeptical to prevent overfitting. As the data is standardized to Cohen's d, all the variables are automatically zero-centered. The distribution of pitch variability is roughly gaussian with some skew towards lower values. A log-transformation skews the data to a similar degree towards higher values. Therefore, we decided

to use Gaussian priors as transforming variables in Cohen's d form to log distribution is above our abilities.

Choosing identical priors of intercepts for both languages reflects our skepticism in the hypothesis, since we do not presuppose a language difference. Nonetheless, by choosing not too restricting priors, we allow the difference to manifest, if there is any. The prior for interaction estimates are zero-centered with a small standard deviation as, in general, the effects in psychology are modest. Furthermore, by setting the same prior for interaction estimates of both languages, we let the data 'talk'. Further prior concerns are the varying effects of participants, which is again zero-centered with a low standard deviation, as we do not expect our participants to vary dramatically. Lastly, a prior for the model's expected average error, sigma, is set to 0.5 as the standard deviation of the outcome variable, pitch variability, is 1 and we do not expect the model to do worse.

$$\begin{aligned}\beta_{Languagedk} &\sim normal(0,.1) \\ \beta_{Languageus} &\sim normal(0,.1) \\ \beta_{Languagedk:DiagnosisTD} &\sim normal(0,.1) \\ \beta_{Languageus:DiagnosisTD} &\sim normal(0,.1) \\ \alpha &\sim normal(0,.1), class = sd \\ \sigma &\sim normal(.5,.0), class = sigma\end{aligned}$$

Modeling with meta-analytic priors

Another approach for the analysis of our study's data is to employ meta-analytic priors. This is a sensible alternative to conservative priors since these priors are informed by past studies, which enables us to be more confident when, for instance, we are setting the standard deviations. The model thereby has more relevant information to start with and any effect detected will generally be more robust.

Meta-analysis

To generate the exact values for the meta-analytic priors, we analyzed 41 studies concerning pitch variability in ASD and controls. This analysis was carried out using the formula below, which estimates the effect size of the autistic diagnosis on pitch variability along with its error term which concerns the measurement uncertainty. Also, we account for population by modeling random intercepts as the same population has been used multiple times.

$$EffectSize | se(StandardError) \sim 1 + (1 | Population)$$

By running this formula using multivariate Bayesian linear modeling and normally distributed priors for the intercept (0,1) and standard deviation (0,3), we get the output, described in Table

1. The mean of the effect size together with the population are used to create informed meta-analytic priors in the following section.

Variable	Estimate	Est. Error	Lower 95% CI	Upper 95% CI
Effect size	-0.43	0.09	-0.60	-0.25
Population variance	0.31	0.09	0.12	0.50

Table 1 - Output of the meta-analysis model

Designing meta-analytic priors

When rerunning the interaction model previously defined in the section regarding conservative priors, we now utilize the information from the meta-analysis to direct our model.

In selecting priors for the interaction model, the first two priors indicate the baselines for the languages: one for Danish, one for English. The priors for interaction effects were based on the effect size of the meta-analysis, while the average expected prediction error of the model was given by variation of populations in the meta-analysis. We, again, decide to utilize Gaussian distributed priors. The formula for the priors is given:

$$\begin{aligned}\beta_{Languagedk} &\sim normal (.2, .3) \\ \beta_{Languageus} &\sim normal (.2, .3) \\ \beta_{Languagedk:DiagnosisTD} &\sim normal(-0.43, .1) \\ \beta_{Languageus:DiagnosisTD} &\sim normal(-0.43, .1) \\ \alpha &\sim normal(0, 1), class = sd \\ \sigma &\sim normal(.32, .1), class = sigma\end{aligned}$$

¹

We repeated the procedure from the conservative priors using the meta-analytic priors and built the same two models as defined previously with pitch variability as the dependent variable and participant ID as a random effect. Furthermore, the baseline model uses diagnosis as a predictor, while the complex model models each language with separate intercepts and slopes dependent on diagnosis.

Baseline model: Pitch variability ~ 1 + Diagnosis + (1 | Participant ID))

¹ We have unfortunately introduced a small human error in our code by defining the sigma prior value as 0.32 instead of 0.31. Considering the rather small size of the error of our typo and the extended running time of the code as well as the amount of spillover effects it would have on our other values, we have decided to continue using sigma = 0.32, .1.

Interaction model: Pitch variability ~ 0 + Language + Diagnosis:Language + (1 | Participant ID))

The results from running the multivariate regression model on the study data using the meta-analytic priors stated above are described in the section below. These results are compared to those obtained from the analysis using conservative priors. We wish to elucidate the discrepancies between the two approaches to setting priors and discuss the applicability as well as the pitfalls of both methods.

Comparative discussion of methods and results

Predictive checks

Both model priors seem reasonable and within range when prior predictive checks are performed. As expected, they do not quite catch the skew of the data and have a center peak around zero. (see Figure 3 and 4 in Appendix). Similarly, both models describe the data quite well when posterior predictive checks are performed (see Figure 3 and 4 in Appendix). Again, the models don't perfectly catch the skew of the data.

Results and estimates

According to the model employing the conservative priors, the mean pitch variability in ASD is much higher in American-English ($b=0.43$, CI's=0.31, 0.55) compared to Danish ($b= -0.14$, CI's= -0.26, -0.01) individuals. Looking at the estimates of the model with informed priors, the mean pitch variability in ASDs is, again, much higher in American English ($b= 0.77$, CI's= 0.62, 0.93) compared to Danish language ($b=-0.05$, CI's= -0.20, 0.10). Additionally, the employment of informed priors has led to an increase in the mean pitch value of ASDs in both languages compared to the skeptical prior estimates. The visual representation of the diagnosis effect on pitch variability across both languages is depicted in Figure 1 and Figure 2 and is discussed in more detail in the hypothesis testing section.

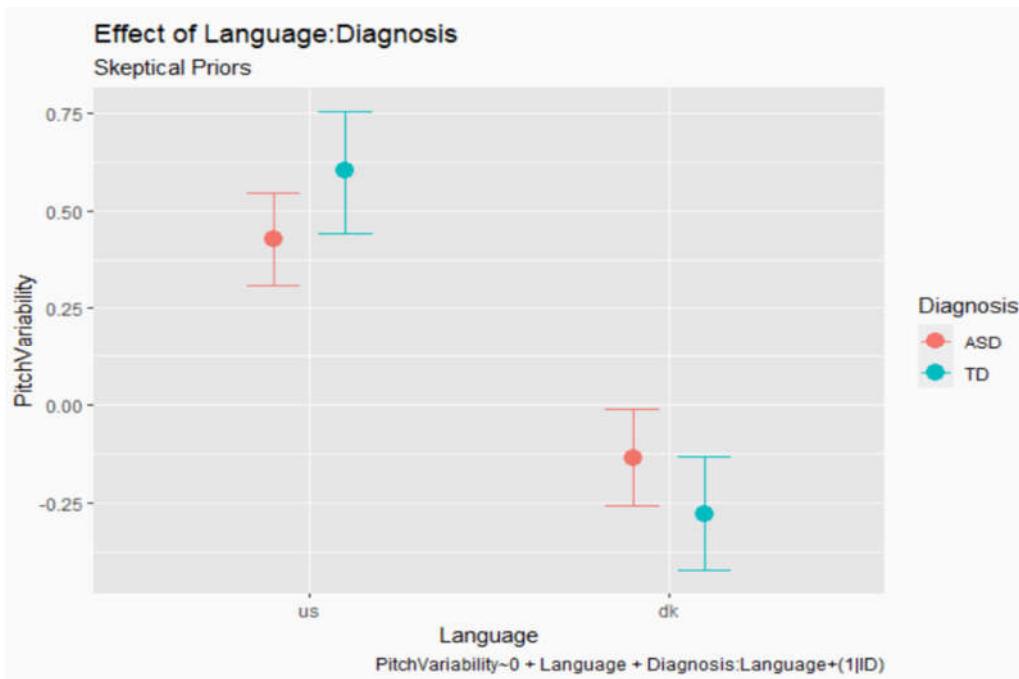


Figure 1 - Pitch variability in two languages by diagnosis using skeptical prior

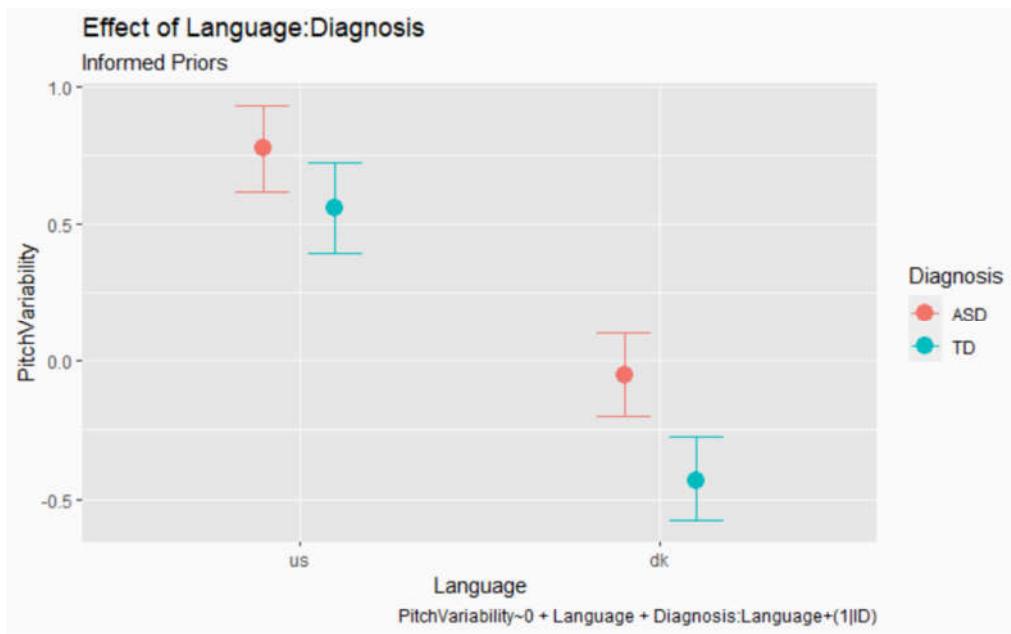


Figure 2 - Pitch variability in two languages by diagnosis using informed prior

	Estimate	Est. Error	Lower 95% CI	Upper 95% CI
Model with skeptical prior				
Languagedk	-0.14	0.06	-0.26	-0.01
Languageus	0.43	0.06	0.31	0.55
Languagedk:DiagnosisTD	-0.14	0.08	-0.29	0.01
Languageus:DiagnosisTD	0.17	0.08	0.03	0.33
Model with informed prior				
Languagedk	-0.05	0.08	-0.20	0.10
Languageus	0.77	0.08	0.62	0.93
Languagedk:DiagnosisTD	-0.38	0.08	-0.54	-0.23
Languageus:DiagnosisTD	-0.22	0.08	-0.37	-0.06

Table 2 - Results of the models with both skeptical and informed priors

Hypotheses testing

Our prediction that the mean pitch variability in individuals with Autism Spectrum Disorder will differ by language as well as in comparison to TD children has been confirmed. When modeling with skeptical priors, Danish neurotypical children display a slightly lower mean pitch variability than their ASD peers ($b=-0.14$, $CI's=-0.27$, -0.01 , $ER=29.77$). In contrast, American neurotypical children show a slightly higher mean pitch variability ($b=0.17$, $CI's=0.05$, 0.31 , $ER=0.01$). However, this effect disappears when we employ our informed prior, with both languages showing a lower mean pitch variability in TD patients. Danish neurotypical children exhibit a much more credible lower mean pitch variability ($b=-0.38$, $CI's=-0.51$, -0.25 , $ER=Inf$) from ASDs than their American peers ($b=-0.22$, $CI's=-0.35$, -0.08 , $ER=284.71$). However, it needs to be pointed out that informed priors for the language-diagnosis interaction was set with mean -0.43 and very small standard deviation of 0.1 . That inevitably leads to exclusively negative estimates. Therefore, testing whether the posterior samples of the interactions are smaller than zero is not informative as the informative prior does not allow this. Hence, instead of consulting evidence ratio (written in italics), which were included for completeness, in this case, one's attention should be focused at credible intervals instead.

