Dana Jensen (DJ) Portfolio

Ft. Riley Anthony (RA), Kenneth Enevoldsen (KE), and Blanka Zana (BZ)

GitHub Assignments 1-4
https://github.com/au558796/ComputationalModelling

GitHubAssignment 5
Sentiment and Analysis: https://github.com/KennethEnevoldsen/BeersAndBitcoin
Market Herding Calculation: https://github.com/au558796/Beers-and-Bitcoin

Assignment 1 (RA, DJ, BZ)

Visual Search Experiment

Introduction

In the following report we will discuss the replication of a previously performed study which sought to investigate the differences underlying top down and bottom up driven visual foraging. Two experimental tasks were used to qualify these phenomena. A task involving counting targets and a task involving finding a hidden target. It is supposed that the counting task is driven by bottom-up cognitive processes, while the target locating condition is a top-down process. It is assumed that patterns of visual foraging work similarly to those of association network mechanisms, illustrated by similar long-tailed distributions. This experiment was set to investigate whether this pattern can be also found when executing mechanically different visual search tasks, designed with two conditions. In one condition, the participant was asked to count targets in a clustered image, while in the other condition the participant was tasked with looking for a star hidden within various images.

Hypotheses

Conceptual: Visual search patterns are affected by task structure (top-down)

Operational: Saccade amplitude distributions are affected by task structure

Methods

Data collection

Eye-tracking data of monocular position and pupil size was collected with a head mounted Eye Link 1000. The data was recorded on the setting of 1000Hz (recording data once per every millisecond). The experiment took place in a basement room with the lighting set to be as consistent as possible for each participant. Participants were seated approximately 30 inches in front of a 30-inch flat panel LCD monitor. The eye-tracker was calibrated prior to the experiment by using the in-built nine-point automated calibration procedure. This was repeated until the validation procedure reported an average error lower than 1 and max error lower than 1.5.

Participants

Six participants were part of the experiment (four female, two male), all university students (between age of 20 and 30). They were not required to meet any specific criteria, excluding neither eyeglasses nor cosmetics. One participant was wearing glasses. Average error of calibration among all participants was 0.87.

Data pre-processing

The collected data was automatically pre-processed using the in-built DataViewer software (including velocities, pupil size and x and y coordinates). Fixations, saccades and eye-blinks were identified, artifacts were removed. Previously mentioned validation data was used to exclude participants and model measurement error. The data was high-pass filtered at a 100 cut off to counter calibration drift. Systematic bias on fixation estimates on fixation crosses was estimated and positions were accordingly adjusted at every trial.

Data analysis

Two hypotheses were tested in order to investigate the main hypothesis. Firstly, we investigated whether differences between the two kinds of tasks could be explained in terms of fixation durations, then in saccade amplitudes. Three models used were selected to best represent the first hypothesis and are as follows:

Duration ~ SearchType + Trial + (1+SearchType |ParticipantID)

Duration ~ SearchType * Trial + (1+SearchType |ParticipantID)

Duration ~ SearchType + (1+SearchType |ParticipantID)

Models to test the second hypothesis are as follows:

Amplitude ~ SearchType + Trial + (1+SearchType | ParticipantID)

Amplitude ~ SearchType * Trial + (1+SearchType | ParticipantID)

Amplitude ~ SearchType + (1+SearchType | ParticipantID)

It is posited that eye fixations reflect cognitive information processing. The assumption is that when the eye fixates on an element in the visual field, the simple information (colour, geometrical shape, etc.) embedded in the element being viewed is processed automatically. Furthermore, when not in a fixation, information in the visual scene is not processed (when, for example, the eye in in the midst of a saccade). The two different visual search tasks were chosen to display the consequences of top-down and bottom-up driven attention, top down meaning the participant's prior beliefs and assumptions direct their attention, whereas in bottom-up, the stimuli in the environment directs the participant's attention. The counting visual task was thus designated to be a bottom-up condition, where the viewer's gaze was guided by the elements in the picture to count, whereas the visual search task was a top-down condition because the viewers search pattern would not be as influenced by elements in the visual field.

On the other hand, with amplitude, the main hypothesis is that the saccade amplitude will change based on the type of the visual task. Therefore, our first predictor is search type. If true, we will see a significant beta value for search type, where the saccade amplitude will be either higher or lower depending on the visual task. Secondly, we suspect that the trial will have an effect on the saccade amplitude. The further into the experiment the participant, the more likely they are to either a) be better at the learned task through experience, and/or b) become fatigued and therefore worse at performing. Therefore, to control for this, we add trial to the equation to account for the difference. In our interaction model (#2) we assume that the effects of the trial the participant is on will also influence the effect that the visual task will have. For the random effects, we include search type and participant, which assumes that each participant will have a different intercept and slope depending on the nature of the visual task they are performing.

We decided to create mixed effects models in order to control for repeated measures (participants taking part in multiple trials) and to account for random influences and individual differences (as explained above). Eye tracking data is not a normal distribution, but an exponential distribution. We therefore log-transform our data to meet the criteria of a normal distribution for the analysis.

$$\text{Fixation Duration} = \beta_0 + \beta_1 * \text{Visual Task} + \beta_2 * \text{Trial} + \beta_3 * \text{Visual Task} * \text{Trial}$$

$$\text{Saccade Amplitude} = \beta_0 + \beta_1 * \text{Visual Task} + \beta_2 * \text{Trial} + \beta_3 * \text{Visual Task} * \text{Trial}$$

We analysed the data using the statistical software of R studio (RStudio Team, 2015). Before performing a cross validation on the models, we observed the *Akaike Information Criterion* (AIC) and the *Bayesian Information Criterion* (BIC) for all the models (shown in table below). Although from these information criteria we would select model #2, we will perform a cross validation in order to retain the out-of-sample error that will determine which model is best at prediction and generalisation.

Model	**AIC**	**BIC**
Duration ~ SearchType + Trial + (1 + SearchType | ParticipantID)	56379.0	56423.4
Duration ~ SearchType * Trial + (1 + SearchType | ParticipantID)	56379.9	56430.7
Duration ~ SearchType + (1 + SearchType | ParticipantID)	56382.5	56420.6
Amplitude ~ SearchType + (1+ SearchType | ParticipantID)	22649.4	22687.4
Amplitude ~ SearchType + Trial + (1 + SearchType | ParticipantID)	22651.1	22695.4
Amplitude ~ SearchType * Trial + (1 + SearchType | ParticipantID)	22642.1	22692.8

We then proceeded to perform a 4-fold cross validation with each model on our dataset. As a result, the fixation duration model with the lowest out-of-sample error was model #3 with the interaction effect of visual task and trial, with an average test error (rmse) of 319.05. In regards to saccade amplitude, the model with the lowest out-of-sample error was model #3 with the interaction effect of visual task and trial, with an average test error (rmse) of 113.45. The second best fixation duration model was #1 with only the effect of visual task, with an average test error (rmse) of 334.25. The second best saccade amplitude model was #1 with only the effect of visual task, with an average test error (rmse) of 117.82. Finally, the worst performing model was model #2 with main effects of visual task and trial, with an average test error (rmse) of 337.75, and the worst performing saccade amplitude model was model #2 with main effects of visual task and trial, with an average test error (rmse) of 132.91.

