PORTFOLIO ASSIGNMENTS

Computational methods of cognitive science

CHRISTOFFER LUNDBAK OLESEN

GitHub link for assignment 1-4: https://github.com/clolesen/methods4_exam

TABLE OF CONTENTS

Christoffer Lundbak Olesen 25/05/2018

Assignment 1

Visual Search

Methods

Eye tracking methods where used while a participant was doing an experiment on a computer. In the experiment participants looked at ten different pictures one at a time for 45 seconds (each period being a trial). The pictures were taken from “National Geographic’s Photo of the Day” and depicted complex visual scenes with many enumerable objects. All pictures were edited so to contain a small and slightly transparent star symbol somewhere in the scene.

	Mean RMSE	RMSE St.dev.
Model 1	778.23	369.87
Model 2	775.44	370.23
Model 3	773.19	365.20
Model 4	776.45	364.07
Model 5	778.08	361.93

Table 1.1 - Results of model cross validation

The ten trials where divided into two tasks (conditions), a counting task and a visual search task. In the counting task the participants were instructed to count the many enumerable objects. In the visual search task the participants where instructed to find the star. The order in which the two tasks were done, was randomized for each participant.

12 students from Aarhus university participated in the study. They where mixed gender and between 20 and 30 years old. Participants where screened for eye-glasses and eye-makeup, which could affect the quality of the data obtained in the study, but not all were excluded on these factors.

Figure 1.1 - Example of scan path in counting task.

The eye-tracking device used in the study was an Eye Link 1000 recording at 1000Hz. The participant were placed in a head mount 30 inches from the computer screen and the eye-tracker was calibrated using the in-build calibration procedure. The experimental paradigm was programmed in PsychoPy and was running on a second computer, which was synchronized with the Eye link 1000. x/y coordinates, velocities and pupil sizes were recorded and preprocessed by the eye-trackers in-build software, which removed artifacts and identified fixations and saccades.

For the analysis, the preprocessed fixation data was used. Three mixed effects linear regression models were validated using a cross validation technique in order to select the best model. The dependent variable of each model was fixation durations and the predictors were chosen to be task and trial. All models had random slope and intercept per participant in the different tasks. The models are described as model 1, model 2 and model 3.

Model 1: $F = \beta_{0i} + \beta_{1i}C + \epsilon$

Model 2: $F = \beta_{0i} + \beta_{1i}C + \beta_{2i}T + \epsilon$

Model 3: $F = \beta_{0i} + \beta_{1i}C + \beta_{2i}T + \beta_{3i}CT + \epsilon$

In the above models F is fixation duration, C is condition (task) and T is trial. The betas are the coefficients and i is a random component of participant. The epsilon is an error term.

In the cross validation procedure the data was arranged in 3 folds in such a way that no data from the same participant was present in more than one fold. For each test set the root-mean-squared-error (RMSE) was calculated. Then the mean and standard deviation of all RMSE values were calculated and compared across models. Based on these data a model was selected and fitted to the full data.

Results

The results of the cross validation are shown in table 1.1. Since the RMSE only differ in the marginals across models the simplest model was chosen (model 1) on the ground that the extra predictors didn't optimize the model.

When the chosen model was fitted to the full data the predictor was not significant ($\beta = -0.18$, se $= 0.09$, $t = -1.92$, $p > 0.05$)

The distribution of fixation durations in the visual search condition was more centered around short fixations than in the distribution of fixation durations in the count condition, which was more long tailed towards longer fixations (see figure 1.2).

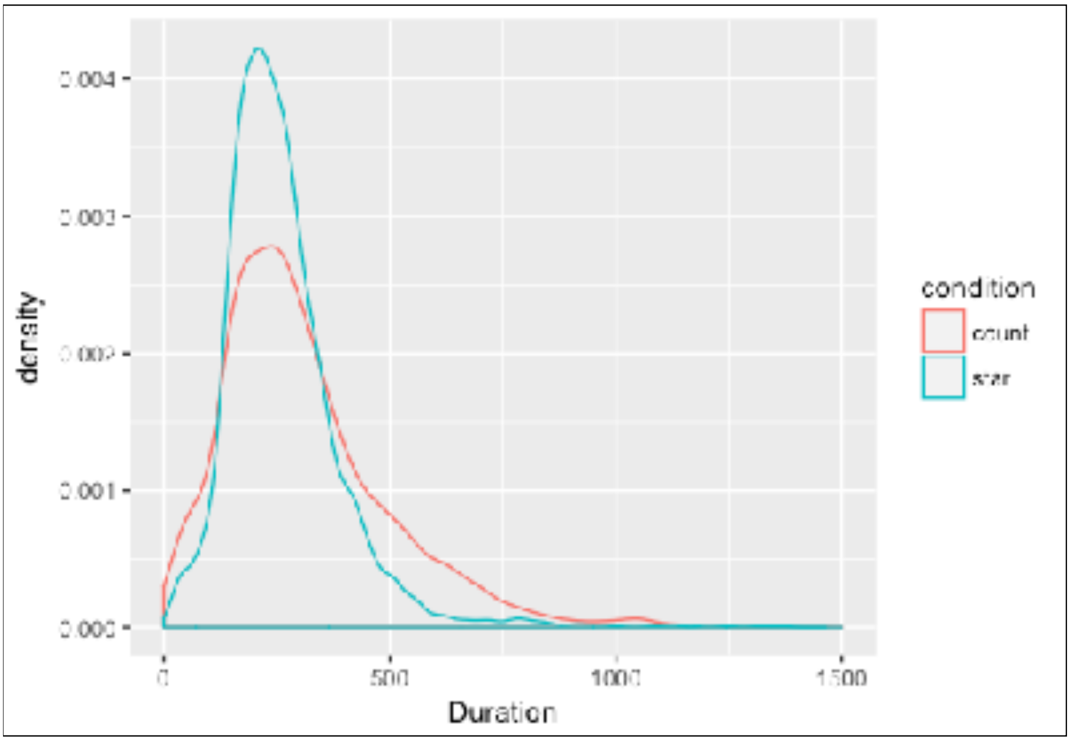


Figure 1.2 - Density plot of the durations of fixations. The blue line represents the visual search condition and the red line represents the count condition. The x axis is cut at 1500ms in favor of visual representation.