When we test the hypothesis whether the pitch variability in ASDs in Danish language is lower than their peers in English American language we get it confirmed with both priors, where the difference is bigger and more credible using skeptical prior ($b=-0.32$, $CI's= -0.5$, -0.13 , $ER=399$) compared to informed prior ($b=-0.17$, $CI's=-0.36$, 0.02 , $ER=12.42$).

The models had no divergences, a Rhat of 1 and Effective sample sizes well above 200 for both Bulk and Tail.

	Estimate	Est. Error	Lower 95% CI	Upper 95% CI	Evid. Ratio
Model with skeptical prior					
Languagedk:DiagnosisTD < 0	-0.14	0.08	-0.27	-0.01	29.77
Languageus:DiagnosisTD < 0	0.17	0.08	0.05	0.31	0.01
Languagedk:DiagnosisTD < Languageus:DiagnosisTD	-0.32	0.11	-0.5	-0.13	399
Model with informed prior					
Languagedk:DiagnosisTD < 0	-0.38	0.08	-0.51	-0.25	<i>Inf</i>
Languageus:DiagnosisTD < 0	-0.22	0.08	-0.35	-0.08	284.71
Languagedk:DiagnosisTD < Languageus:DiagnosisTD	-0.17	0.11	-0.36	0.02	12.42

Table 3 - Hypotheses testing of the models with both skeptical and informed priors

Comparing methodological approaches

Model	Weight
With skeptical prior	0.413
With informed prior	0.587

Table 4 - Leave-one-out cross-validation of the models with both skeptical and informed priors

In a small world environment where our two models with different priors, skeptical and informed, respectively, are the only possible models, the leave-one-out cross-validation predicts that there is a 59% probability that the model with informed priors is the best model (see *Table 4*). However, we should not blindly assume that the informed priors offer the best, most robust results every time, because of the possible drawbacks associated with meta-analytic priors, which were mentioned previously in the introduction. Furthermore, in this specific example, the LOOCV weight is quite close to chance, which also indicates that we don't have sufficient evidence to be able to confidently choose one model over the other. On the other hand, as long as we are aware of the reliability of our sources, which are used for setting the informed prior, it is considered an applicable practice to aptly make use of the accumulation of scientific knowledge.

Using skeptical priors, the effect of diagnosis differs across languages. Observing the American-English subjects, we can detect lower pitch variability in ASD patients, whereas when looking at the Danish subjects, ASD patients appear to have higher pitch variability than neurotypical controls. However, when the meta-analytic (informed) priors are employed, the effect of diagnosis is consistent across languages. ASD subjects were found to have higher pitch variability regardless of their spoken language.

There are several possible interpretations of what we witness in our case. One explanation could be that the idiosyncrasies of the samples from the two studies, which were used to motivate our skeptical priors, might be larger than usual. When informed priors are used, this increased sample variance was corrected, and the “true” effect thus emerged. Another explanation could be that the effect of ASD diagnosis on pitch variability is, in fact, not consistent across languages. As the informed priors are based predominantly on the English-speaking population (26 studies out of 41 in the meta-analysis), the “true” effect in our analysis could be therefore falsely corrected, when relying on meta-analytic priors.

Given the specific context of our study, it is difficult to generalize the efficiency of using meta-analytic priors. However, there is still a lesson to be learned. Accounting for cultural and language differences is certainly as important of an aspect when employing meta-analytic priors as it is when we are using skeptical ones. Conceptually, the usage of meta-analytic priors is well-reasoned, and attempts of its implementation should continue, however, we need to be sure that we are standing on the shoulders of the right giant.

In terms of our research question, we cannot draw any resolute conclusions, since it's not obvious which model is best. If we want to investigate the matter further, we suggest redoing the meta-analysis, defining that the effect of diagnosis on pitch variability should be analyzed individually for each language in the meta-analysis. In the case of our study, when we defined the model with informed priors, we set the same mean and standard deviation for the effect of language on diagnosis. If we truly want to utilize the cumulative nature of science, it seems natural to take language into account from the start. Otherwise we might be restricting our outcome to an unnecessary degree.

Conclusion

To conclude, when evaluating which priors to use, conservative and meta-analytic priors can generate different results. Our hypothesis H1 is validated when we model our data with both skeptical and meta-analytic priors. The directions of the effects, however, change depending on the chosen priors. When using skeptical priors, the model predicts a decrease in pitch variability within American ASD children and increase pitch variability within Danish ASD children. Although, this effect changes when employing informed priors, as this model predicts an increase in pitch variability regardless of the language.

Evidently, it is important to substantiate the theoretical reasoning behind choice of prior, and when in doubt, it could be relevant to assess what effect the use of conservative versus informed priors contributes to the analysis. In theory, meta-analytic priors are preferred, since they prioritize robustness as well as capitalize on years of hard work from other researchers. In practice, however, they are susceptible to human error. Furthermore, they might impress a false sense of confidence if applied inconsiderately. On the other hand, the informed priors prevent us

from being too rash when encountering new evidence, which conflicts from the accepted paradigms. Furthermore, because of the immense amount of past research we can assume non-zero prior values for the mean of the effect size, enabling the model to gain more confidence than achieved by applying skeptical priors, given the new evidence is consistent with the meta-analytic effects. The reliance on other researcher's data using meta-analytic priors is thus a double-edged sword, which generates reliable results if wielded thoughtfully.

Appendix

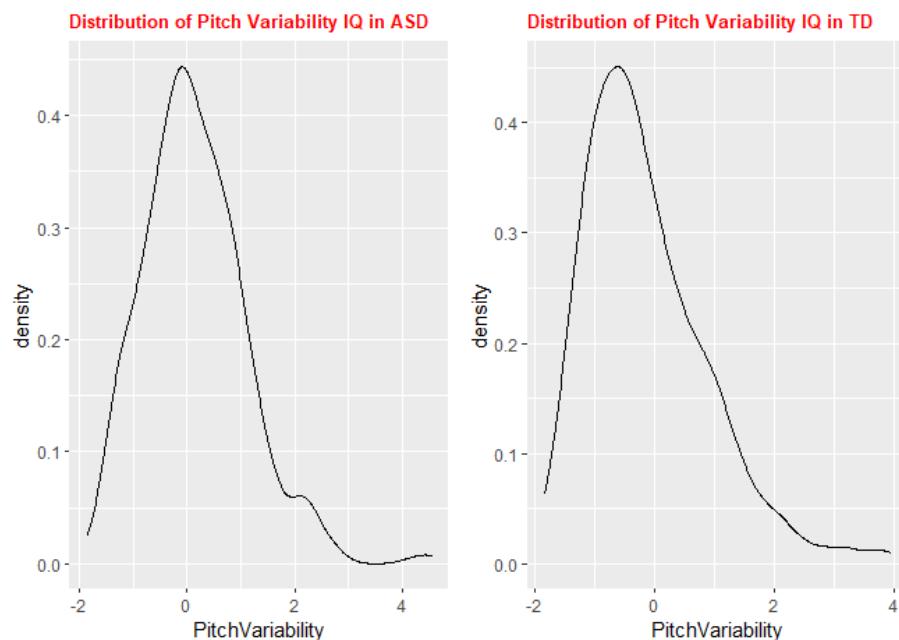
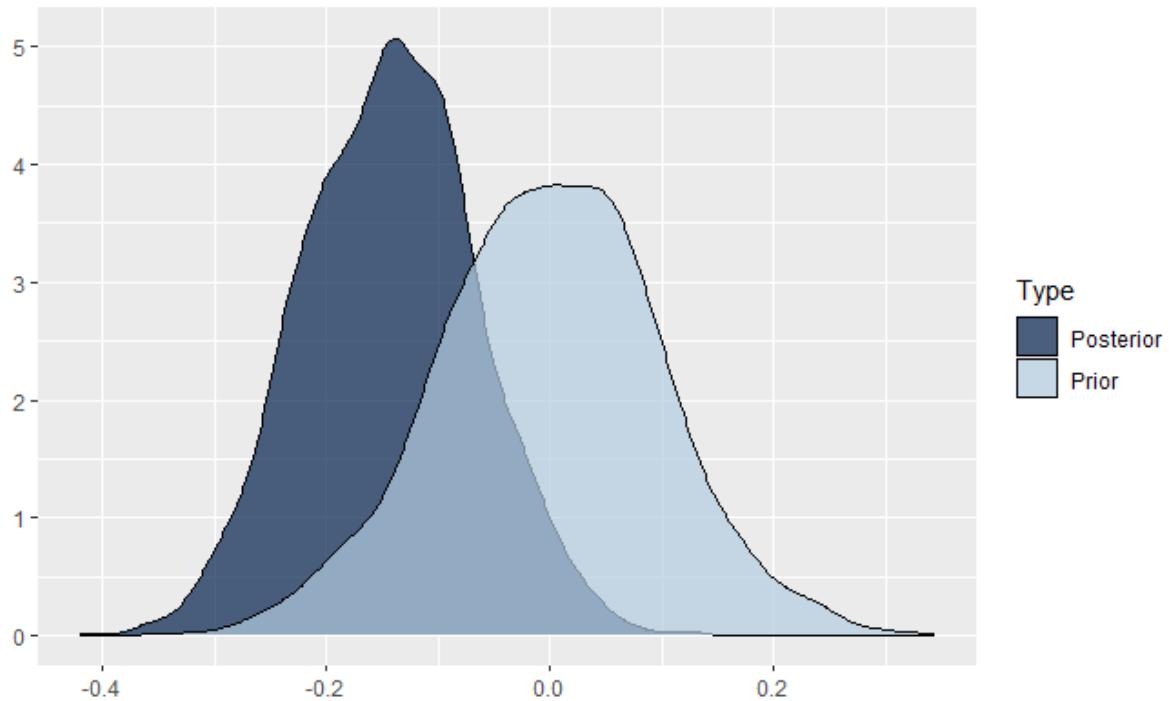