	perf_test_list	perf_train_list	temp_model
1	268.4409	359.9391	glmer(Duration~SerachType + Trial + (1+SearchType |Partici...
2	385.8417	313.6724	glmer(Duration~SerachType + Trial + (1+SearchType |Partici...
3	412.2505	334.0085	glmer(Duration~SerachType + Trial + (1+SearchType |Partici...
4	284.4835	365.8500	glmer(Duration~SerachType + Trial + (1+SearchType |Partici...
5	186.2325	350.3721	glmer(Duration~SearchType * Trial + (1+SearchType |Partici...
6	344.8411	345.8674	glmer(Duration~SearchType * Trial + (1+SearchType |Partici...
7	412.2505	334.0085	glmer(Duration~SearchType * Trial + (1+SearchType |Partici...
8	332.8822	352.8444	glmer(Duration~SearchType * Trial + (1+SearchType |Partici...
9	390.3502	332.1822	glmer(Duration~SearchType + (1+SearchType |ParticipantID)...
10	393.7190	318.4651	glmer(Duration~SearchType + (1+SearchType |ParticipantID)...
11	268.4409	359.9391	glmer(Duration~SearchType + (1+SearchType |ParticipantID)...
12	284.4835	365.8500	glmer(Duration~SearchType + (1+SearchType |ParticipantID)...

	perf_test_list	perf_train_list	temp_model
1	34.55868	164.44468	glmer(Amplitude~SerachType + Trial + (1+SearchType |Parti...
2	302.36964	84.30423	glmer(Amplitude~SerachType + Trial + (1+SearchType |Parti...
3	114.97578	152.74402	glmer(Amplitude~SerachType + Trial + (1+SearchType |Parti...
4	86.60286	180.10063	glmer(Amplitude~SerachType + Trial + (1+SearchType |Parti...
5	105.22844	157.35478	glmer(Amplitude~SearchType * Trial + (1+SearchType |Partic...
6	302.36964	84.30423	glmer(Amplitude~SearchType * Trial + (1+SearchType |Partic...
7	114.97578	152.74402	glmer(Amplitude~SearchType * Trial + (1+SearchType |Partic...
8	47.12694	187.62484	glmer(Amplitude~SearchType * Trial + (1+SearchType |Partic...
9	123.74246	147.58673	glmer(Amplitude~SearchType + (1+SearchType |ParticipantI...
10	219.01334	78.31866	glmer(Amplitude~SearchType + (1+SearchType |ParticipantI...
11	34.55868	164.44468	glmer(Amplitude~SearchType + (1+SearchType |ParticipantI...
12	89.23597	172.70326	glmer(Amplitude~SearchType + (1+SearchType |ParticipantI...

Report

Because model #3 (model 2 in the above table) had the lowest out-of-sample error of the three (319.05) we chose it as the best performing model, and continued to use it to analyse the full data set.

	temp_model	mean_test_rmse
1	glmer(Duration~SerachType + Trial + (1+SearchType |Partici...	337.7541
2	glmer(Duration~SearchType * Trial + (1+SearchType |Partici...	319.0516
3	glmer(Duration~SearchType + (1+SearchType |ParticipantID)...	334.2484

	temp_model	mean_test_rmse
1	glmer(Amplitude~SerachType + Trial + (1+SearchType |Parti...	132.9126
2	glmer(Amplitude~SearchType * Trial + (1+SearchType |Partic...	113.4508
3	glmer(Amplitude~SearchType + (1+SearchType |ParticipantI...	117.8176

Fixation Duration

```
Fixed effects:
                       Estimate Std. Error t value Pr(>|z|)
(Intercept)            5.713337   0.122447   46.66   <2e-16 ***
SearchTypeSearch      -0.309674   0.147232   -2.10   0.0354 *
Trial                  0.007844   0.011377    0.69   0.4905
SearchTypeSearch:Trial 0.015267   0.014754    1.03   0.3008
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The first main effect of the visual search type had a significant influence on the fixation duration (β = -0.30, SE = 0.15, $p = .03$), meaning that when the visual task was the foraging (search) task, the fixation duration was 300ms faster than when the visual task was counting. Moreover, trial did not have a significant influence on fixation duration, (β = 0.01, SE = 0.11, $p = .49$), meaning there was not a substantial influence of expertise or fatigue on performance as discussed earlier under the data analysis section. Finally, the interaction effect of visual search type and trial did not have a significant influence on fixation duration, (β = 0.02, SE = 0.15, $p = .30$), meaning the trial number did not have a substantial influence on fixation duration depending on which visual task was being performed.

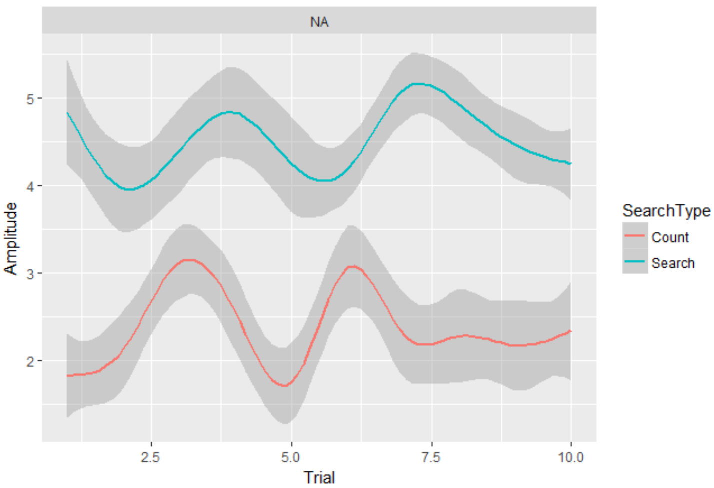
Saccade Amplitude

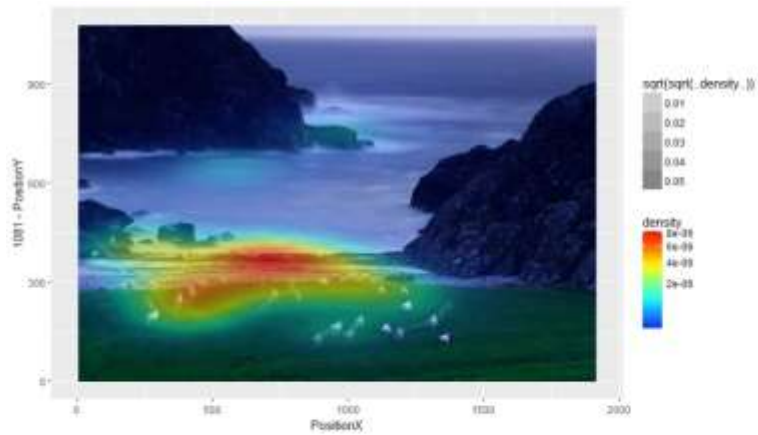
```
Fixed effects:
                       Estimate Std. Error t value Pr(>|z|)
(Intercept)             0.81242    0.10358   7.843 4.39e-15 ***
SearchTypeSearch        0.70820    0.11993   5.905 3.52e-09 ***
Trial                   0.01035    0.01537   0.674    0.500
SearchTypeSearch:Trial -0.01457    0.02067  -0.705    0.481
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

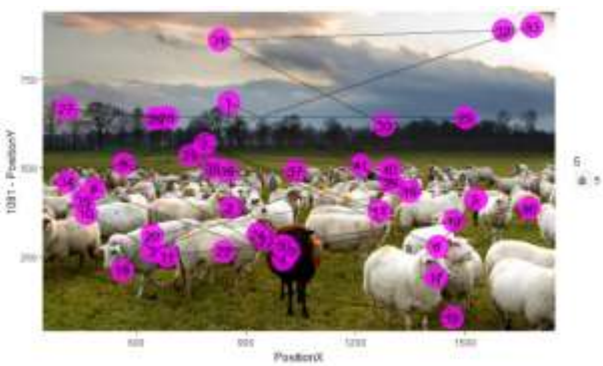
The first main effect of the visual search type had a significant influence on the saccade amplitude (β = 0.71, SE = 0.12, $p = <.001$), meaning that when the visual task was the foraging (search) task, the saccade amplitude was millimetres longer than when the visual task was counting. Moreover, trial did not have a significant influence on saccade amplitude, (β = 0.01, SE = 0.15, $p = .5$), meaning there was not a substantial influence of expertise or fatigue on performance (as discussed earlier under the data analysis section) reflected in saccade amplitude.