Discussion

This study failed to find any significant predictor of the type of tasks involved in the study and the cross-validation results suggested that adding predictors would not improve the model. There was only 12 participants doing 5 trials in each condition and because this is not a sufficient amount of data to draw any conclusions, the present study is not providing any definitive answers,. The distributions of fixation durations reveal a difference between the tasks. The longer fixations in the count condition could be due to cognitive load on counting. The reason the distributions of fixation durations in the visual search condition is more centered around shorter fixations could be an indication of foraging patterns where an small area is searched by many fixations in a short time until a jump is made to another area.

Christoffer Lundbak Olesen 25/05/2018

Social Engagement

Methods

Eye tracking methods where used while a participant was doing an experiment on a computer. In the experiment participants were passively watching videos on the computer screen. There were 32 videos, each with a duration of 5 seconds of an actor sitting at table with a coffee cup in a black room. In all the videos the actor puts forward the cup. The videos varied along three dimensions, namely the gender of the actor, the direction the actor puts forward the cup and the direction of the actors gaze. The last two are the variables being tested in this study. Direction of the action is either directed towards the participant or directed away from the participant and the gaze is like wise either at eye-contact with the participant or down at the table. In the study the variables are called direction (the direction of the coffee cup) and ostention and make a total of four conditions.

12 students from Aarhus university participated in the study. They where mixed gender and between 20 and 30 years old. Participants where screened for eye-glasses and eye-makeup, which could affect the quality of the data obtained in the study, but not all were excluded on these factors.

The eye-tracking device used in the study was an Eye Link 1000 recording at 1000Hz. The participant were placed in a head mount 30 inches from the computer screen and the eye-tracker was calibrated using the in-build calibration procedure. The experimental paradigm was programmed in PsychoPy and was running on a second computer, which was synchronized with the Eye link 1000. x/y coordinates, velocities and pupil sizes were recorded and preprocessed by the eye-trackers in-build software, which removed artifacts and identified fixations and saccades.

For the analysis, the pupil size data was used on the sample level. Five mixed effects linear regression models were validated using a cross validation technique in order to select the best model. The dependent variable of each model was pupil size and the predictors were chosen to be ostention, direction and time. All models had random slope and intercept per participant for ostention and direction variables. Since the distribution of pupil sizes were skewed, this data was log transformed before doing the analysis. The models are described as model 1-5.

Model 1: $P = \beta_{0i} + \epsilon$

Model 2: $P = \beta_{0i} + \beta_{1i}D + \beta_2 T + \epsilon$

Model 3: $P = \beta_{0i} + \beta_{1i}O + \beta_2 T + \epsilon$

Model 4: $P = \beta_{0i} + \beta_{1i}O + \beta_{2i}D + \beta_3 T + \epsilon$

Model 5: $P = \beta_{0i} + \beta_{1i}D + \beta_{2i}O + \beta_{3i}DO + \beta_4 T + \epsilon$

In the above models P is pupil size,D is direction, O is ostention and T is time. The betas are the coefficients and i is a random component of participant. The epsilon is an error term.

In the cross validation procedure the data was arranged in 3 folds in such a way that no data from the same participant was present in more than one fold. For each test set the root-mean-squared-error (RMSE) was calculated. Then the mean and standard deviation of all RMSE values were calculated and compared across models. Based on these data a model was selected and fitted to the full data.

Results

The results of the cross validation are shown in table 1.2. Although the standard deviations of the RMSE values for all the models are very large and the mean RMSE values are very similar, the model with the lowest mean RMSE were picked (model 3). This is also one of the simpler models.

	Mean RMSE	RMSE St.dev.
Model 1	347.54	17.00
Model 2	347.50	17.05
Model 3	347.50	17.02

Table 1.2 - Table showing results from the cross validation.

The model results are shown in table 1.3.
Direction was not a significant predictor of pupil size, but time was a significant predictor of pupil size.

Fixed effects	β	SE	t-value	p-value	Random effect	SD
Direction	67.02	62.35	1.07	$p > 0.5$	**Direction**	152.7
Time	78.32	0.33	236.44	$p < 0.5$		