Figure 3 - Distribution of pitch variability in ASD and TD

Languagedk:DiagnosisTD < 0

Skeptical Priors



PitchVariability~0 + Language + Diagnosis:Language+(1|ID)

Figure 4 - Hypothesis testing with skeptical priors

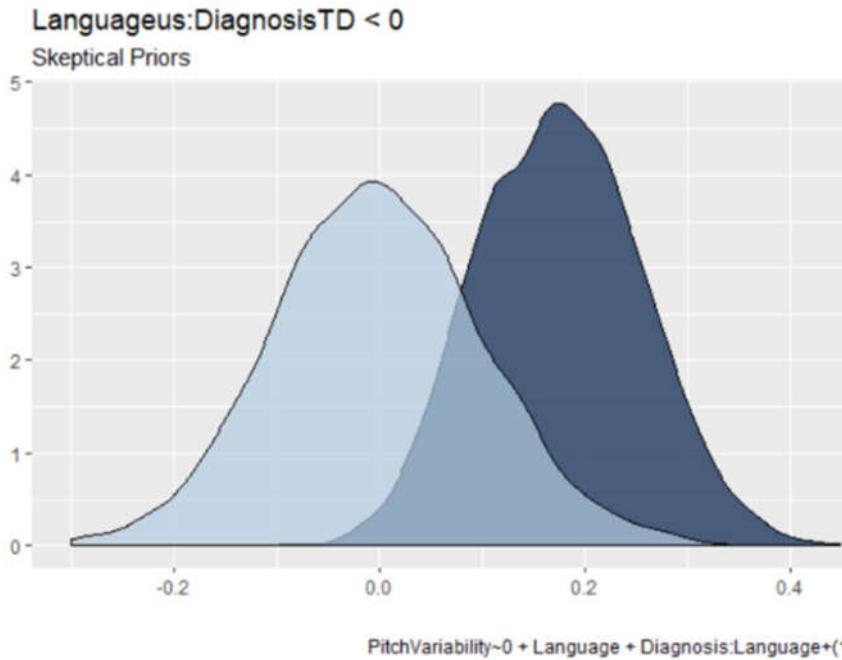


Figure 5- Hypothesis testing with skeptical priors

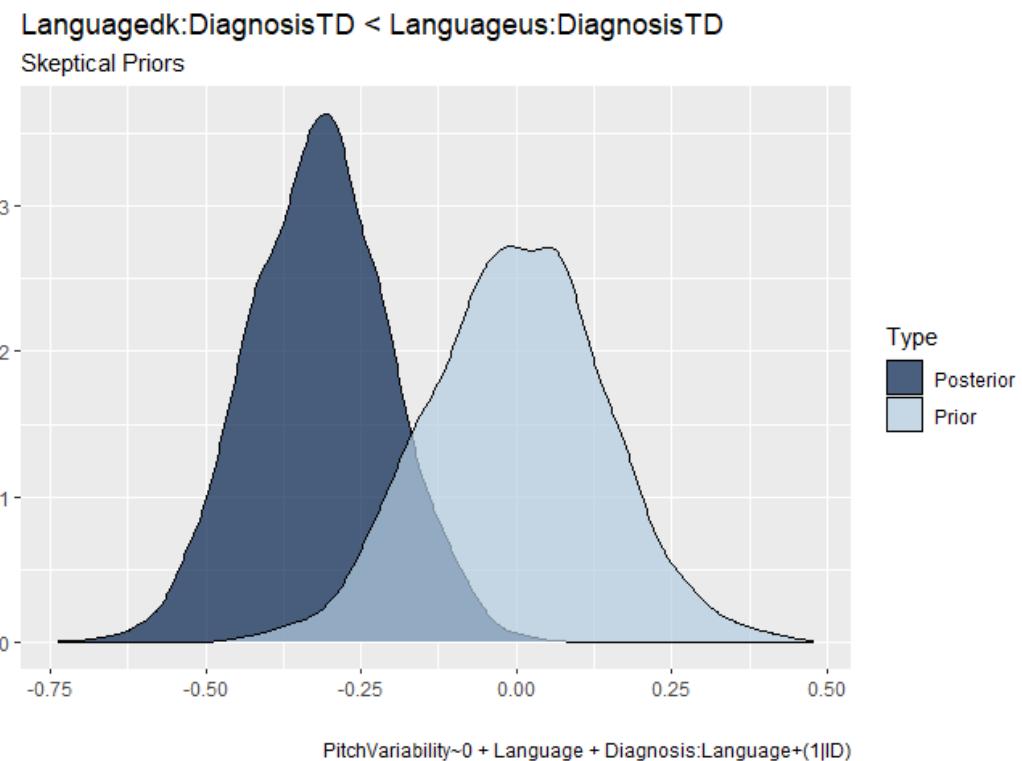


Figure 6 - Hypothesis testing with skeptical priors

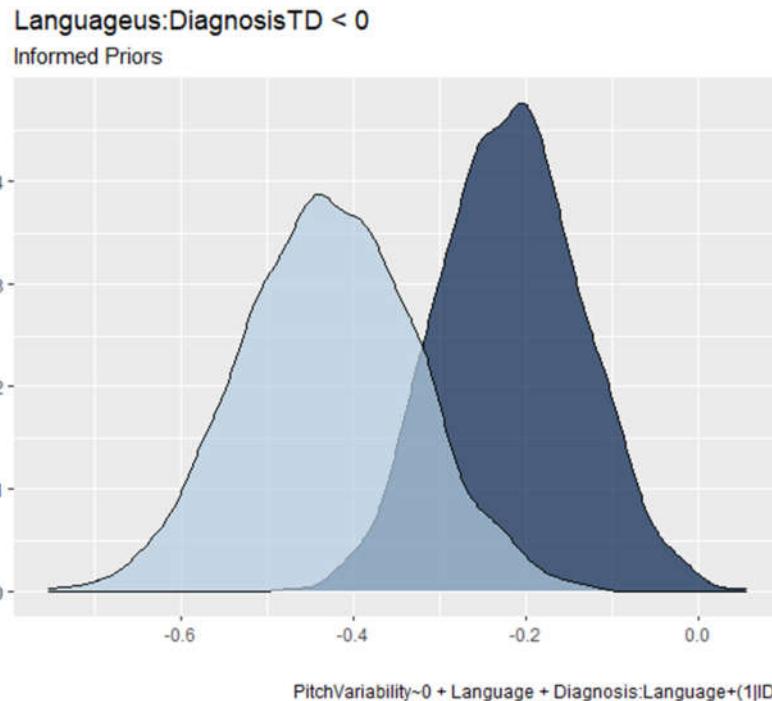


Figure 7 - Hypothesis testing with informed priors

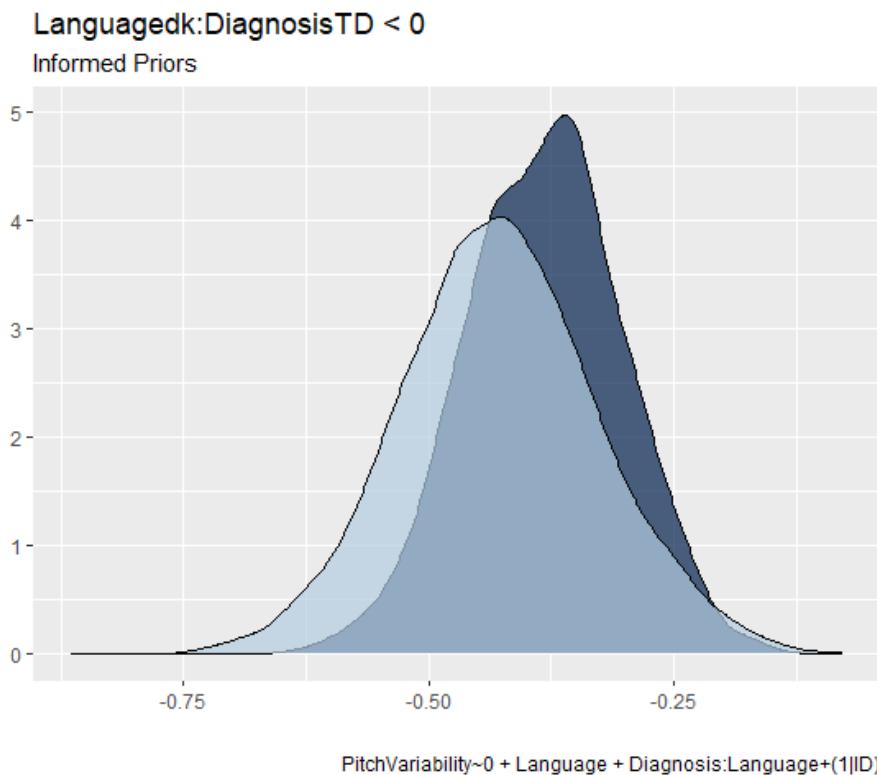


Figure 8 - Hypothesis testing with informed priors

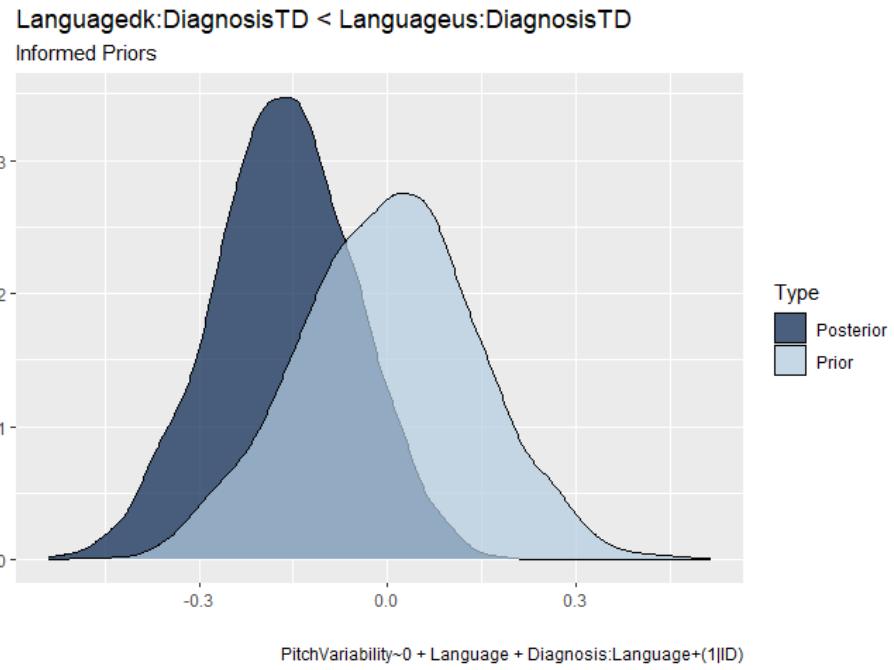


Figure 9 - Hypothesis testing with informed priors

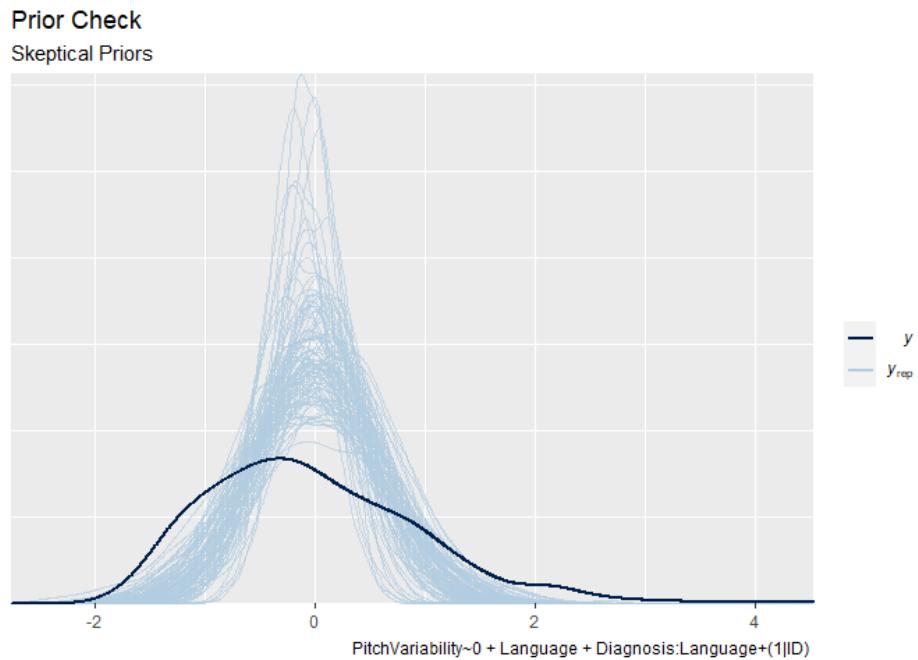


Figure 10 - Prior Predictive Check with skeptical priors

Posterior Check

Skeptical Priors

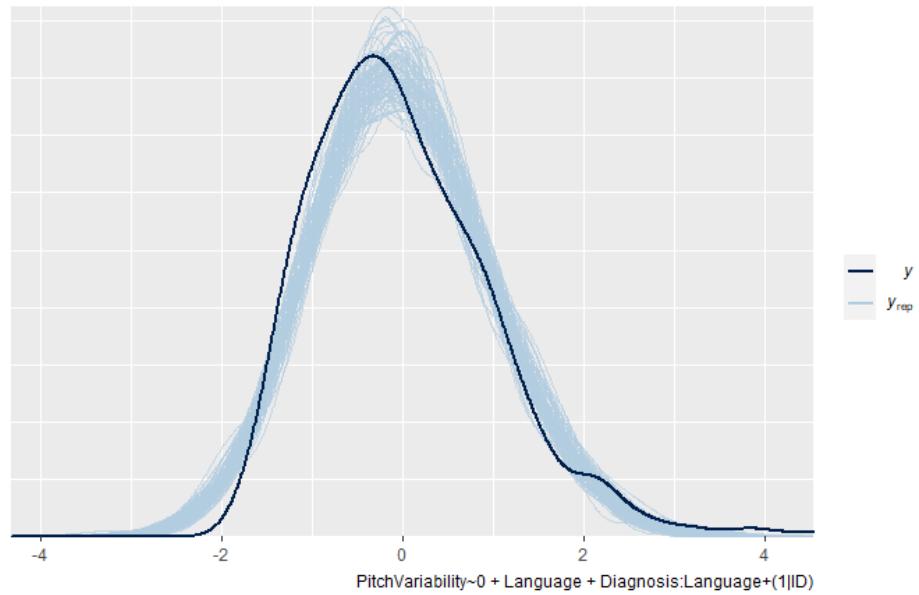


Figure 11 - Posterior Predictive Check with skeptical priors

Prior Check

Informed Priors

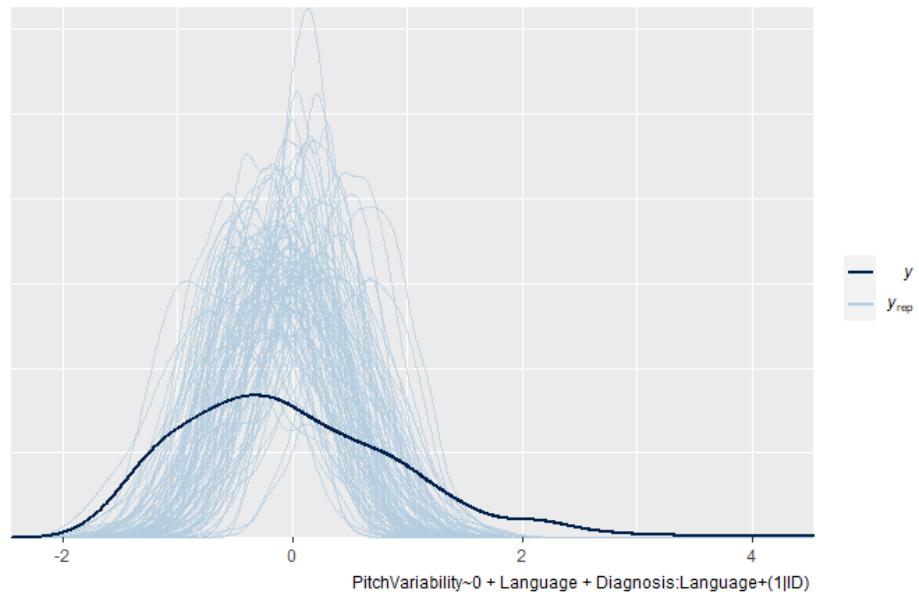


Figure 12 - Prior Predictive Check with informed priors

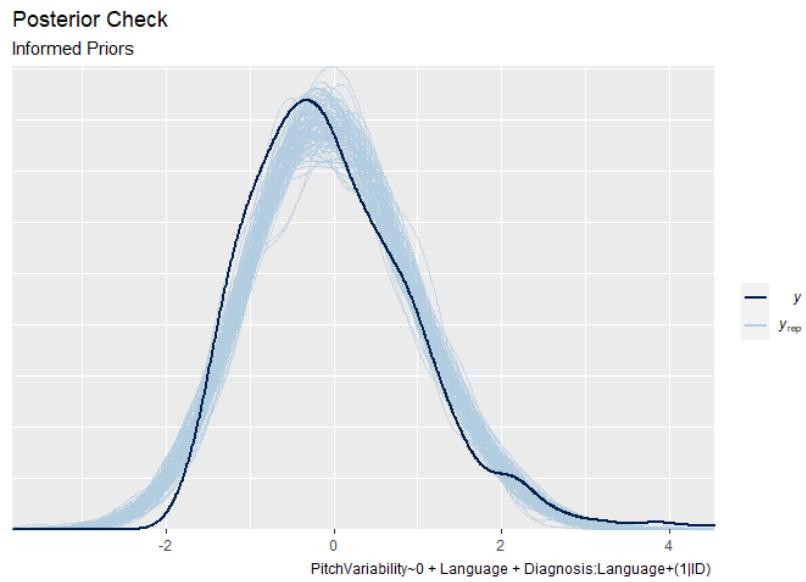


Figure 13 - Posterior Predictive Check with informed priors

Assignment 5

Investigating echo-chamber formation in large social networks using Agent-Based Modelling.

Contents

Contact Information.....	1
INTRODUCTION	2
METHODS.....	2
Model setup	2
Network Structure	3
Making the model go: Searching for connections	3
Sampling information from links and updating confidence	3
Updating beliefs and assessing belief purity	4
Conclusions of the previous study.....	4
Our model design	4
RESULTS.....	5
Replication of previous findings.....	5
Preventing echo chambers with censorship.....	9
Investigating the two-party formation	10
DISCUSSION.....	11
Our findings.....	11
Limitations.....	12
Appendix	13
References	12

Contact Information

Asger Lakkenborg Christiansen, student number: 201804579, email : 201804579@post.au.dk

Peter Andreas Mikkelsen Thramkrongart, student number: 201806892, email : 201806892@post.au.dk

Github: https://github.com/PeterThramkrongart/ABM_SocKult

INTRODUCTION

In order to investigate the formation and break-up of echo chambers, we used an existing model of echo chamber generation as our starting point, in which agents make connections to each other and share information in an attempt to find an objectively true value (Madsen et al., 2018).