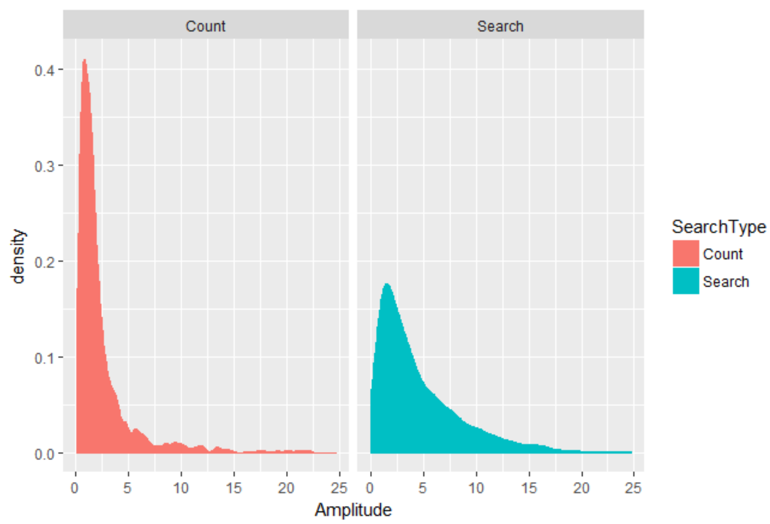
Finally, the interaction effect of visual search type and trial did not have a significant influence on saccade amplitude, (β = -0.01, SE = 0.02, p = .48), meaning the trial number did not have a substantial influence on saccade amplitude depending on which visual task was being performed.

Illustrations









Discussion

The goal of the experiment to analyze and differentiate ocular motion and attention as it relates to top-down and bottom-up processes. Visual foraging patterns are assumed to be linked with bottom-up attentional processing whereas task-structured visual patterns are a result of top-down processing. The results of the fixation duration analysis support our hypothesis that visual search patterns are influenced by visual task structure. The significance of our model indicates that search type is a substantial predictor of fixation duration controlling for the random effects of trial and participant. Moreover, the significance of the parameter search type indicates that search type is also a considerable predictor of saccade amplitude. For future researchers wishing to replicate this experiment, it would be recommended to have more participants. It is unlikely that having six participants is adequate to fully illustrate the effect described.

Social Engagement Experiment

Introduction

In the following report we will discuss the replication of a previously performed study which sought to investigate whether being involved in a social interaction elicits higher levels of emotional engagement. Emotional engagement was hypothesized to be reflected in changes in pupil size, so as the more involved one is, the bigger one’s pupil size will be. Participants were tasked with viewing various videos which demonstrated different aspects of social engagement. Eight videos were used in the experiment. Half of the videos involved a female actor and the other half a male actor. The videos were further separated into trials of high and low ostentation (which for the purpose of this experiment will mean the enthusiasm of the actor) and directedness and non-directedness (which will mean whether the actor looks at the participant or not). Each actor performs each of the trials.

Hypothesis

Conceptual: Emotional engagement is a response to the ostentation and the directedness of an interlocutor.

Operational: The magnitude of pupil dilation will correlate with emotional engagement.

Methods

Data collection

Eye-tracking data of monocular position and pupil size was collected with a head mounted Eye Link 1000. The data was recorded on the setting of 1000Hz (recording data once per every millisecond). The experiment took place in a basement room with the lighting set to be as consistent as possible for each participant. Participants were seated approximately 30 inches in front of a 30-inch flat panel LCD monitor. The eye-tracker was calibrated prior to the experiment by using the in-built nine-point automated calibration procedure. This was repeated until the validation procedure reported an average error lower than 1 and max error lower than 1.5.

Participants

Six participants were part of the experiment (four female, two male), all university students (between age of 20 and 30). They were not required to meet any specific criteria, excluding neither eyeglasses nor cosmetics. One participant was wearing glasses. Average error of calibration among all participants was 0.87.

Data pre-processing

The collected data was automatically pre-processed using the in-built DataViewer software (including velocities, pupil size and x and y coordinates). Fixations, saccades and eye-blinks were identified, artifacts were removed. Previously mentioned validation data was used to exclude participants and model measurement error. The data was high-pass filtered at a 100 cut off to counter calibration drift. Systematic bias on fixation estimates on fixation crosses was estimated and positions were accordingly adjusted at every trial.

Data analysis

The four models used were created to best represent the hypothesis and are as follows:

PupilSize ~ Ostention + Directionality + Trial + (1+Ostention + Directionality | ParticipantID)

PupilSize ~ Ostention * Directionality * Trial + (1+Ostention + Directionality| ParticipantID)

PupilSize ~ Ostention + Directionality + (1+ Ostension + Directionality | ParticipantID)

PupilSize ~ Ostention * Directionality + (1+ Ostension + Directionality | ParticipantID)

The assumption the experiment is based on (and is testing) is that pupil contraction and dilation are products of emotional arousal incited by the ostensiveness of visual stimuli. Therefore, we predict that this phenomenon will be most apparent when the direction of ostentation is either fully towards or fully away from the participant. Thusly, the variables used in the model (ostentation and directionality) are directed related to our hypothesis.

We chose to use mixed effects model in order to control for repeated measures (participants taking part in multiple trials) and to account for random influences and individual differences. Although there is theoretical support for adding time within the trial as a predictor and trial and trial as a random effect, we made the attempt but the model became too complex and did not converge. Eye tracking data is not a normal distribution, but a exponential distribution. We therefore log-transform our data to meet the criteria of a normal distribution for the analysis.

We analysed the data using the statistical software of R studio (RStudio Team, 2015). Before performing a cross validation on the models we observed the *Akaike Information Criterion* (AIC) and the *Bayesian Information Criterion* (BIC) for all the models (shown in table below). Although from these information criteria we would select the three-way interaction model, we will perform a cross validation in order to retain the out-of-sample error that will determine which model is best at prediction and generalisation.