Table 1.3 - Table showing the results of the model 3

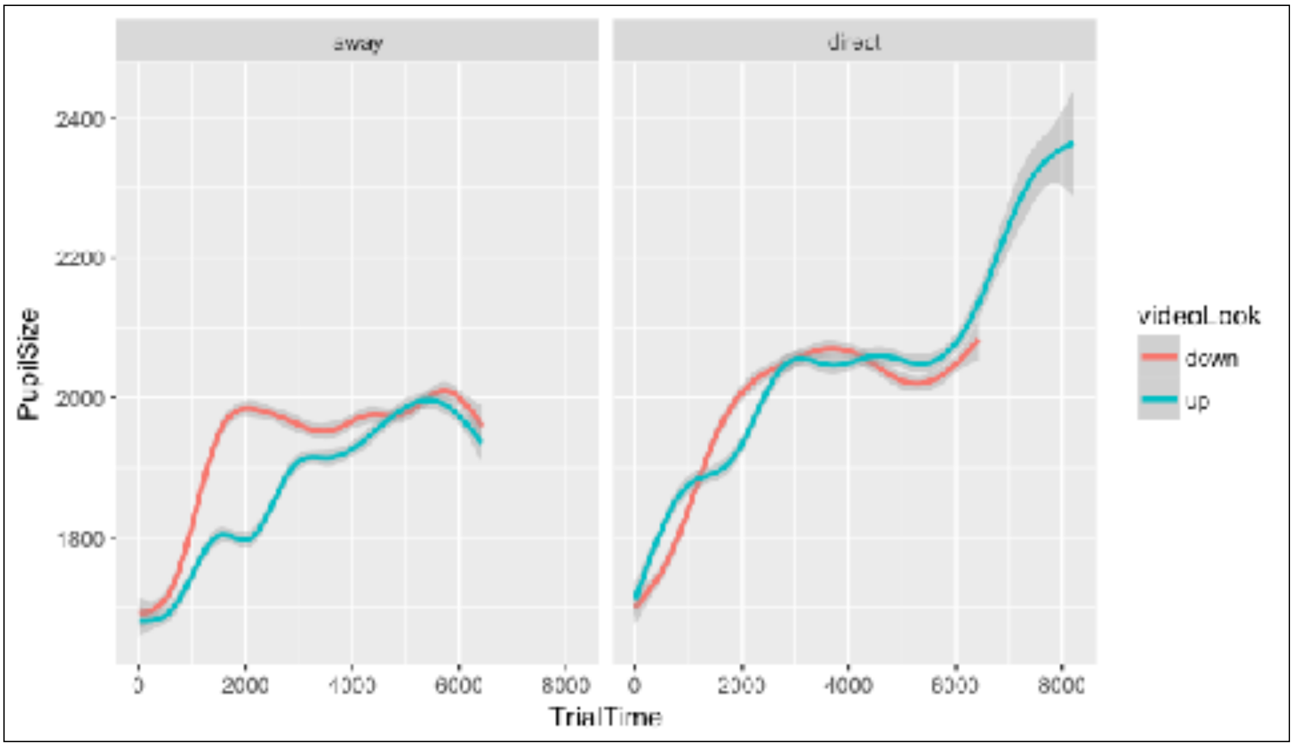


Figure 1.3 - Plots showing the development of pupil size over trials. The left plot is representing all trials were the direction was away from the participant and the right plot represents the trials were the direction was towards the participant. The red line represents the trials were the actor looked down to the table and the blue line represent trials were the actor looked up.

Discussion

Direction was not a significant predictor but time was. When looking at the plots in figure 1.3 this makes much sense. The pupil size is starting vary low on each trial and gets bigger over the trial. this is might have something to do with the eyes adjusting to the video. Both plots looks much the same, except the blue line in the right hand plot. It rises higher than the others in the end. This could be an effect emotional arousal as the actor is looking up at the participant and handing the cup towards the participant, but since ostention was not in the model chosen this is hard to say much about.

Assignment 2

1.1

There is a 50% probability that RF's CogSci knowledge is above chance. This was calculated by summing up all the posterior probabilities above 0.5 in the grid. This also makes perfect sense when looking at the plot in figure 2.1, where the distribution centers around 0.5 and looks normally distributed. The quadratic approximation estimates the mean to be 0.5 with a standard deviation of 0.2.

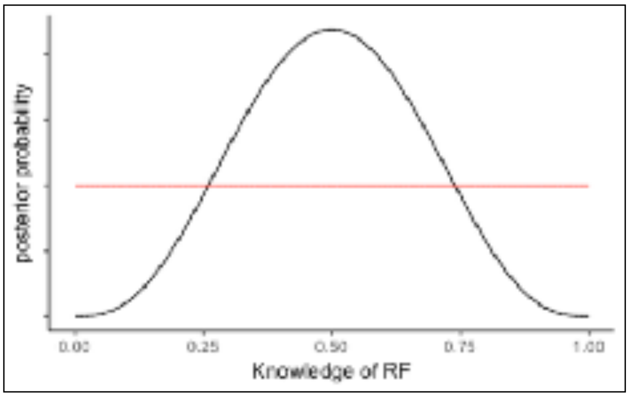


Figure 2.1 - Plot showing the posterior distribution (black line) and the prior (red line) of the model for RF.

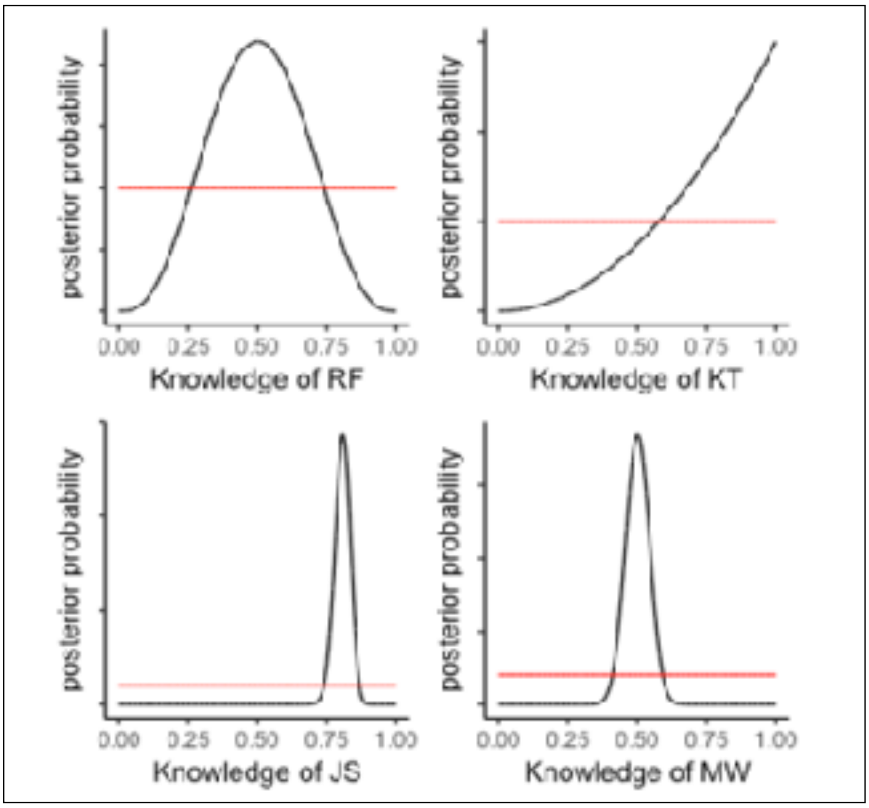
1.2

The percentage of posterior probabilities of a CogSci knowledge above 0.75 was calculated for each teacher and shown in table 2.1 below. The 0.75 threshold was chosen because it is the half way point between random chance and full CogSci knowledge and so this upper part should reflect who has the highest probability of being the most knowledgeable. This means that the model suggests that JS possesses the most CogSci Knowledge.