Our research goals were twofold. Firstly, we wanted to check if the conclusions of Madsen et al. (2018) would replicate across a wide range of existing model parameters. Namely, the conclusion that even idealized Bayesian agents would end up holding demonstrably wrong beliefs with great confidence, and refuse to communicate with other agents who didn't share their worldview. In other words, that they would naturally form echo-chambers. This happened despite the fact that these agents update their beliefs in a rational, Bayesian fashion, and lack a number of the pitfalls of ordinary human communication, such as deception, bias and faulty memory.

Secondly, we wished to investigate how this pattern of echo chamber formation could be broken or counteracted with the introduction of new parameters. In other words, what changes could be made to the model environment to counteract the formation of echo chambers.

We experimented with a two new parameters. The first made agents seek out new connections more often, so as to simulate a highly social or active debating environment; perhaps if humans were less insular, and continuously sought new debating partners, we would more often find the truth? Unfortunately, pilot testing proved this variable so ineffective that, in order to save computing power, we did not vary it in the final version of the model.

The second new introduction to the model was a censorship parameter, prohibiting agents who deviate too much from global consensus from expressing their opinions. This was done to simulate the fairly common practice of censorship, specifically the kind of censorship often found on internet forums which prohibits the expression of certain opinions which are deemed to be too extreme or hurtful for the community. This can be everything from bigotry, such as racism and sexism, to contrary political opinions.

METHODS

The model explores how agents interact with other agents, sample information from others, and use this information to update their confidence in their own beliefs in a Bayesian fashion. These beliefs are represented by values between 0 and 1; the goal for each agent is to figure out the 'truth', represented by the value 0.5, by sampling evidence from other agents.