Model	**AIC**	**BIC**
PupilSize ~ Ostension + Directionality + (1 + Ostention + Directionality | Participant)	5813.1	5853.3
PupilSize ~ Ostension * Directionality + (1 + Ostention + Directionality | Participant)	5811.0	5855.2
PupilSize ~ Ostension + Directionality + Trial + (1 + Ostention + Directionality | Participant)	5725.7	5769.9
PupilSize ~ Ostension * Directionality * Trial + (1 + Ostention + Directionality | Participant)	5681.3	5741.5

We performed a 4-fold cross validation with each model on our dataset. As a result, the model with the lowest out-of-sample error was model #1 with the interaction effect of ostentation, directionality, and trial with an average test error (rmse) of 940.68. The second best model was #4 which used the interaction between ostentation and directionality with an average test error (rmse) of 953.61.

	perf_test_list	perf_train_list	temp_model
1	724.7841	1042.4127	glmer(PupilSize~Ostension + Directionality + Trial + (1+Ost...
2	1240.0865	873.1223	glmer(PupilSize~Ostension + Directionality + Trial + (1+Ost...
3	837.6482	1024.9314	glmer(PupilSize~Ostension + Directionality + Trial + (1+Ost...
4	960.2056	1008.4322	glmer(PupilSize~Ostension + Directionality + Trial + (1+Ost...
5	837.6482	1024.9314	glmer(PupilSize~Ostension * Directionality * Trial + (1+Oste...
6	765.0087	1108.0205	glmer(PupilSize~Ostension * Directionality * Trial + (1+Oste...
7	1871.2860	857.9471	glmer(PupilSize~Ostension * Directionality * Trial + (1+Oste...
8	960.2056	1008.4322	glmer(PupilSize~Ostension * Directionality * Trial + (1+Oste...
9	852.5022	1022.0672	glmer(PupilSize~Ostension + Directionality + (1+Ostension ...
10	765.0087	1108.0205	glmer(PupilSize~Ostension + Directionality + (1+Ostension ...
11	1871.2860	857.9471	glmer(PupilSize~Ostension + Directionality + (1+Ostension ...
12	952.9462	1012.0087	glmer(PupilSize~Ostension + Directionality + (1+Ostension ...
13	837.6482	1024.9314	glmer(PupilSize~Ostension * Directionality + (1+Ostension ...
14	1272.6234	859.3714	glmer(PupilSize~Ostension * Directionality + (1+Ostension ...
15	800.4535	1033.7098	glmer(PupilSize~Ostension * Directionality + (1+Ostension ...
16	903.7457	1033.8759	glmer(PupilSize~Ostension * Directionality + (1+Ostension ...

Results

	temp_model	mean_test_rmse
1	glmer(PupilSize~Ostension + Directionality + Trial + (1+Ost...	940.6811
2	glmer(PupilSize~Ostension * Directionality * Trial + (1+Oste...	1108.5371
3	glmer(PupilSize~Ostension + Directionality + (1+Ostension ...	1110.4358
4	glmer(PupilSize~Ostension * Directionality + (1+Ostension ...	953.6177

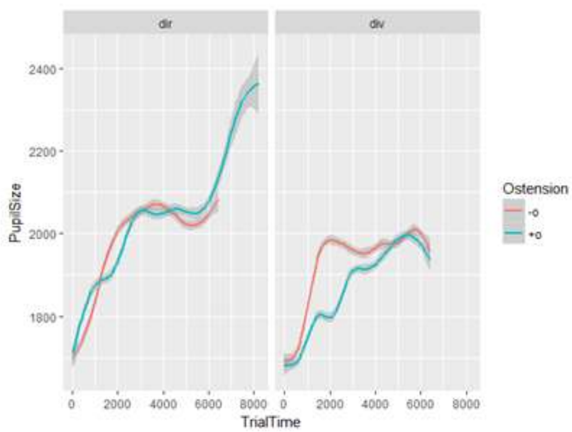
Model #1 had the lowest mean out of sample error out of the four (940.8), so we chose it as best performing model and continued to use it to analyse the full data set.

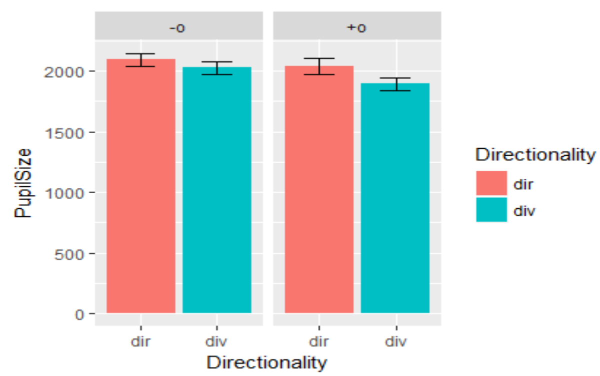
$$\text{Pupil Dilation} = \beta_0 + \beta_1 * \text{Ostentation} + \beta_2 * \text{Directionality} + \beta_3 * \text{Trial}$$