Teacher	%
RF	7,1
KT	57,8
JS	97
MW	<0,1

Table 2.1 - Percentage of posterior probabilities above 0,75 knowledge per teacher.

Since both RF and MW answered half of the questions right, both of their posterior probability distributions centers around 0.5. The difference is that we have a lot more data on MW than RF, which makes the curve of the distribution narrower. This means that the model is surer that MW's knowledge is around random chance than RF's knowledge, which is also reflected by the numbers in table 2.1.



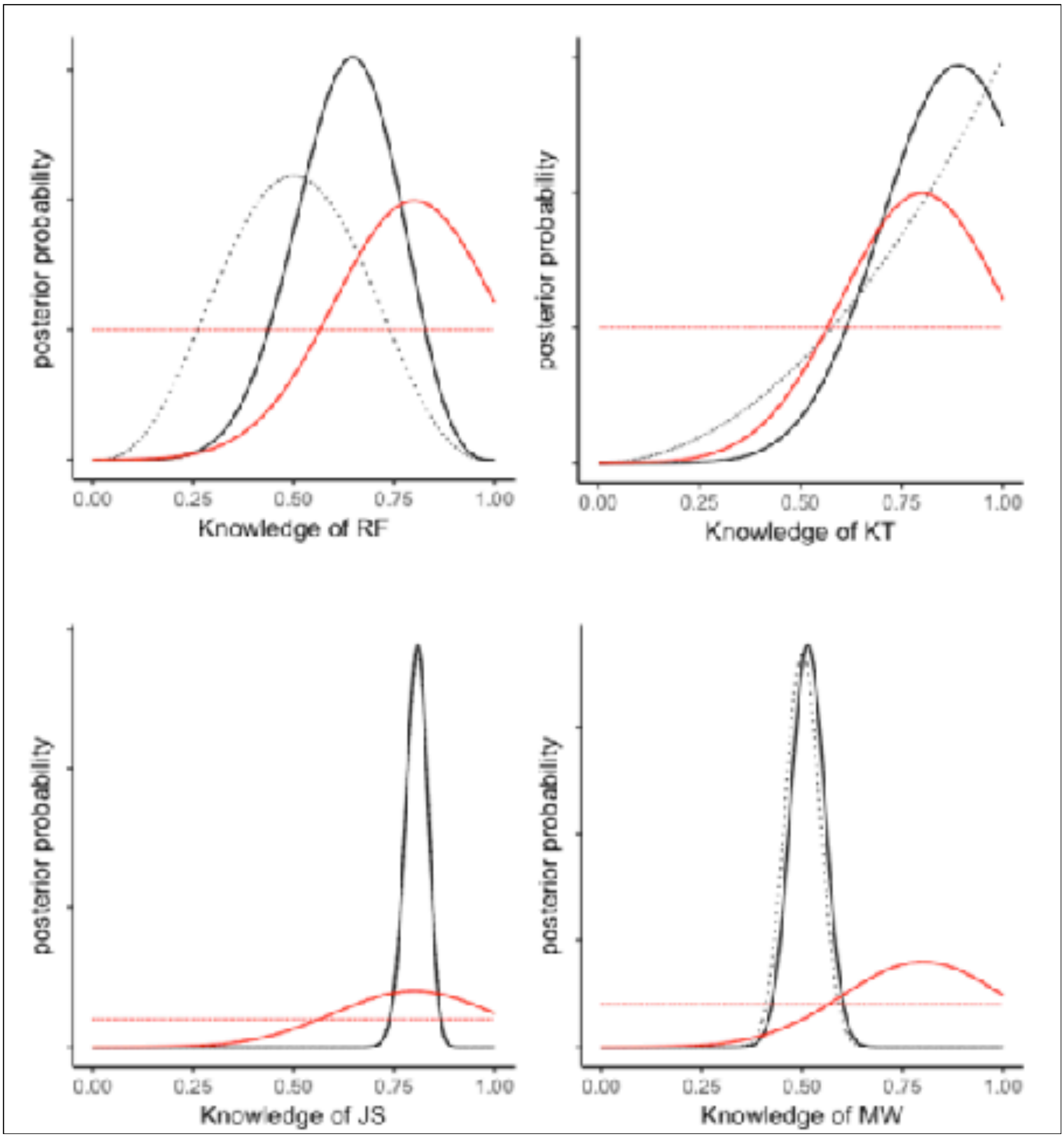
Plot 2.2 - Plot showing the posterior distribution (black line) and the prior (red line) of the model for each teacher.

1.3

The table below (table 2.2) is an updated version of table 2.1. Here the numbers calculated from the model with a non-flat prior is also shown. The new prior affects the results for RF and KT in a positive direction but for the others are barely affected. This is also clearly demonstrated in plot 2.3. The reason for this is that there is considerably less data available for RF and KT and therefore the prior information weights relatively more in these cases than the in the others. The model is then more likely to skew the posterior probability distribution towards the prior.

Teacher	% (flat prior)	% (non-flat prior)
RF	7,1	17,1
KT	57,8	68,4
JS	97	97,1
MW	<0,1	<0,1

Table 2.2 - Percentage of posterior probabilities above 0,75 knowledge per teacher and for models with different priors.