Model setup

During model setup, each agent samples 5 values from a normal distribution centered around the ground truth. This distribution is given by mean $\mu_{true} = 0.5$ and standard deviation $\sigma_{true} = 0.25$ (representing noise in the evidence). The agent uses the mean of these five values μ_{own} to form its initial belief, or hypothesis H , about the true value of μ . The standard deviation of the five values, σ_{own} , is used by the agent to keep track of the uncertainty in the evidence it has acquired.

In addition to its hypothesis, each agent has a *confidence level*, $P(H)$, which is the agent's own estimate of the probability of its hypothesis H being true. We initialized this value at 0.01, corresponding to each agent having a 1% confidence rating in their hypothesis at model start. If an agent reaches $P(H) = 1$, representing 100% confidence, it will stop sampling information. Thus, when all agents reach 100% confidence, the model stops.

In a hypothetically perfect case, all agents would converge from their random initial beliefs to μ_{true} with 100% confidence in this value (Madsen et al., 2018).

Network Structure

At the start of each simulation, agents are generated and distributed across a two-dimensional space. Madsen et al. (2018) tested two different methods for distributing agents; either the agents' positions were random, or they were distributed in accordance with a 'scale-free' social network structure, in which agents' position relative to each other is governed by a power law (Amaral et al., 2000). Since the findings of Madsen et al. (2018) were generally robust across the two network structures, we only used scale-free networks in our model, in order to decrease computational load, and because these are similar to real social networks (Baronchelli et al., 2013).

Making the model go: Searching for connections

The first step in the model is for agents to make connections, or *links*, with each other. At each model tick, each agent searches for other agents to interact with. This search is constricted by two parameters. The first of these is the *search distance* parameter α , which defines how far across the model space an agent can reach. If $\alpha = 0.5$, the agent can reach 50% of the model space. If $\alpha = 1$, the agent has access to the entire space, and therefore all other agents.

The second parameter affecting agent search is the pruning parameter, β . In essence, β determines the range of beliefs that a given agent will accept. An agent will only accept beliefs that fall within a window around its own hypothesis, the size of which is influenced by β and σ_{own} . Whether an agent will accept another agent is given by $\mu_{own} \pm \beta\sigma_{own}$. The model can also be set to run with 'stochastic' agents, who do not have a pruning parameter at all. Instead, these agents will randomly search for other agents within the search distance α .

Once the agent has determined which agents around it satisfy the constraints of α and β , it chooses one of these agents at random and forms a *link* with it, representing a social connection and allowing them to exchange information. This link remains for subsequent ticks, unless broken later. If possible, an agent will usually make one new link per tick.

Sampling information from links and updating confidence

Once the search procedure is over, the next stage in the model is the actual information sampling. The agent chooses a random agent among *all* of its current links (including ones from previous ticks), and compares that agent's hypothesis to its own. It then adjusts its confidence in its own hypothesis using Bayes Theorem:

$$P(H|E) = \frac{P(H) \cdot P(E|H)}{P(E)}$$

$P(H)$ is the agent's current confidence level (goes from 1% to 100%), and $P(H|E)$ is its confidence after considering the evidence E (the other agent's hypothesis). The probability of the evidence, $P(E)$ is given by a normal distribution's probability density curve, calculated using the true normal distribution ($\mu = 0.5$, $\sigma = 0.25$):

$$P(E) = \frac{1}{\sigma_{true} \cdot \sqrt{2\pi}} \cdot e^{-\frac{(X-\mu_{true})^2}{2\sigma_{true}^2}}$$

$P(E|H)$, the probability of the evidence E given the agent's own hypothesis H , is likewise calculated with a probability density curve, this time using the agent's own distribution:

$$P(E|H) = \frac{1}{\sigma_{own} \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-\mu_{own})^2}{2\sigma_{own}^2}}$$

Thus, if $P(E|H) > P(E)$, the agent's confidence $P(H|E)$ increases, and vice versa.

Updating beliefs and assessing belief purity

Once this is done, the agent recalculates μ_{own} and σ_{own} while including the new evidence, thus updating its hypothesis. Finally, the agent prunes any connections to agents that were previously acceptable, but now are not, either because their beliefs have become too extreme, or because the agent's pruning window has decreased.

Additionally, at each tick, all agents assess the 'belief purity' of their local network; that is, how different their hypothesis μ_{own} is from the hypotheses of its links. Belief purity ϵ is assessed by each agent as:

$$\epsilon = 1 - \frac{\sum_{i=1}^n \Delta_i}{n}$$

Where Δ_i is the difference $\mu_{own} - \mu_i$ for each linked agent i . In other words, belief purity ϵ is the mean difference in belief between an agent and its links. Globally, belief purity is calculated as the average ϵ across all agents.

Conclusions of the previous study

Madsen et al. (2018) found that agents gradually and consistently increased their confidence over time, and global belief purity consistently reached above 90%. Moreover, many agents would reach complete confidence in a demonstrably wrong hypothesis ($\mu_{own} \neq 0.5$). In other words, agents would gather in echo chambers, sharing information about wrong hypotheses (albeit, hypotheses clustered closely around 0.5). This pattern persisted even when a certain percentage of all agents were anti-confirmatory (deliberately seeking out contrary opinions), and even when the agents repeatedly received a 'broadcast' of the ground truth, 0.5, at regular intervals. Broadcasts were observed to be helpful, but did not prevent all echo chamber formation. (Madsen et al., 2018)

Our model design

When designing our own model, we varied many of the same parameters as in the original paper. We tested several different levels of α (0.05, 0.1, 1) and β (stochastic, 0.1, 1, 2). We also varied the presence (every 5 ticks) or absence of broadcasts. Due to computational limitations, and because their overall impact on the model was judged to be quite small (Madsen et al., 2018), we elected to exclude anti-confirmatory agents from the model entirely. Instead, we experimented with two new parameters.

The first was a *searching parameter*, which modified the way agents search for new links. In the original model, agents will search less often the more confident they become (the probability of seeking new links decreases as $P(H)$ approaches one). We wished to investigate what would happen if agents searched for new links every tick. Thus, the searching parameter had two levels. The first was random search, which functioned like the original model (Agents have a probability of $1 - P(H)$ of searching each tick). The second was constant search (Agents always search for new links unless $P(H) = 1$).

Early pilot testing, however, showed almost no impact of changing the search parameter, except for a negligible effect on the speed with which agents reached 100% confidence. This is probably because, even if agents stop making *new* links, they still sample information from *existing* links on every tick. Since the impact of this change

was very small, we chose not to test both levels of this parameter in our final run of the model, as this would have doubled the amount of simulations. This of course meant that we had to decide whether our final model should use random or constant search. In the end, we chose constant search, in the hope that this would make the model run slightly quicker.

The second new parameter we wanted to test was the impact of *censorship*. To do this, we created a censorship parameter C which constrains (based on the ongoing global consensus) whom agents are allowed to interact with. Concretely, if censorship is present, each agent will be prohibited from forming links with, or sampling information from, other agents whose beliefs fall outside a certain window around the global consensus. This window is based on the mean (μ_{global}) and standard deviation (σ_{global}) of all agents' hypotheses, and the censorship parameter C :

$$\mu_{global} \pm (\sigma_{global} \cdot C)$$

The censorship parameter thus functions much like the pruning parameter, except for two critical differences:

- It works at a global level rather than at the level of the individual agent.
- Censorship doesn't remove existing links between agents, it merely prevents the forming of new ones. Likewise, it only prevents sampling *from* censored agents. Censored agents can still sample from non-censored agents. Otherwise, censorship would render them unable to participate at all.

We tested four different levels of C (None, 0.5, 1, 2) for all levels of α and β , with and without broadcasts. Furthermore, for each parameter combination, we repeated the simulation 10 times. Finally, we varied the number of agents; 100 and 1000. Thus, we ran $3 * 4 * 4 * 2 * 2 = 192$ unique parameter combinations, for a total of 1920 simulations. However, we will only report on the simulations using 1000 agents in this paper, since the results are very similar and we were mainly interested in large networks.

We limited each simulation to 50 ticks, as we observed that for most parameter settings, global confidence would reach 1, or stabilize close to 1, in this interval.

The model was adapted from Netlogo 5.2.1 and run in Netlogo 6.0.4, using the RNetlogo and nlexperiment packages (Bergant, 2015/2019; Thiele, 2017).

RESULTS

Replication of previous findings

This section details our findings from simulations where censorship was absent, reminiscent of the original model. Like Madsen et al. (2018), we observe a general tendency for agents' beliefs to cluster around the objective mean (figure 1 and 4), and in all but four cases (which will be discussed later), we observed swift growth in global confidence, fluctuating near 100% (figure 2). This happens even though many agents end the simulation holding a wrong belief. In the end, many agents have grown absolutely sure about an objectively wrong belief, and we do not see all agents converge at the objective truth under any parameter combination.

Generally, we observe that the *search distance* parameter α does not seem to make a noticeable difference in what belief agents end up with. As demonstrated by figure 1, we see similar belief dispersion at all settings of α and no substantial interactions with other parameters. As α did not seem to have much impact on neither belief purity nor confidence over time, figures 2-5 only display results from simulations where $\alpha = 0.1$.