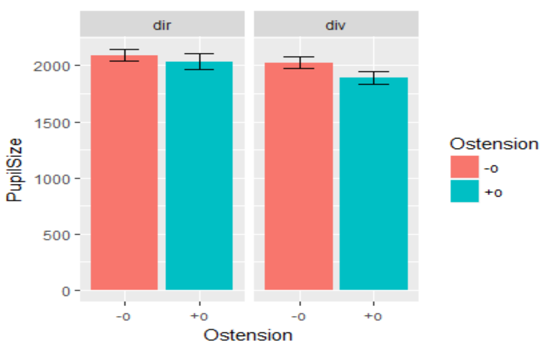
```
Fixed effects:
                          Estimate Std. Error t value Pr(>|z|)    
(Intercept)               7.603322   0.128543   59.15  < 2e-16 ***
OstensivenessOstensive   -0.032176   0.015909   -2.02   0.0431 *  
DIRECTIONTowards you      0.016256   0.020521    0.79   0.4283    
Trial                    -0.018992   0.003347   -5.68 1.39e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

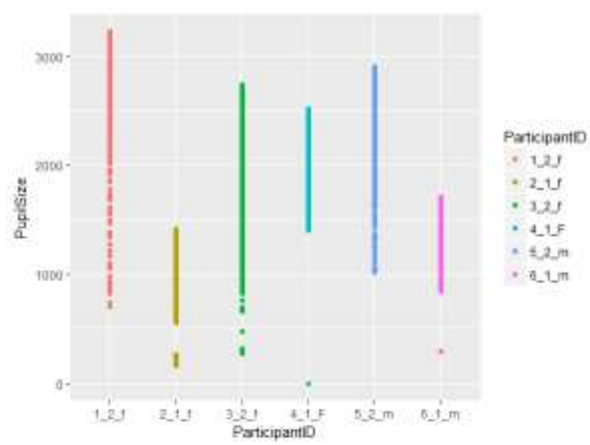
The fixed effects for the best model show significance for ostentation (β = -0.03, SE = 0.02, p = <.05) and trial (β = -0.02, SE = .003, p = <.001). This suggests that ostentation seems to negatively correlate with pupil dilation as well as with trial. There was not, however, a significant influence on account of directionality, (β = 0.02, SE = 0.02, p = >.05).

Illustrations









Discussion

The outcome of the analysis shows that the findings from the experiment are counter to our original hypothesis. The model that performed the best in the cross validation displayed a negative correlation between ostentation and pupil dilation. With pupil size being the measure we're assuming for emotional engagement, this finding is counter to the anticipated positive correlation between pupil size and emotional engagement.

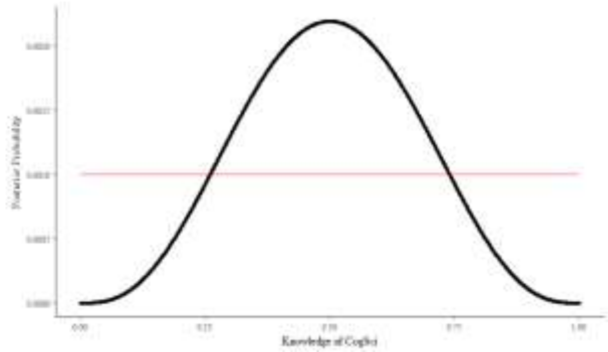
Additionally, the lack of influence seen from directionality of the actor does not support our hypothesis that this will have a significant influence on pupil size. This may be because, among other things, the participants were viewing a video and not engaging in a person in real life. Future researchers can investigate this hypothetical. The main drawback of this experiment is the lack of variation in and number of participants. Future researchers may want to take this into consideration and include a more varied and numerous group of participant data.

Assignment 2.1 (RA, DJ, BZ)

1. What's Riccardo's estimated knowledge of CogSci? What is the probability he knows more than chance (0.5) [try figuring this out. if you can't peek into chapters 3.1 and 3.2 and/or the slides]?

With a uniform prior, the total probability that Riccardo would have performed better than chance is 50% (since he does perform at chance level). This is because our uninformed prior assumes nothing and therefore assigns equal probabilities to all possible outcomes. Therefore, since Riccardo had an outcome of 3/6 (0.5), this is uninfluenced by the prior. However, there is uncertainty around this notion because of i) not a lot of data to go by and, ii) an uninformed prior which both contribute to a relaxed standard deviation and therefore uncertainty in our estimated posterior.

a. **First implement a grid approximation (hint check paragraph 2.4.1!) with a uniform prior, calculate the posterior and plot the results**



b. **Then implement a quadratic approximation (hint check paragraph 2.4.2!).**
   - **N.B. for the rest of the exercise just keep using the grid approximation (we'll move to quadratic approximations in two classes)**

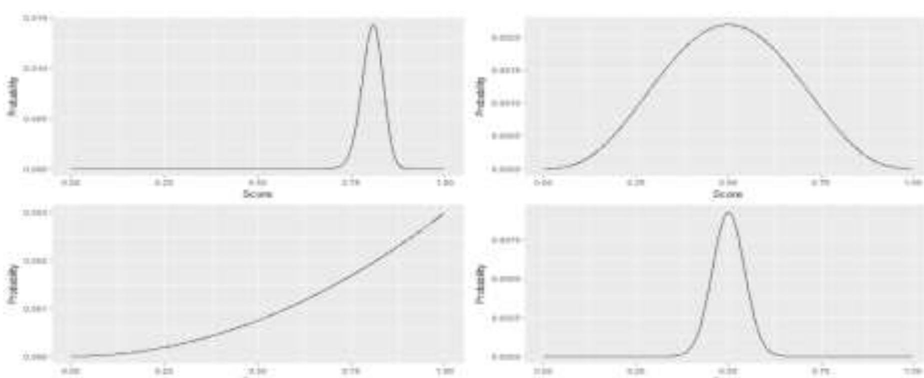
```
  Mean StdDev 5.5% 94.5%
p  0.5    0.2 0.17  0.83
```

2. Estimate all the teachers' knowledge of CogSci. Who's best? Use grid approximation. Comment on the posteriors of Riccardo and Mikkel.

Calculating the posteriors from an uniform prior, it can be deduced that Joshua is the most educated when it comes to Cognitive Science. This is the case because a) He has a lot of available data which makes his 'score' more reliable, and b) his score is high compared to the others. Saying Kristian is the most educated at Cognitive Science would be naïve as we did not collect enough data to have a meaningful comparison to Joshua. We can still estimate a distribution for Kristian, but as seen below, it is less reliable that Joshua's. That being said, it should be noted that it should be considered what the criteria for 'best' is. We are fairly certain Joshua knows something, but indications for Kristian point in the direction that he may be a cognitive science rock-star. 'Who is best' is dependent on the accompanying question, 'what is best for what'.

a. Produce plots of the prior, and posterior for each teacher.

Uninformed Posterior (JS, RF, KT, MW)



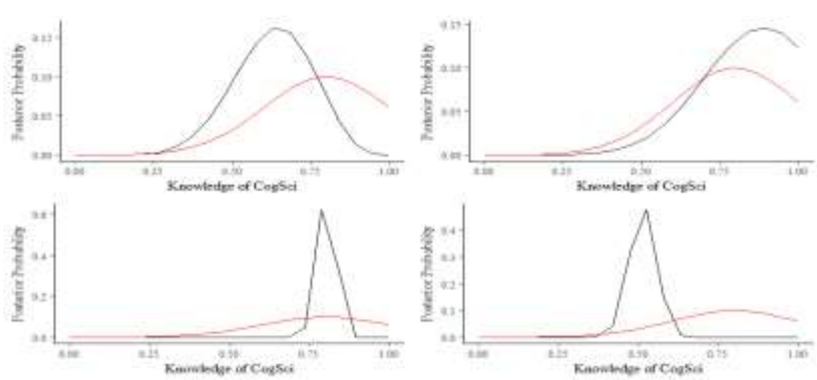
Both Mikkel's and Riccardo's uninformed posterior is around chance (mean 0.5). However, since we have more empirical data from Mikkel (132 answered questions) compared to Riccardo (6 answered questions), Mikkel's standard error is much smaller than Riccardo's. In other words, the probability of Mikkel's knowledge being close to chance is higher than of Riccardo's. We would need more data from Riccardo in order to be more certain about his knowledge.

3. Change the prior. Given your teachers have all CogSci jobs, you should start with a higher appreciation of their knowledge: the prior is a normal distribution with a mean of 0.8 and a standard deviation of 0.2. Do the results change (and if so how)?