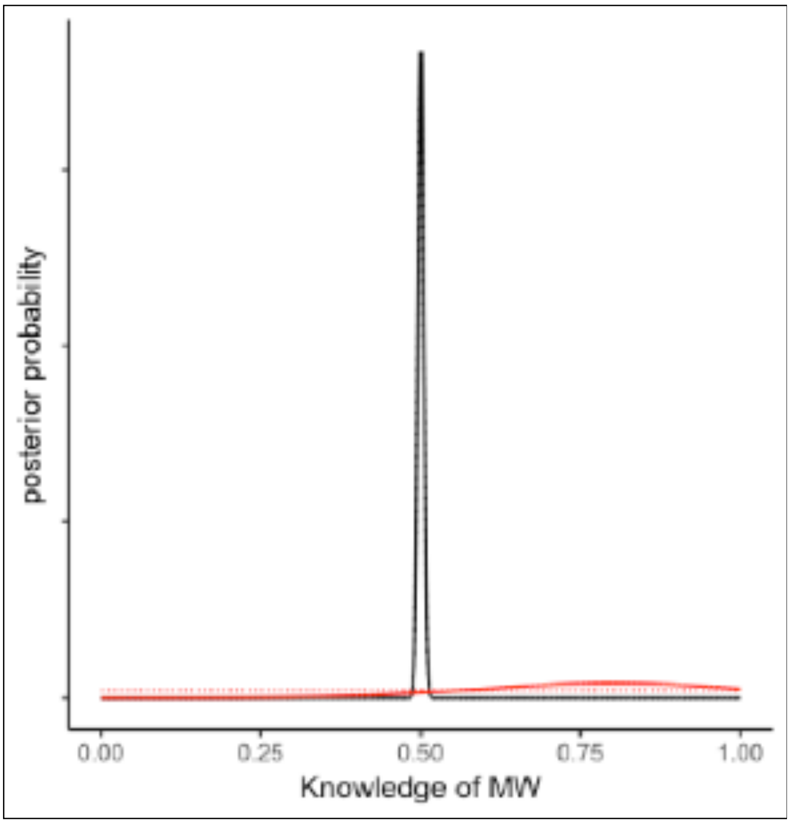
Plot 3 - Plot showing the posterior distribution (black lines) and the prior (red lines) of the models for each teacher. The dotted lines represent the models with a flat prior as also shown in plot 2. The solid lines represent the models with a non-flat prior. Note that in the case of JS the two posterior distributions are so close that they appear to overlap completely in the plot.

1.4

Table 2.3 resembles table 2.2 but the numbers are calculated from the new and bigger set of data. It is quite clear that the results are much more extreme than before and that the priors doesn't have an effect. The posterior has become very narrow with the new data set (demonstration in plot 2.4). Again, the low difference between models with different priors and the narrowness of the posterior probability distribution curves are a result of more data. The model is now much more sure of the teachers respectable levels of knowledge.

Teacher	% (flat prior)	% (non-flat prior)
RF	<0,1	<0,1
KT	100	100
JS	100	100
MW	0	0

Table 2.3 - Percentage of posterior probabilities above 0,75 knowledge per teacher and for models with different priors. These numbers are calculated from a bigger data set.



Plot 2.4 - Plot showing the posterior distribution (black lines) and the prior (red lines) of the models for MV. The dotted lines represent the models with a flat prior as also shown in plot 2. The solid lines represent the models with a non-flat prior. This plot is made from a bigger data set.

1.5

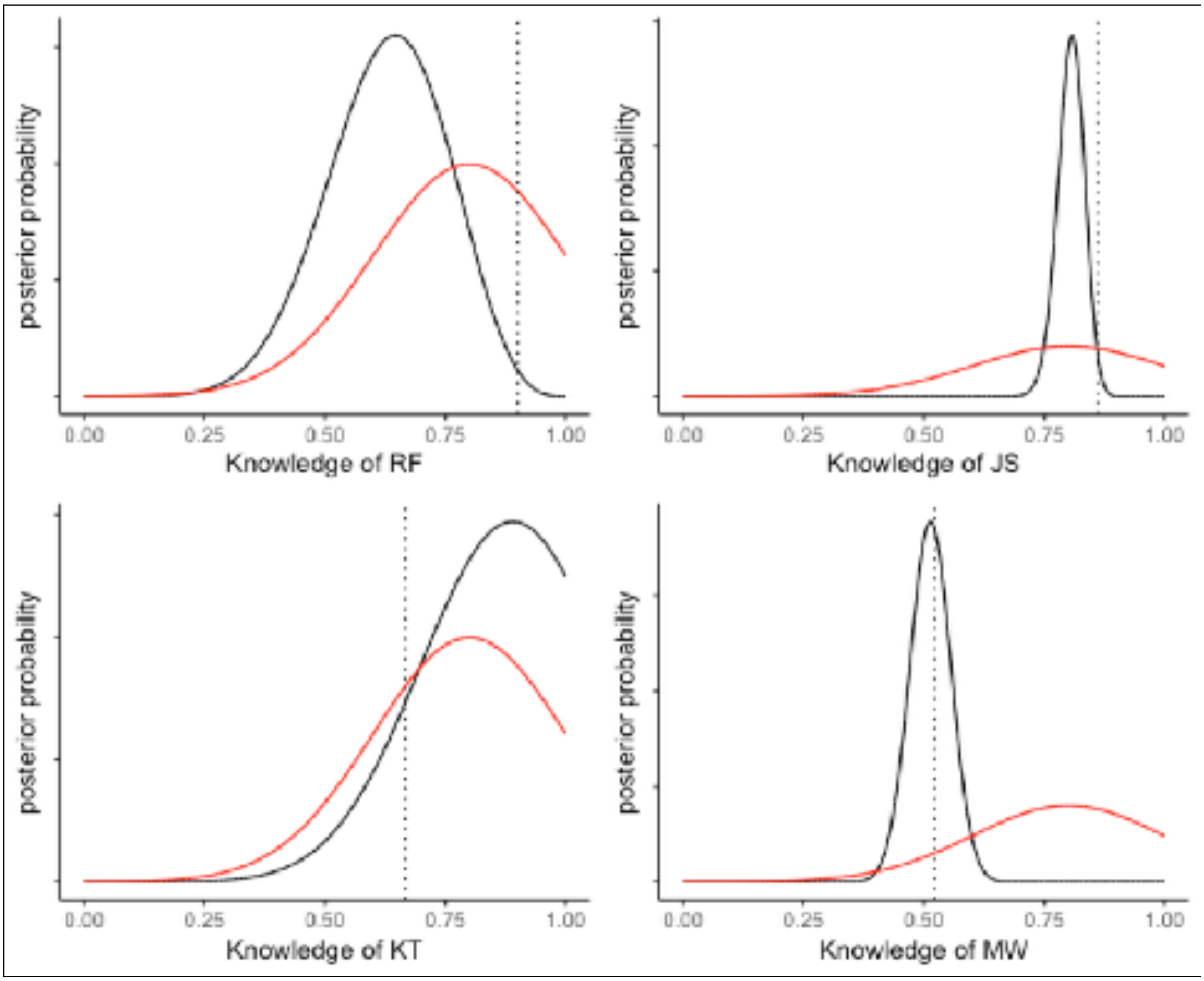
One way is to make the prior distribution a normal distribution with a mean of 0.5. This would reflect the belief that the teachers are at chance level and thus they must answer more questions right in order to convince us that they are worthy CogSci's.