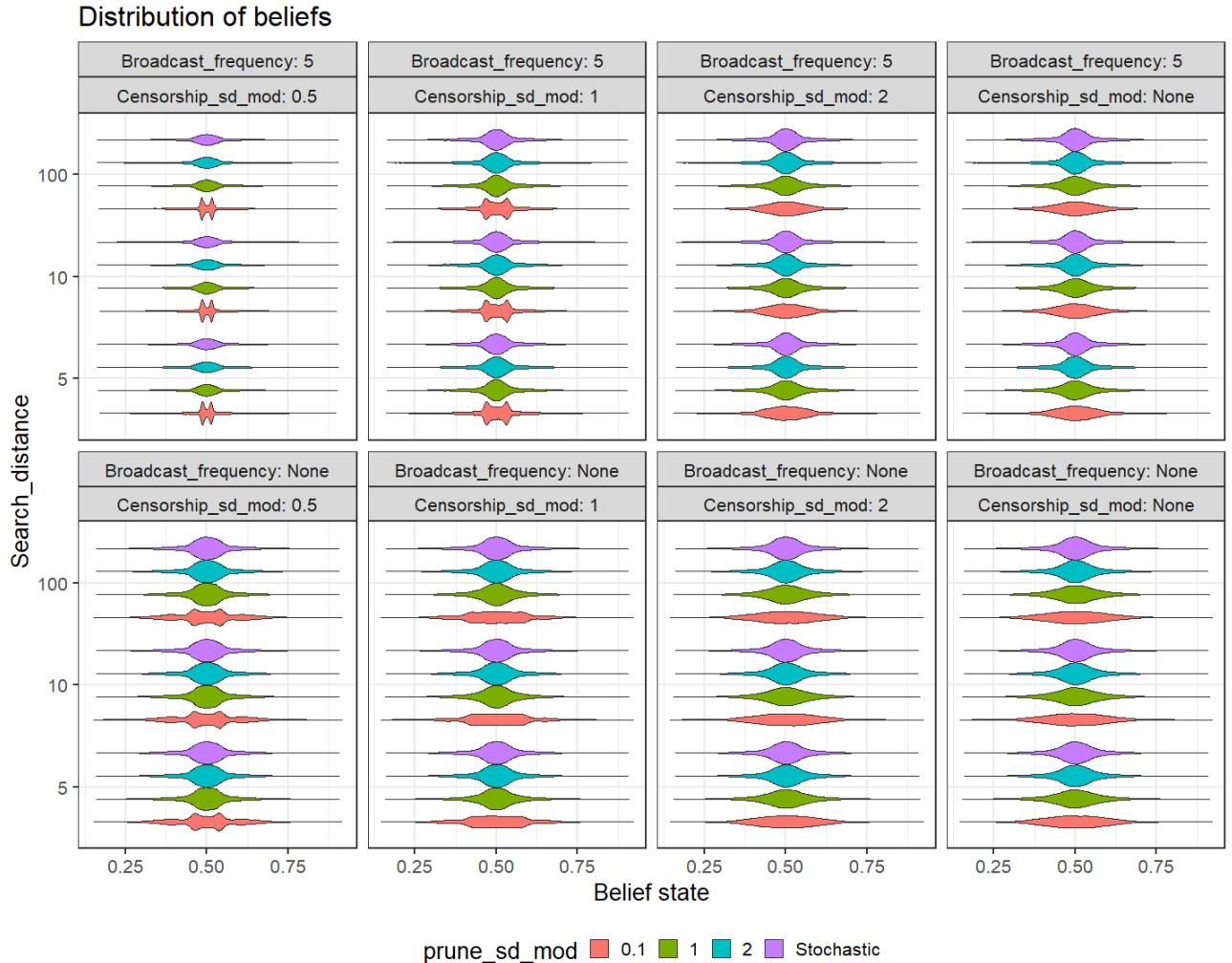


Figure 1: The distribution of beliefs (value of μ_{own} for each agent) at tick 50 over 10 repetitions. The y-axis maps the search distance α measured in percentages and the colors map the setting of pruning parameter β . Plots are sorted by degree of censorship C , and by frequency of broadcasts (every 5th tick or none at all). This figure includes all 96 parameter combinations.

We observe that broadcasts of the objective truth every 5th tick does somewhat narrow the belief distributions, but it does in no way prevent echo chambers all together. Broadcasting does not have a noticeable effect on the growth of confidence (See figure 2), however, which means that once confidence reaches 1, broadcasting will become next to useless.

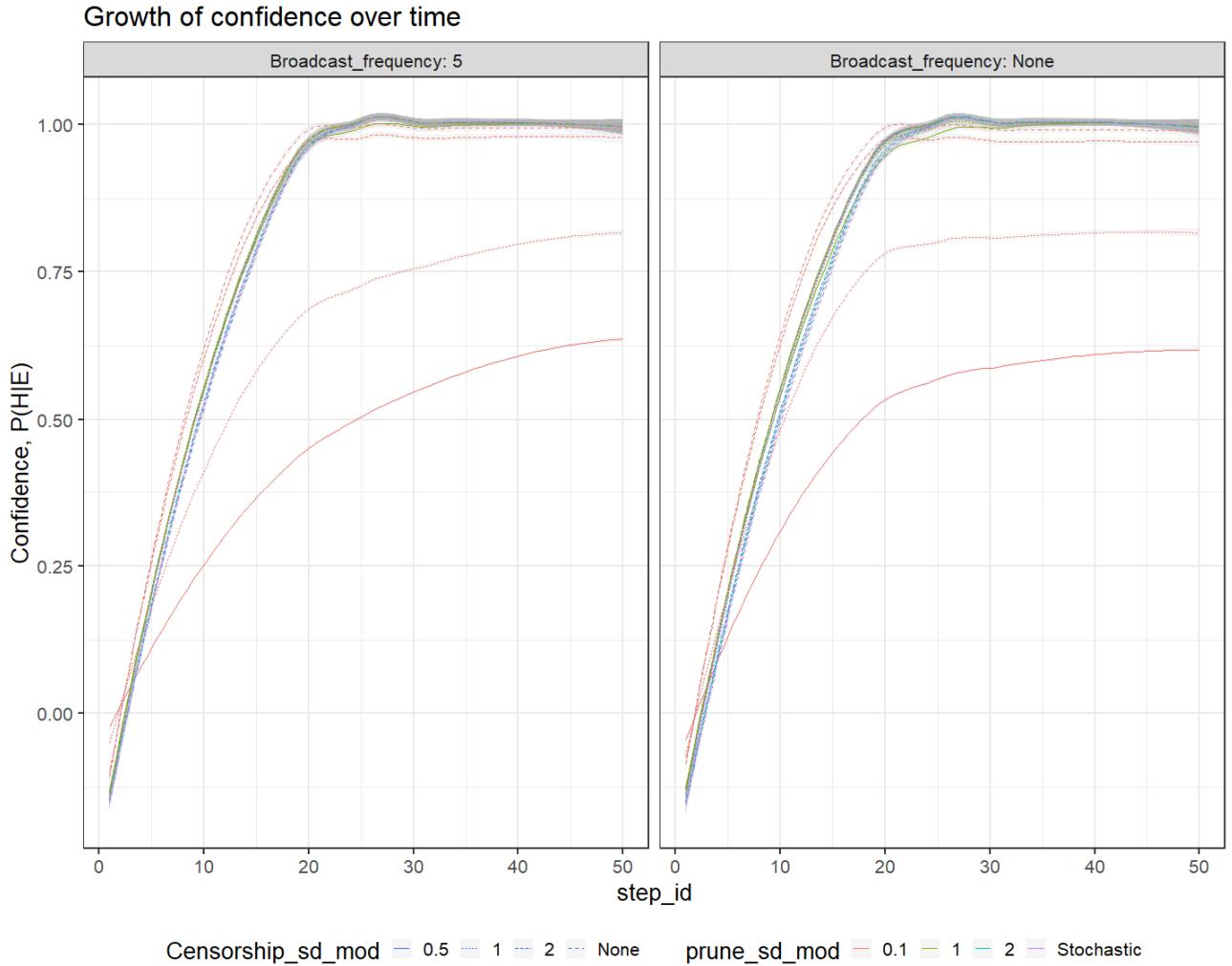


Figure 2: The growth of confidence $P(H|E)$ over 10 repetitions, with search distance α set to 10%. Colors map the pruning parameter β , and linetypes map the censorship modifier C . Plots are sorted by search distance measured in percentages and by frequency of broadcasts (either every 5th tick or none at all).

With regards to belief purity (displayed in Figure 3), we observe consistently high values (above 90% regardless of parameter combination), and also see that networks with more confirmatory pruning parameters, meaning a low β value, form echo chambers with higher belief purity than more open minded agents with a higher β , or even no β (stochastic- agents).

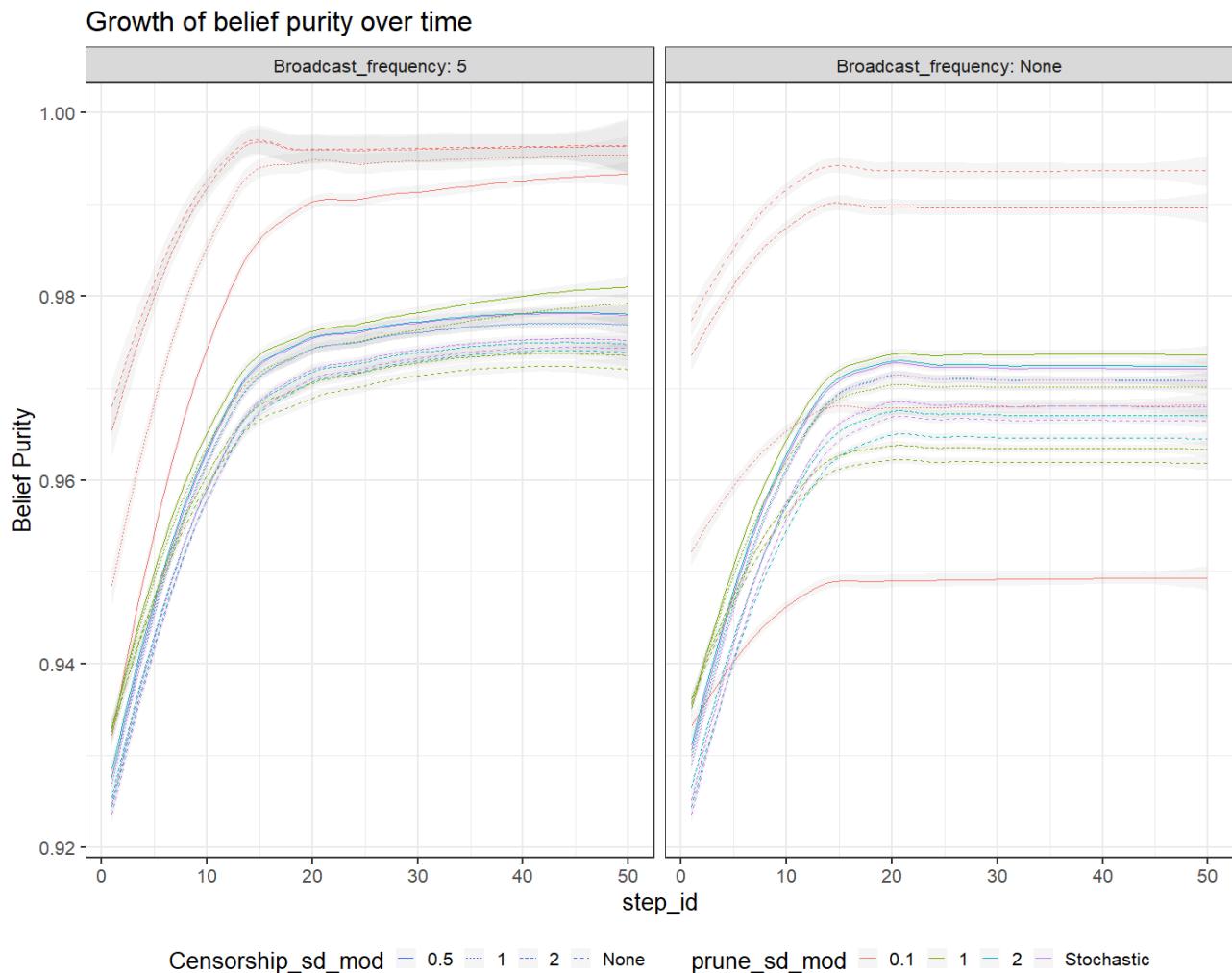


Figure 3: Growth of belief purity over time, across 10 repetitions, with search distance α set to 10%. Colors map the pruning parameter β and linetypes map the censorship modifier C . Plots are sorted by frequency of broadcasts (either every 5th tick or none at all).