There is still a difference in each teacher's posterior. Due to the fact that although we have now added an informed prior, the data is not totally disregarded and therefore still has an influence on the posterior. Because the teachers still have very different results than each other, they will have different posteriors. Although, they will have slightly more similar posteriors in the sense that they will be more inclined to be towards the right side of the plot (high values), since we defined the prior as having a high mean of 0.8 and a standard deviation of 0.2. The more 'weight' we give the prior (smaller standard deviation), the less influence the data will have on the posterior. In other words, if our prior had a mean of 0.8 and a standard deviation of 0.01, and Riccardo had a score which was on a tail of this distribution, we would be less likely to believe that Riccardo's score actually reflects his true knowledge of Cognitive Science.

a. Produce plots of the prior and posterior for each teacher.

Informed Posterior (RF, KT, JS, MW)

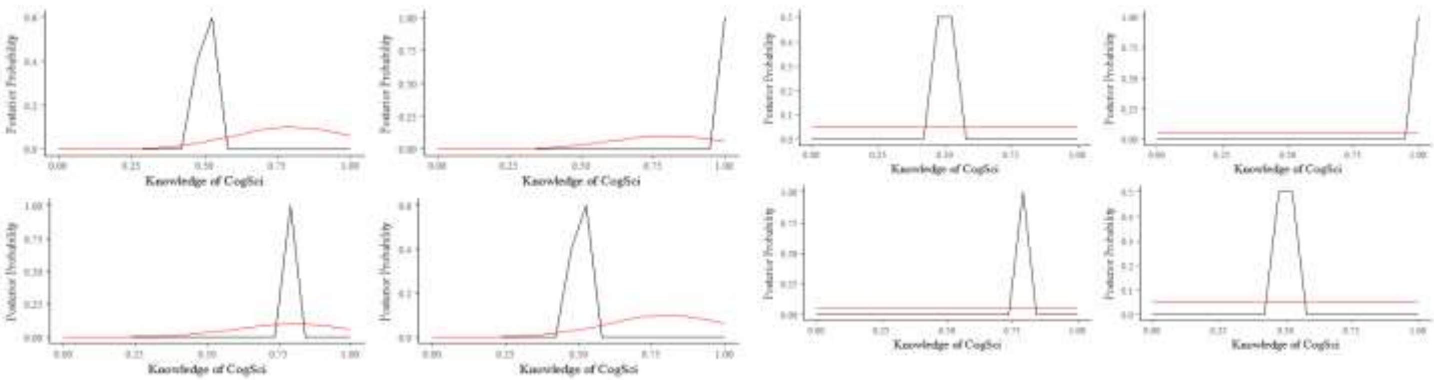


4. You go back to your teachers and collect more data (multiply the previous numbers by 100). Calculate their knowledge with both a uniform prior and a normal prior with a mean of 0.8 and a standard deviation of 0.2. Do you still see a difference between the results? Why?

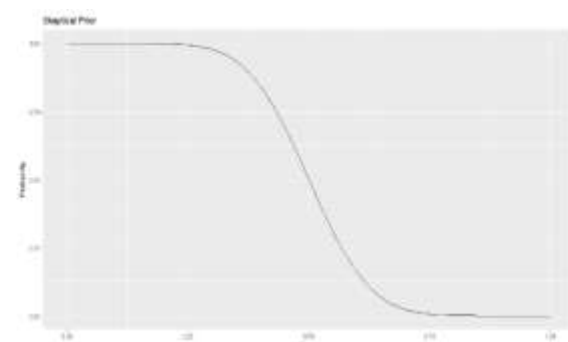
There is not a large difference between the uninformed and informed posteriors with the newly collected data. This occurs because the more data we collect, the smaller the standard error becomes by itself (without help of the prior), and the less influence the informed prior has (since the prior now has a larger standard error than the data).

Additional Data Informed Posterior (RF, KT, JS, MW)

Additional Data Uninformed Posterior (RF, KT, JS, MW)



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?



You would want to change the prior from a normal distribution with a mean of 0.8 and an SD of 0.2 to a distribution in which lower values have high probability and high values have low probability. This is because if we infer the teachers know nothing, our prior belief should be that they are more likely to score poorly than to score well. Additionally, if they know nothing, then any score above chance level will be less probable (since they will be guessing the answers). You can do this by changing the values in the defined variable 'prior' to have a mean of 0.5 (chance level), and an SD of 0.1.

Assignment 2.2 (RA, DJ, BZ)

1. Write a paragraph discussing how assessment of prediction performance is different in Bayesian vs. frequentist models

In prediction performance one could argue that Bayesian models take a more “bottom-up” approach to data analysis. Absconding from the narrow viewpoint of an individual experiment, Bayesian analyses take into account priors from all previously collected data. Thusly, we find that Bayesian models tend to be more generalisable. Frequentism, on the other hand, deals only with data available from the current source (e.g. experiment, study).

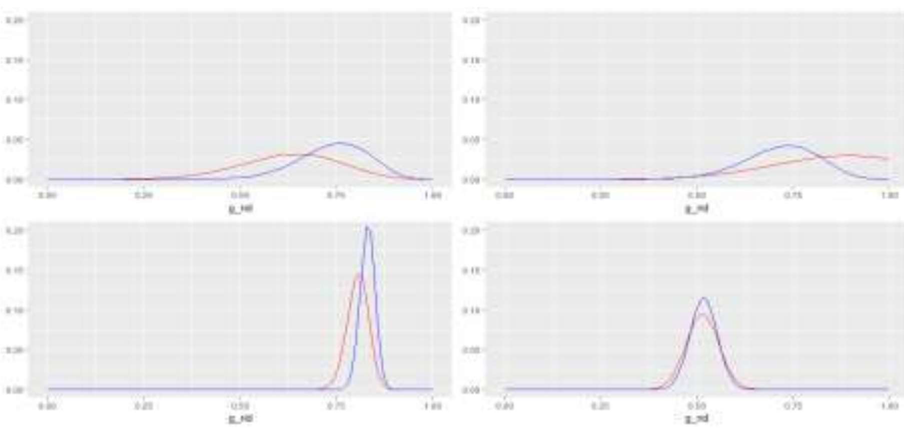
Additionally, since Bayesian statistics deals in the realm of distributions and not absolute values, prediction error considers the comparison of the previously calculated distribution and the posited distribution of the parameter estimation. The difference in these distributions can be assessed qualitatively (visualizing distributions) or quantitatively with a quadratic approximation. Frequentist statistics produces an estimated beta value attached with a probability of that value being the truth. Prediction error, then, is the difference in the previously found value and the predicted value represented as an out-of-sample-error. This can be calculated using root-mean-square-error to assess the error between the two values. This is much easier for the frequentist researcher than for the Bayesian researcher.

Moreover, Bayesian style analyses have the possibility of performing an analysis on very few data. Frequentist analysis has the need for enough data in order to gain enough ‘power’ for a reliable analysis. Bayes, on the other hand, can take any amount of available information at hand and asses it based on prior beliefs, giving a ‘this is the best estimate we can make given what we know right now’ type of answer.

2. Provide at least one plot and one written line discussing prediction errors for each of the teachers.

N.B. The prior used for the old posterior is the informed prior (m = 0.8, sd = 0.2)

Old (red) and New (blue) Posteriors (RF, KT, JS, MW)



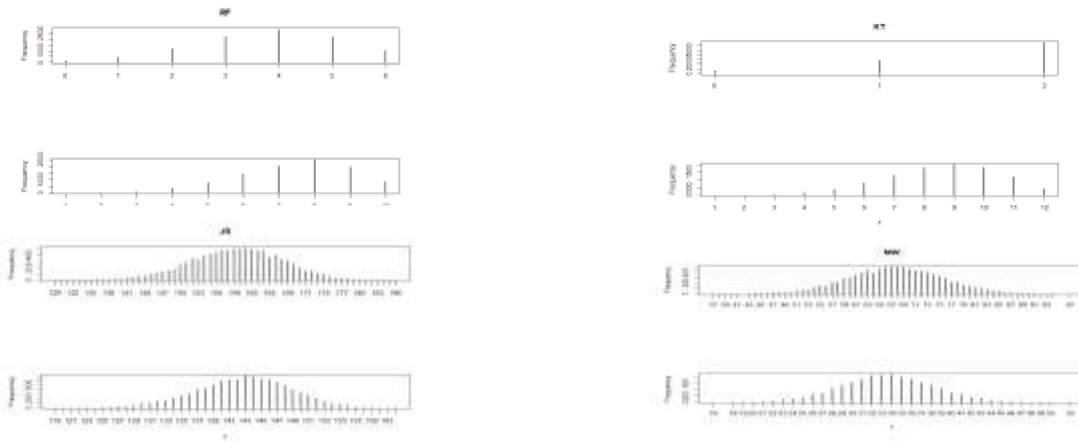
Old Data

Teacher	Correct	Questions
RF	3	6
KT	2	2
JS	160	198
MW	66	132

New Data

Teacher	Correct	Questions
RF	9	10
KT	8	12
JS	148	172
MW	34	65

Old vs. New Predictive Posteriors



Riccardo: Riccardo's new posterior has a smaller standard deviation (which makes sense, since we have collected more data the second round). However, the prior for this new distribution (old posterior) has had an influence on the new posterior and dragged the distribution to lower values. It can also be seen there is a higher distribution on the left tail of the new posterior in order to account for this.