2.1

In frequentist statistics assessment is based upon point estimates like root mean square error (RMSE), which is the average error between the model prediction and the observed data. These estimates can be assessed using model validation techniques like cross validation. This way it is relatively easy to compare model predictions, since they are usually represented by single numeric values.
In the Bayesian framework, models are assessed by comparing the posterior probability distribution to the observed data. This makes the job of comparing model predictions less straight forward than it can be in the frequentist tradition. On the other hand, the Bayesian model prediction is also less arbitrary to interpret since it captures the full uncertainty of the model.

2.2

In plot 5 the models with a Gaussian prior is shown for each teacher. They are the same as those also shown in plot 2.3, but here there is a vertical dotted line representing newly collected data. For RF, KT and JS the model did not predict the new observations very well, since there is not much space under the posterior curve at the point of the observation. In the case of KT the model seems to do a lot better than RF or JS, but it is worth considering that the distribution for KT is very wide due to low sample size and one should therefor be skeptical when assessing anything from this. Common for all three is that the new observation isn't to far of the prior distribution, which suggest that the teachers might actually meet our expectations of them. For MW it is another story. The new observation for MW is close to the predictive posterior mean and far from the mean of the prior. This suggests that MW does not meet our expectations of how knowledgeable a teacher of cognitive science should be but instead strengthens the notion that MW only perform at the level of chance.



Plot 2.5 - Plot showing the posterior distribution (black line) and the prior (red line) of the model for each teacher. The dotted line represents the new observation (number of correct answers divided by number of questions).

Assignment 3

Before any modeling was done, a subset was made of the data set containing only the data collected at visit one. Then the three different IQ measures was z-scored in order to make them fit the same scale. For the first and second part only data from kids with ASD was used.

3.1

Three models were made, one for each type of IQ. Besides the IQ variable the models are the same and are summarized as model 3.1. In the models ADOS is described as a normal distribution, with a mean (mu) and a standard deviation (sigma), where the mean is a linear function of IQ score. The prior for the intercept (alpha) was chosen to be a very broad normal distribution. The prior for the beta was chosen to be only weakly regularizing because it might easily dominate the parameter too much, due to the lack of data (only 34 observations). A flat prior between 0 and 30 was chosen for sigma.

$$A_i \sim \text{Normal}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta Q_i$$
$$\alpha \sim \text{Normal}(0,10)$$
$$\beta \sim \text{Normal}(0,2)$$
$$\sigma \sim \text{Uniform}(0,30)$$

Model 3.1 - A is ADOS and Q is the IQ score

Verbal IQ

	Mean	StdDev	5.5%	94.5%
α	13.81	0.54	12.94	14.68
β	-3.12	0.54	-3.97	-2.26
σ	3.17	0.39	2.55	3.79

Non-Verbal IQ

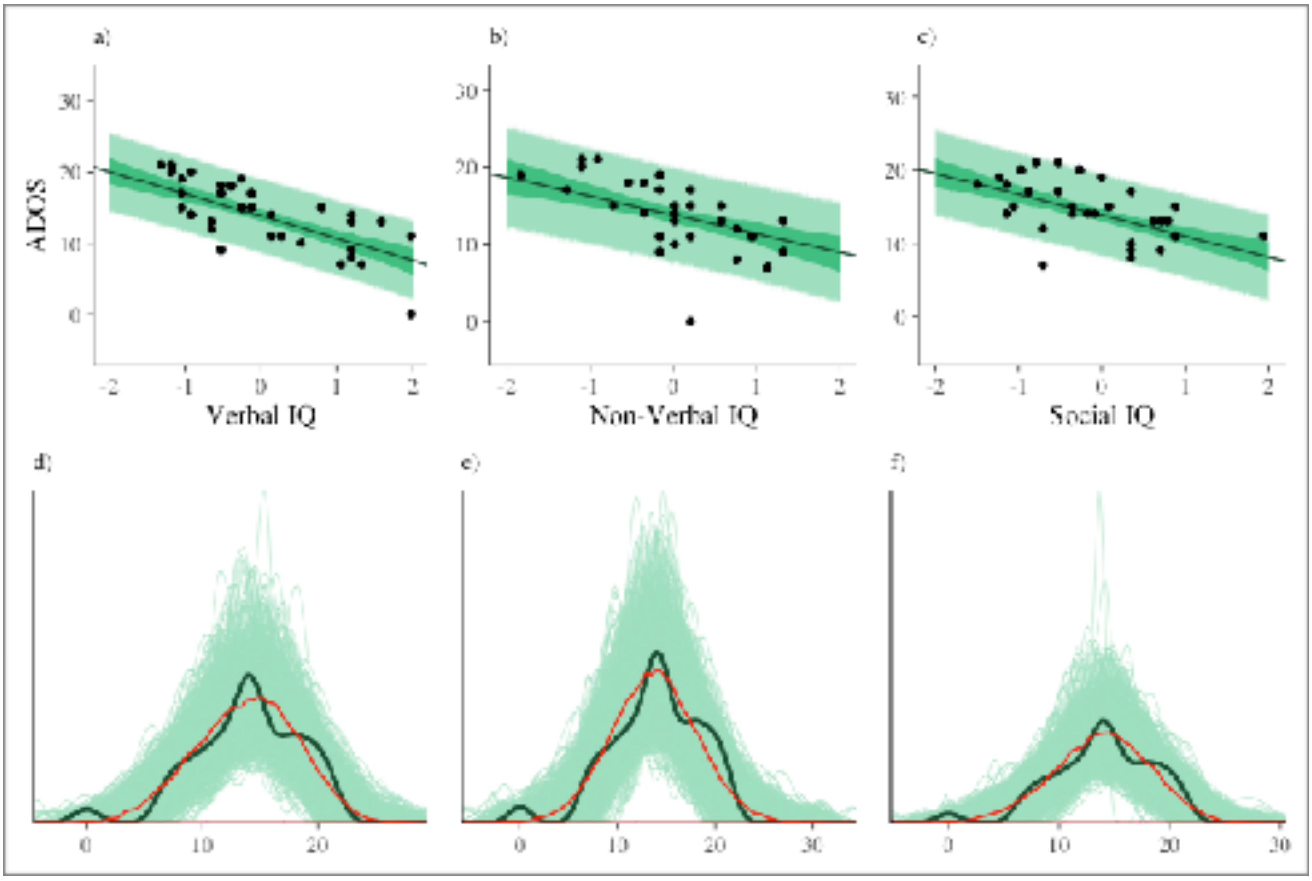
	Mean	StdDev	5.5%	94.5%
α	13.8	0.64	12.77	14.82
β	-2.43	0.62	-3.42	-1.43
σ	3.74	0.46	3.01	4.47