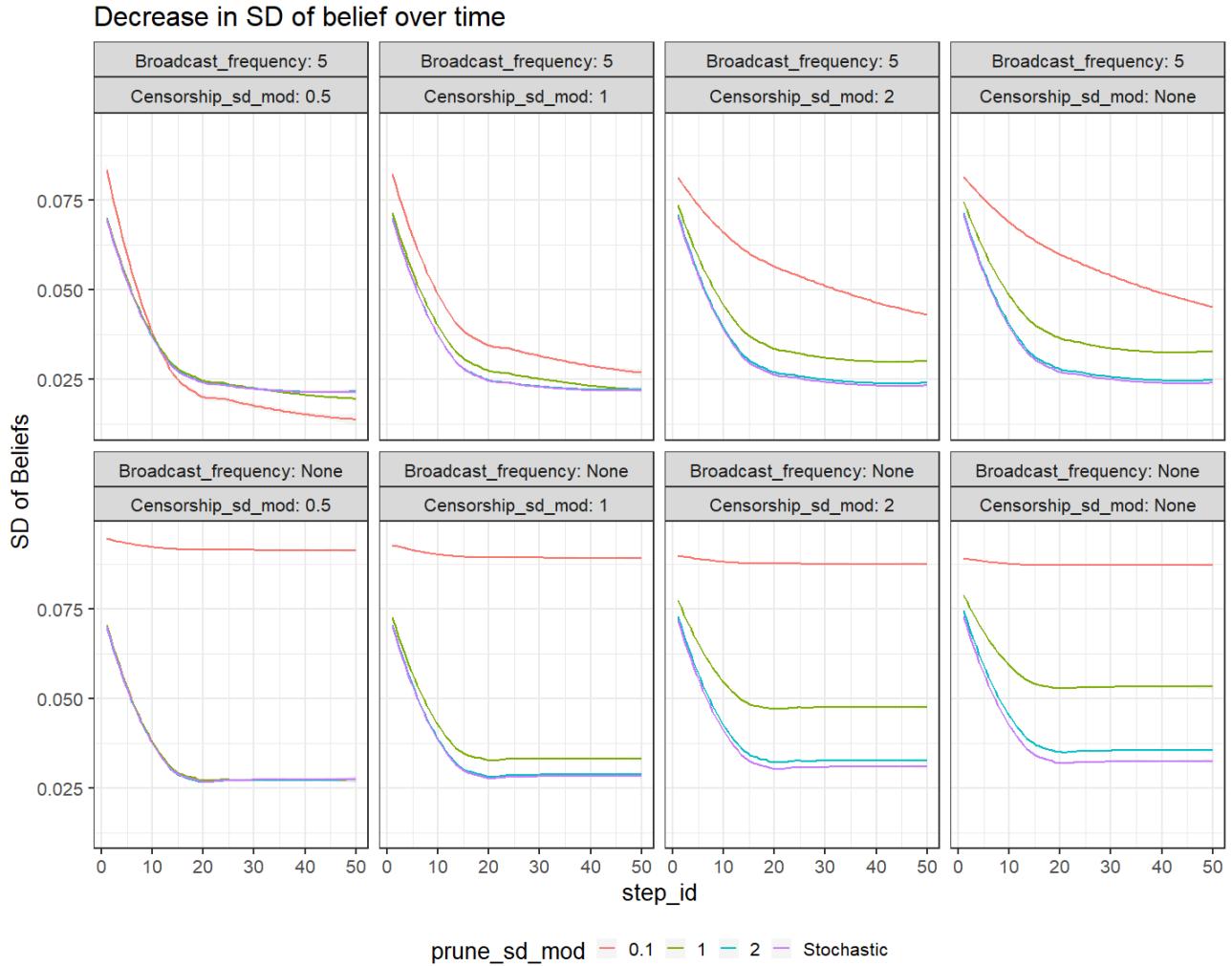


Figure 4: Global SD of belief over time, across 10 repetitions, with search distance α set to 10%. Colors map the pruning parameter β . Plots are sorted by the censorship modifier C and by broadcast frequency (every 5th tick or none at all).

Preventing echo chambers with censorship

At the most reasonable setting ($C = 2$, meaning a window of 2 SD's from μ_{global}) censorship has very little effect across all combinations. This means that silencing extreme outliers, or ‘extremists’, has a minuscule effect on the outcome. The effect is more dramatic when we lower the threshold for the censorship. When broadcasting is absent (bottom row of figure 1) we see a slightly narrower belief distribution in proportion to the censorship threshold ($\beta = 0.5, 1$). This effect is much stronger when broadcasting is present, however (top row of figure 1). In combination with frequent broadcasting of the true state (0.5), censorship proves rather effective. We still observe that many agents stick with wrong beliefs, but the distributions are much narrower under these conditions. These results are very different with narrow minded agents. This will be discussed in the following section.

Investigating the two-party formation

As previously mentioned, four edge cases did not reach close to 100% confidence after 50 ticks. Those were the cases in which $\beta = 0.1$, and strict censorship ($C = 0.5, 1$) was applied. Furthermore, as Figure 5 demonstrates, these cases ended with a very peculiar belief distribution.

Evidently the effect of harsh censorship is very different on very narrow minded agents ($\beta = 0.1$). Here we see that the distribution of beliefs no longer form gaussian distributions, but rather different versions of symmetric bimodal distributions, much akin two an extreme two-party system with two very similar, but not identical, parties. Very few agents actually find the true mean of 0.5, and the two clusters of agents never seem to merge. The presence or absence of broadcasts mainly seem to affect the number of outliers in the distribution. Interestingly, when broadcasts do occur, outliers more or less vanish completely, leaving two echo chambers right around the mean.

Distribution of beliefs

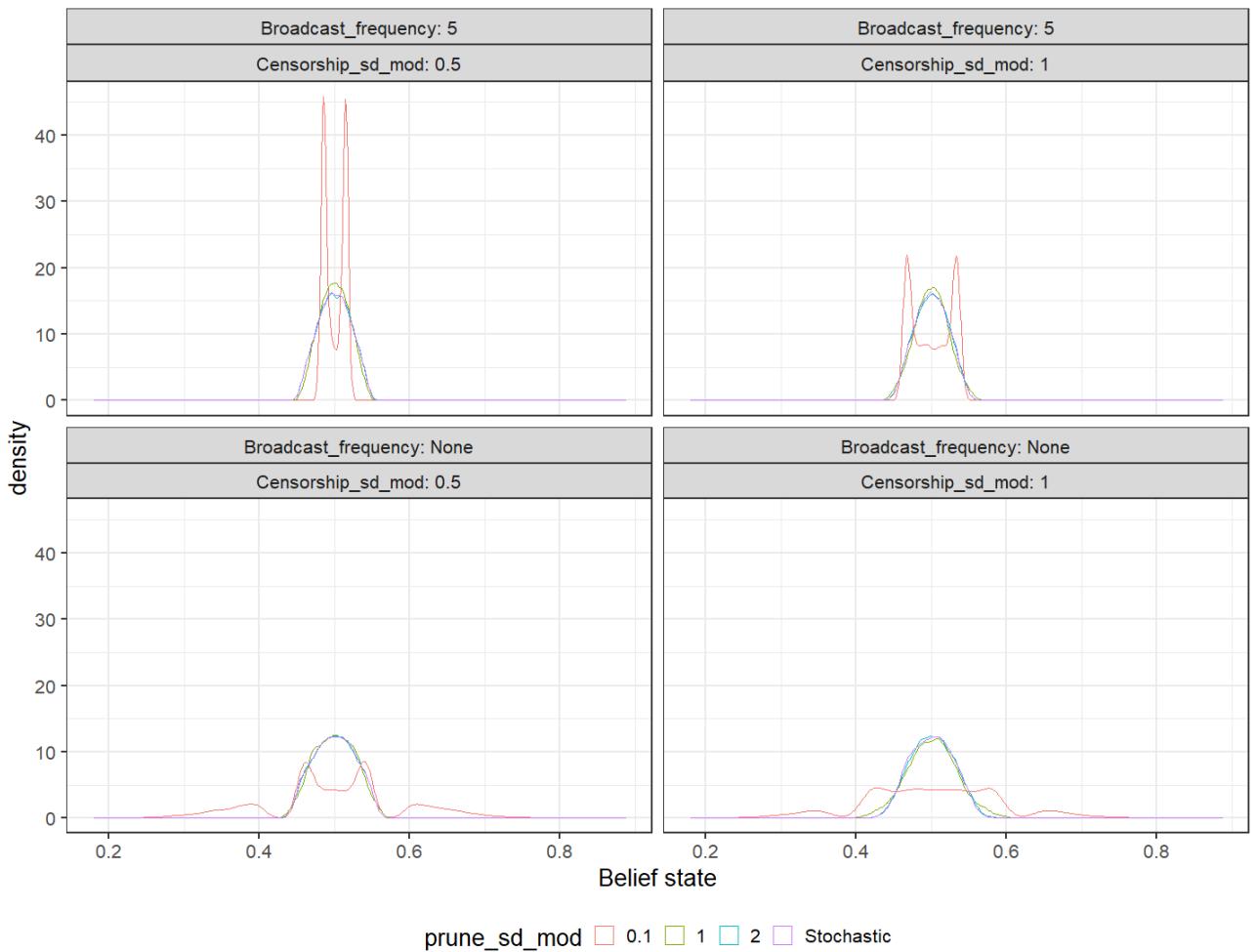


Figure 5: The distribution of beliefs (value of μ for each agent) at tick 50 over 10 repetitions with the search distance set to 10% of the network. Colors map the pruning parameter (SD). Plots are sorted by degree of censorship (SD) and by frequency of a broadcast stating the true state of the world (0.5). Broadcast frequency is either every 5th tick or no broadcast at all.

This pattern was so interesting that we investigated it further. Perhaps if we gave the agents more time, this bimodal structure would vanish? After all, they hadn't reached 100% confidence after 40 ticks. As such, we ran those specific simulations again, this time with 200 ticks and 5 repetitions (the results of which is displayed in Figure 6). We found that the pattern persisted over time, even when helped along by broadcasts. In short, harsh censorship paired with narrow-minded agents seem to cause partisanship; i.e., two extremely potent echo chambers.