Kristian: The biggest difference in Kristian's posteriors is the probability of him being Cogsci omniscient has gone down dramatically. Since more data has been collected, his distribution also has a smaller standard deviation and has shifted down values on the x-axis. His new posterior resembles more a normal distribution.

Joshua: Josh's posterior has become even more certain, with a smaller standard deviation. Since his last posterior is also very close to his new score, he also has a very high probability for these values (the new posterior includes the 'top half' of the old posterior, giving the mean of the last posterior the same probability and higher probabilities to the rest of the values).

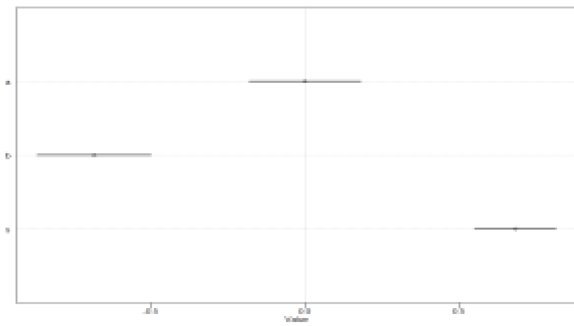
Mikkel: Since we have collected more data about Mikkel, his new posterior has a higher probability, but we are still as almost as uncertain of his true score as we were of his old score (standard deviation hasn't changed much).

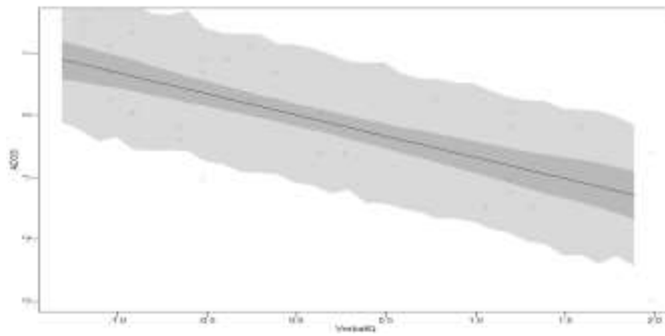
Assignment 3 (RA, DJ, BZ)

1. Assess the relation between symptom severity and IQ (focus on visit 1 and children with ASD) and report the model, a plot of the model, a couple of lines describing the quality of the model and interpreting the results. P.S. Should you scale? P.P.S. Don't forget to motivate your priors. P.P.P.S. At least one plot for results and a plot for quality of each model (here and in the next questions) would be appreciated.

We decided to scale ADOS and well as all our predictor variables (verbal IQ, nonverbal IQ, and social IQ) in order to make them comparable to each other by comparing increments of standard deviations.

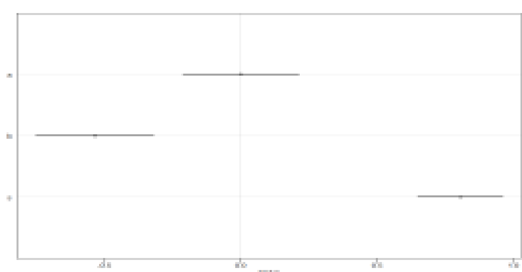
1.1. Verbal IQ and ADOS

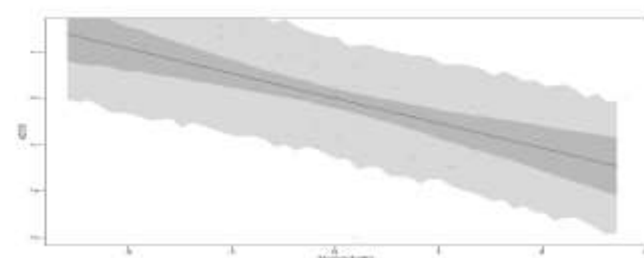




Assuming that ADOS is an accurate measure of symptom severity, we observe a negative correlation between ADOS score and VerbalIQ. Having scaled VerbalIQ, the estimated beta is -0.68 with a standard deviation of 0.12.

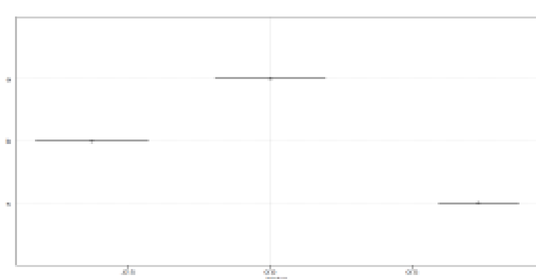
1.2. Non Verbal IQ and ADOS

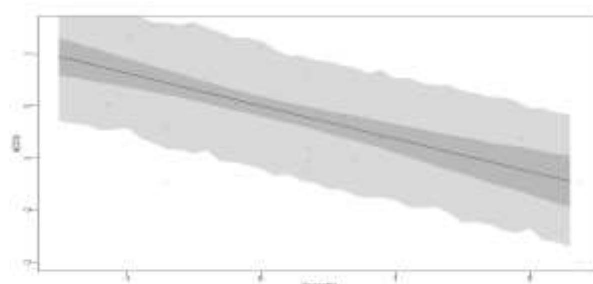




Again, we observe a negative correlation between ADOS score and NonVerbalIQ. Having scaled NonVerbalIQ, the estimated beta is -0.75 with a standard deviation of 0.14.

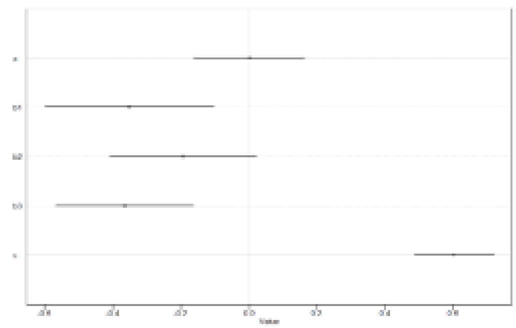
1.3. Social IQ and ADOS





Once more, we observe a negative correlation between ADOS score and SocialIQ. Having scaled SocialIQ, the estimated beta is -0.66 with a standard deviation of 0.13.

2. Do the different aspects of IQ account for different portions of the variance in ADOS?

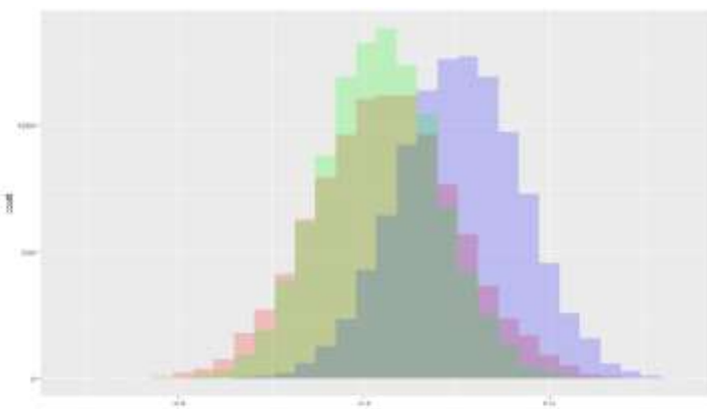


As can be inferred from the graph, it seems that the individual aspects of IQ do not account for different portions of the variance in ADOS score, although we cannot conclude this decisively because there is high uncertainty of the estimates (although all seem to be certainly negative.

2.1. Does it make sense to have all IQ measures in the same model? First write a few lines answering the question and motivating your answer, including a discussion as to what happens when you put all of them in the same model. Then build a model following your answer. If your answer is "no", you are not free, you still have to answer: are there alternative ways of answering the question?

Although individually each variable is negatively correlated with ADOS score, when all in the same model, their explained variance overlaps, making interpretation difficult. The reason to have all three IQ measures in the same model would be due to the assumption that they can each individually account for different phenomena which have an influence on ADOS score; the shared variance seen above casts a shadow on this assumption.

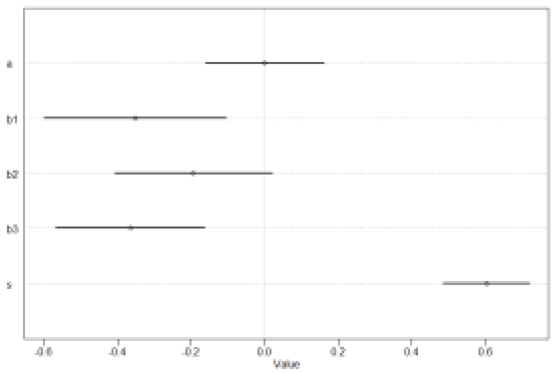
2.2. Build the model, assess its quality, write a few lines interpreting the results.

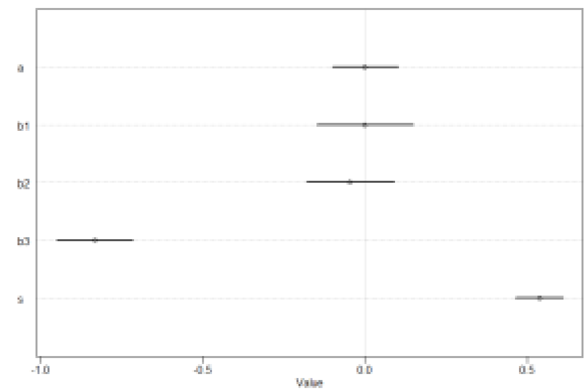


Including all three IQ variables in the model, we yield an alpha of 0 (-0.16, 0.16) and a sigma of 0.6 (0.49, 0.72). The beta estimates include for verbal IQ -0.35 (-0.60, -0.10), non-verbal IQ -0.19 (-0.41, 0.02), and social IQ -0.37 (-0.57, -0.16).

3. Let's now include also the TD children. Does it make sense to ask whether IQ and ADOS are related? Motivate your answer. In any case, if you wanted to build a model to answer that question, which model would you build? Run the model, assess its quality, write a few lines interpreting the results.

From left to right: beta estimates with only ASD children, beta estimates including TD children





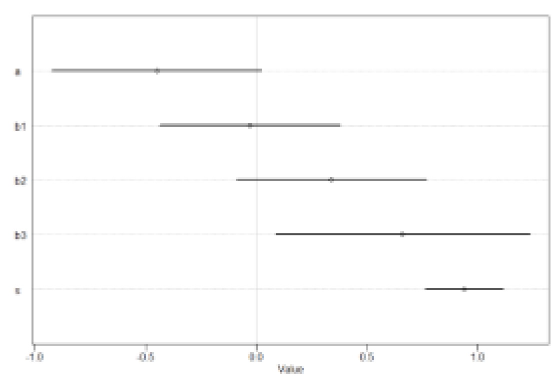
When running an initial study, it would make theoretical sense to include all three different types of IQ measures, as it is sensible to assume they measure and explain different phenomena. Viewing the results, however, we see that verbal and non-verbal IQ overlap in their uncertainty and beta estimates when including typically developing subjects, with social IQ becoming separated from the other estimates. This may be an indication that within the ASD group there is little difference between these three IQs, but comparatively with the TD group social IQ is a strong indicator of ADOS.