Social IQ

	Mean	StdDev	5.5%	94.5%
α	13.81	0.58	12.87	14.74
β	-2.86	0.57	-3.77	-1.95
σ	3.41	0.42	2.74	4.07

Table 3.1 - table showing parameter estimations of the three models.

In table 3.1 the parameter estimates of the different models are shown along with the standard deviation and the lower (5,5%) and upper (94.5%) boundaries of the 89th percentile interval of the parameter distributions. From this table it is evident that the parameter estimations are quite similar across the models. IQ has a negative effect in all the models and this is also evident from plot a, b and c in plot 3.1 below. The largest effect is found in the model with verbal IQ and the smallest in the model with non-verbal IQ. The later is also the model with most uncertainty, but in general non of the models have too much uncertainty and non of the betas 89th percentile interval overlaps with zero. The density plots in plot 3.1, shows that the posterior is describing the data quite well, but with some uncertainty.



Plot 3.1 - a, b and c are plots where the observed IQ measures are plotted against ADOS, one for each type of IQ test. The dark line is the model fit. The dark green shade is the highest posterior density interval of the simulated mean values. The light green shade is the 89th percentile interval of data simulated from the model posterior distribution. The plots d, e and f are density plots showing the density of observed ADOS values (dark line), the combined density of 1000 simulations from the model posterior distribution (red line) and the 1000 individual densities of those simulations (light green).

3.2

To put all the different IQ measures in the same model, would be to assume that the measures doesn't hold the same information. If one assumes that one general intelligence is the driver behind the three measures, either one of them would be sufficient as a measure of intelligence (assuming that the tests are equally well designed), and thereby that the measures hold the same information. From the similarity of the previous models it seems that this could be the case, but this would mean that all the IQ measures are highly correlated. When inspecting these correlations, the correlation coefficient of non-verbal and social IQ (0.37) are much lower than the others (0.67 and 0.61), which are not so different. This could suggest that such a general intelligence is not the case here, but also that these test doesn't reflect three distinct components of intelligence, since verbal IQ is highly correlated with both of the other IQ measures, which doesn't seem to be driven by the same underlying mechanism. To this end, it would make sense to put at least the two less correlated IQ measures into the model, since these seem to hold different information. However, it is hard to say if the last IQ measure (verbal IQ) would contribute with any new information or not, but one way to test that is to include it in the model.

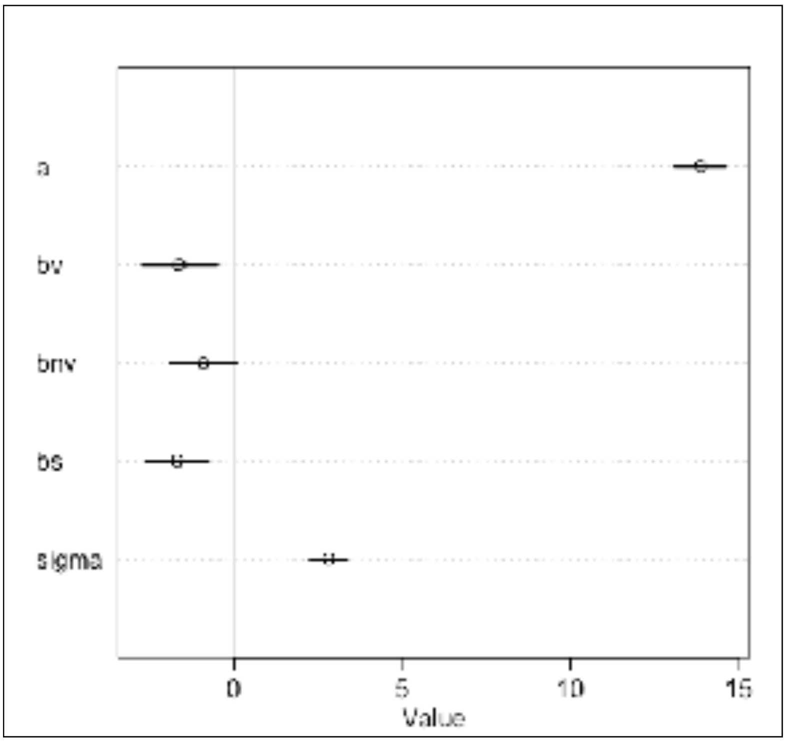
$$A_i \sim \text{Normal}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta_v v_i + \beta_n n_i + \beta_s s_i$$
$$\alpha \sim \text{Normal}(0{,}10)$$
$$\beta_v \sim \text{Normal}(0{,}2)$$
$$\beta_n \sim \text{Normal}(0{,}2)$$
$$\beta_s \sim \text{Normal}(0{,}2)$$
$$\sigma \sim \text{Uniform}(0{,}30)$$