Distribution of beliefs

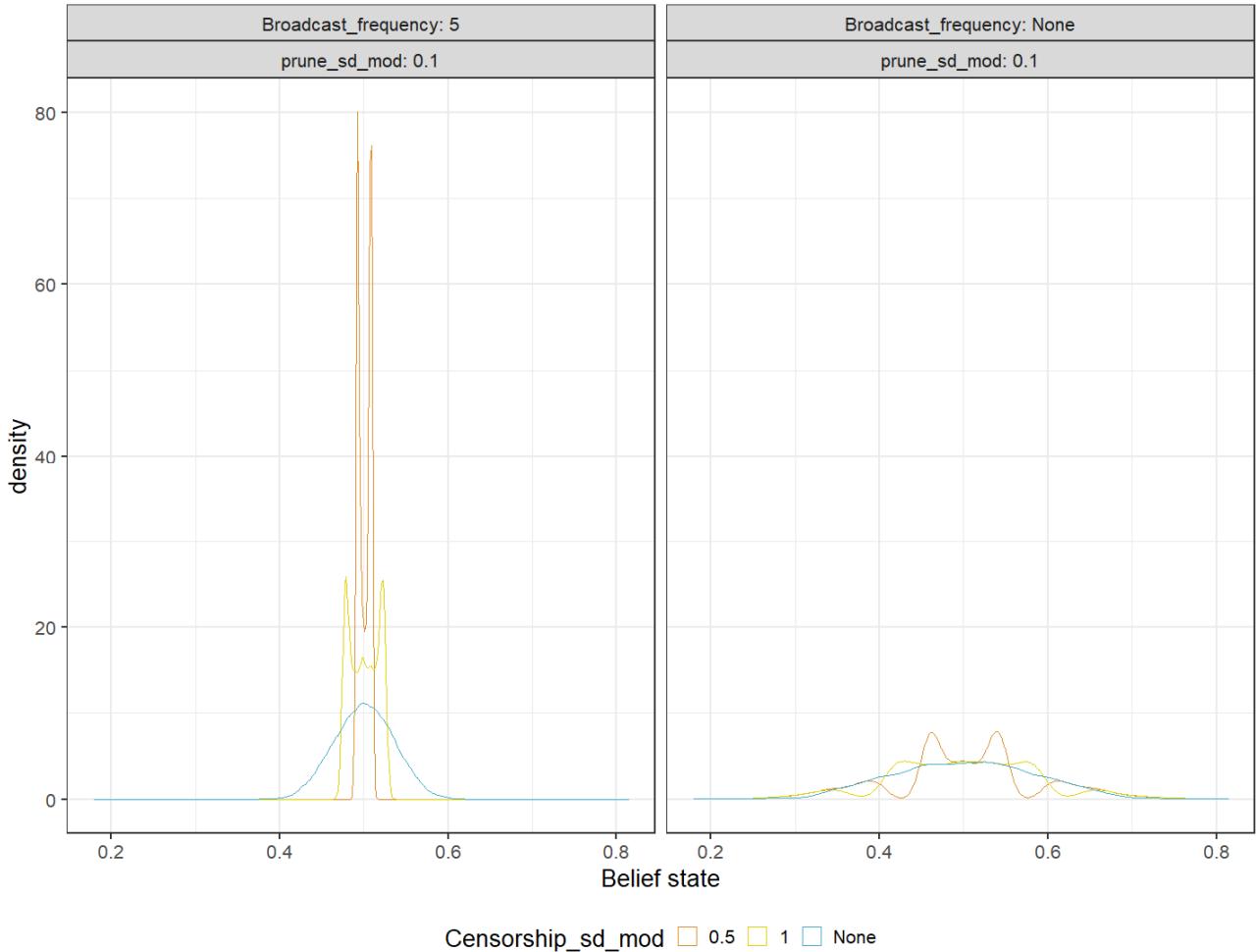


Figure 6: The distribution of beliefs (value of μ for each agent) at tick 200 over 10 repetitions with search distance α set to 10%. Colors map the pruning parameter β . Plots are sorted by search distance measured in percentages and by frequency of broadcasts. Broadcast frequency is either every 5th tick or no broadcast at all.

DISCUSSION

Our findings

We can conclude that within the space of our models persistent echo chambers with wrong beliefs do seem to form in large social networks. Open minded, error-free bayesian agents consistently form tightly knit echo chambers that epistemically deviate from the true mean (0.5). We also saw that frequent educational

broadcasting of the true state of the world was able to pull many agents closer towards the truth. All of these findings align closely with those of Madsen et al. (2018).

Furthermore, we found that light censorship very slightly helped most agents getting closer to the true mean, and that combining that censorship with educational broadcasts helped even further. We consider this finding to be encouraging as light censorship and education are realistic intervention strategies in preventing echo chamber formation on platforms like digital social media. We also found that harsher degrees of censorship was mainly helpful against open minded agents; it arguably had the wrong impact on very narrow minded agents, who under such conditions rarely find the true mean, but instead form a two-party system very close to the mean. We do want to stress, that this sort of harsh censorship is of course an ethically unviable intervention strategy.

Limitations

The main limitation to our model was, as has already been mentioned, computing power. As more parameters are introduced, the amount of simulations that has to be run grows massively. As such, we left several potential variables of interest (such as anti-confirmatory agents) out of the final model. Although care was taken to choose the variables which appeared to be of small consequence to the outcome, it is entirely possible that we might have missed out on interesting variable interactions.

Additionally, the generalizability of our findings to real-world scenarios is obviously limited. First of all, real people are not idealized Bayesian agents and many subjects on which echo chambers form (e.g. politics or religion) do not have an objectively true belief. Secondly, censorship is rarely so efficient as to completely block out dissenting voices from speaking. Finally, in our model, the pruning parameter β and search distance α is shared by the entire distribution of agents, and remains fixed over time. Everyone has equal reach and is equally conservative, whereas in the real world, these dimensions would vary across people, and would quite conceivably change over time. An interesting variation of this would be to assign β -values to individual agents using a normal or log-normal distribution rather than setting a fixed value for all agents.

References

- Amaral, L. A. N., Scala, A., Barthelemy, M., & Stanley, H. E. (2000). Classes of small-world networks. *Proceedings of the National Academy of Sciences*, 97(21), 11149–11152.
- Baronchelli, A., Ferrer-i-Cancho, R., Pastor-Satorras, R., Chater, N., & Christiansen, M. H. (2013). Networks in cognitive science. *Trends in Cognitive Sciences*, 17(7), 348–360.
- Bergant, D. (2019). *Bergant/nlexperiment* [R]. <https://github.com/bergant/nlexperiment> (Original work published 2015)
- Madsen, J. K., Bailey, R. M., & Pilditch, T. D. (2018). Large networks of rational agents form persistent echo chambers. *Scientific Reports*, 8. <https://doi.org/10.1038/s41598-018-25558-7>

Thiele, J. C. (2017). *RNetLogo: Provides an Interface to the Agent-Based Modelling Platform “NetLogo”*

(Version 1.0-4) [Computer software]. <https://CRAN.R-project.org/package=RNetLogo>

Appendix

Growth of confidence over time

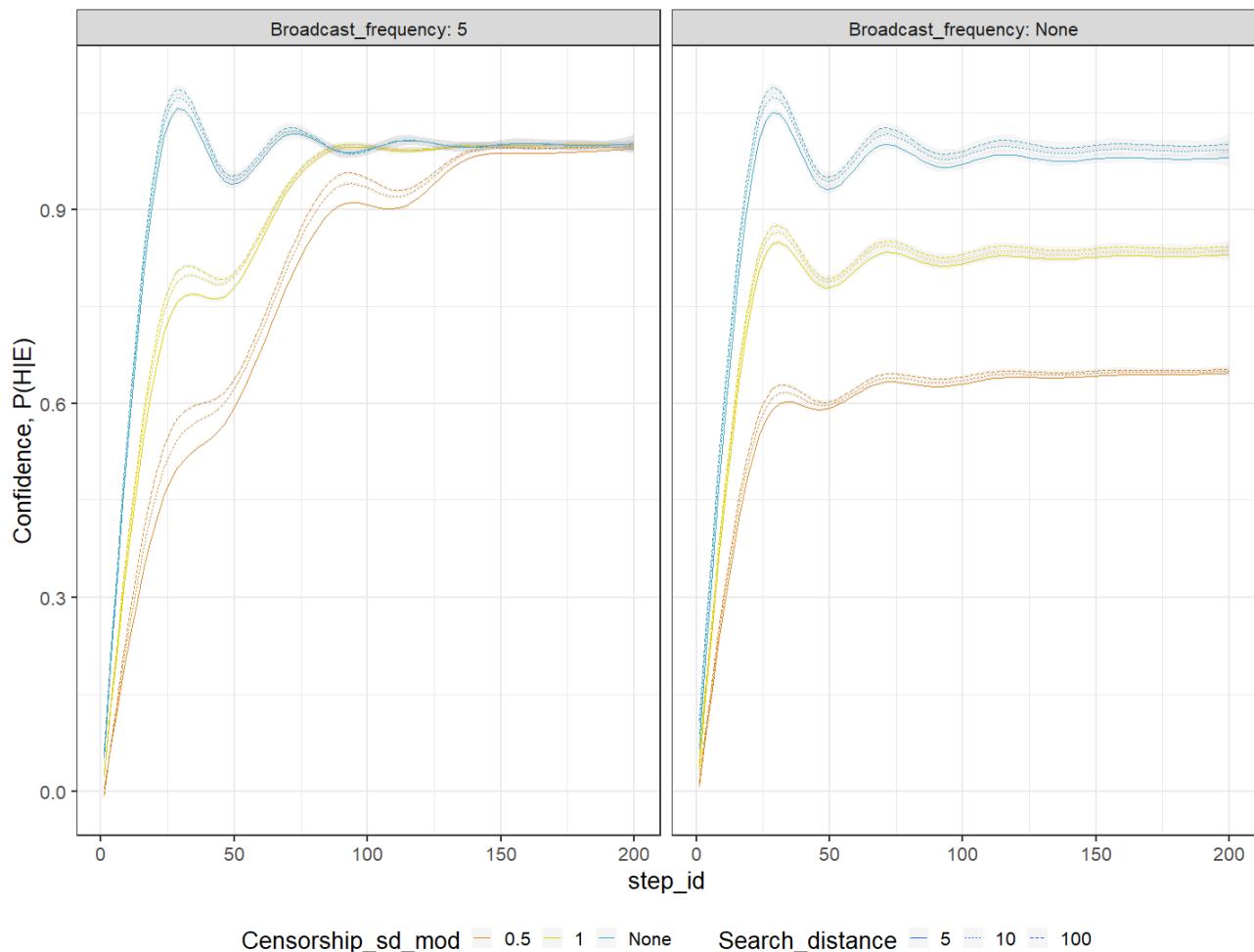


Figure 7: The growth of confidence over time, with pruning parameter β set to 0.1 over 5 repetitions. Colors map the degree of censorship C and the linetypes map the the search distance α measured in percentages. Plots are sorted by frequency of broadcasts (either every 5th tick or no broadcast at all)