4. Let's discuss contents:

4.1. You have three scores for IQ, do they show shared variance?

Is that the same in TD and ASD? What does that tell us about IQ?

See above.



From the above plot, it is seen that social IQ is the only beta estimate which does not overlap with the positive/negative threshold, making it the most reliable of the three. This is the same thing we see in the plot with only ASD participants. This tells us that Social IQ is the only reliable measure of ADOS score in this model, whereas Verbal and Nonverbal IQ share so much variance that it is unclear whether they explain different information.

4.2. You have explored the relation between IQ and ADOS. How do you explain that relation from a cognitive perspective? N.B. You can present alternative hypotheses.

The cognitive systems that seem to pertain to ASD, according to the data provided, are most likely systems related to social learning and development. The overall g factor for intelligence seems to be inconclusive as a TD participant can have low scores in one form of IQ and still not register on the autism spectrum and vice-versa for ASD participants. A causal link is still to be determined between the three IQs and ADOS: does inhibited social learning also hinder verbal and non-verbal IQ (and as an effect, ADOS)? Is low verbal, non-verbal, and social IQ a reflection of a mechanistic blockade of ASD? More experimental research, modelling, and theory need to be developed to further answer these questions.

Assignment 4 (RA, DJ, BZ)

1. What are the consequences of using a meta-analytic prior?

Using a meta-analytic prior allows us to assess the reliability of this study's results in relation to the aggregate of all relevant past findings. This brings us closer to representing the "truth" of the mechanisms involved in Schizophrenia and its relation to pitch. This meta-analytic prior is the same as it would be if each study were conducted successively, using the previous posterior as the following study's prior. Using an uninformed prior each time would be throwing away useful and hard-worked for data that would keep steering us away from the 'true' value, as mentioned above.

2. Evaluate the models with conservative and meta-analytic priors.

The model with a conservative prior yielded an alpha estimate of 0.12 (sd = 0.12), a beta estimate of -0.30 (sd = 0.16), and a sigma estimate of 0.96 (sd = 0.07). The second model with the meta-analytic data has an alpha estimate of 0.01 (sd = 0.13), a beta estimate of -0.04 (sd = 0.20), and a sigma estimate of 0.99 (sd = 0.08).

3. Discuss the effects on estimates. (I'll interpret this as: "Discuss what affects the estimates).

With the conservative prior, we retained a larger effect size (-0.30) than the effect size with the meta-analytic prior (-.04). This is because the meta-analytic prior and the conservative prior both had relatively the same standard deviation, however the analytic prior told the model to estimate the beta estimate closer to 1 than to the conservative prior, which told the model to estimate the beta estimate closer to 0.

4. Discuss the effects on model quality.

The model with the conservative prior had lower WAIC value, however the difference is really small. This implies that the probability of overfitting is smaller for the sceptical prior.

5. Discuss the role that meta-analytic priors should have in scientific practice.

The use of meta-analytic priors (and Bayesian statistical methods in general) should be more widespread in scientific practice, as it is more informative to use already acquired information.

6. Should we systematically use them?
7. Do they have drawbacks?
8. Should we use them to complement more conservative approaches? How does the use of meta-analytic priors you suggest reflect the skeptical and cumulative nature of science?

Systematically using meta-analytic priors would be good in some aspects, as it will combine all previous findings to try to find the 'true' value. However, it would be misleading if all previous studies are biased in the same ways, therefore not allowing to escape the funnel of bad science and bad results. Therefore, uniform priors, or even skeptical prior, should be used from time to time to try to break away from consensus and test current assumptions/results. In a utopian world, rigorous, honest, well-done science could be accompanied by meta-analytic priors systematically. Unfortunately, this is not the case. A third option may be to use a multitude of different priors (with theoretical reasoning) to investigate the impact of previous literature/theoretical background, a researcher's own initiative, and an agnostic approach on findings. If the there is a large difference in posteriors of the three approaches, why is that? More specifically, if the meta-analytic prior and current-study prior are vastly different, what makes the current study different from the rest of the literature? Can this be solved by re-analysing past literature with a skeptical prior (or a prior from the current study)? So many options, so little time. Better get going.