Model 3.2 - A is ADOS, v is verbal IQ, n is non-verbal IQ and s is social IQ

A model similar to model 3.1 was made, but with all the IQ measures as predictors (model 3.2). The priors remained the same. The resulting parameter estimations are shown in table 3.2 and visualized in plot 3.2. In this model there is a negative effect for all beta values, but it is smaller than in any of the single predictor models above. This is not surprising since all the predictors where somewhat correlated and therefore likely to explain some of the same variance. Each of the beta values reflect how much more we can gain by adding the predictor ones we know the other predictors and so the effect gets smaller when they share some variance. The uncertainty around the means are not much different than in the single predictor models, but the 89th percentile interval for the beta of non-verbal IQ is just overlapping with zero, and it is also the beta with the weakest effect of them all. The beta of social IQ has the largest effect on the model and is also the beta with least uncertainty. The effect of verbal IQ are similar to that of social IQ but is more uncertain. This is more surprising since the correlations between the variables suggests that verbal IQ should contribute with less information once all other variables are known, compared to the other IQ measures, due to the stronger correlation and thereby also more potential shared variance between verbal IQ and the other IQ measures. One interpretation of this is that the severity of the ASD kids symptoms are more closely linked to a lower score on social IQ, than the other types of IQ.

	Mean	StdDev	5.5%	94.5%
α	13.85	0.48	13.08	14.62
βv	-1.62	0.7	-2.75	-0.49
βn	-0.91	0.62	-1.89	0.08
βs	-1.68	0.58	-2.61	-0.75
σ	2.8	0.34	2.26	3.35

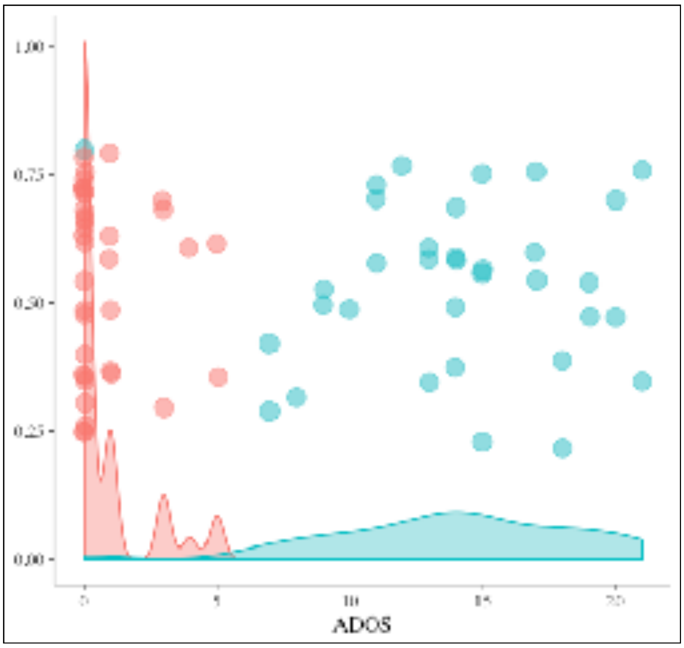
Table 3.2 - table showing parameter estimations of model 3.2.



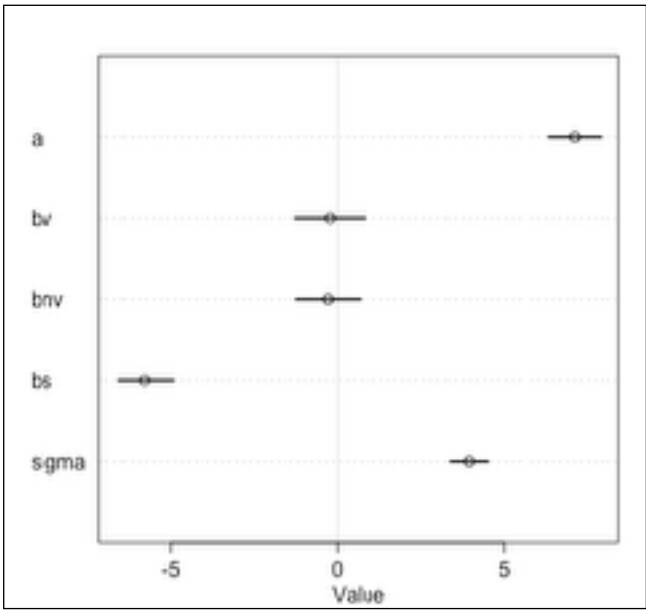
Plot 3.2 - Plot of the model estimates shown in table 3.2

3.3

When including the typical developing children in the data, asking questions about the degree of ADOS becomes less meaningful since almost all typical developing children have an ADOS score of 0 (see plot 3.3). A model that predicts ADOS from different IQ scores will then most likely just tell us something about how kids with ASD are different from kids without the diagnosis, because if the typical developing kids are generally better or worse at any of the IQ tests the beta of that predictor will pull the slope of the model fit towards the respective mean IQ score of those kids around an ADOS of 0. And so in this case, the beta should not be directly interpreted as how intelligence affect the severity degree of symptoms but how kids with symptoms are different from those without. Inspecting the differences in mean IQ scores across diagnosis reveals that the difference in social IQ (1.55) is much larger than the difference in verbal IQ (0.28) and non-verbal IQ (0.2). When fitting model 3.2 on the data including typical developing kids, the beta for social IQ is -5.76 and the other betas are very close to zero with the 89th percentile interval hugely overlapping with zero (see plot 3.4). Including diagnosis as a predictor did not change the picture much, although it has a positive effect on ADOS, but this should be evident from plot 3.3 alone. The model is, for the reasons mentioned above, indirectly predicting diagnosis and it turns out that out of the three types of IQ, social IQ is the most valuable predictor for this job.



Plot 3.3 - Density plot of ADOS where red represents the typical developing kids and blue represents the kids with ASD. The dots each represent one kid on the ADOS scale. The dots position on the y-axis is meaningless and are only scattered along this axis for visual effect. The dots underline the point that most typical developing kids have an ADOS score of 0 and helps to illustrate the lack of overlap across diagnosis.



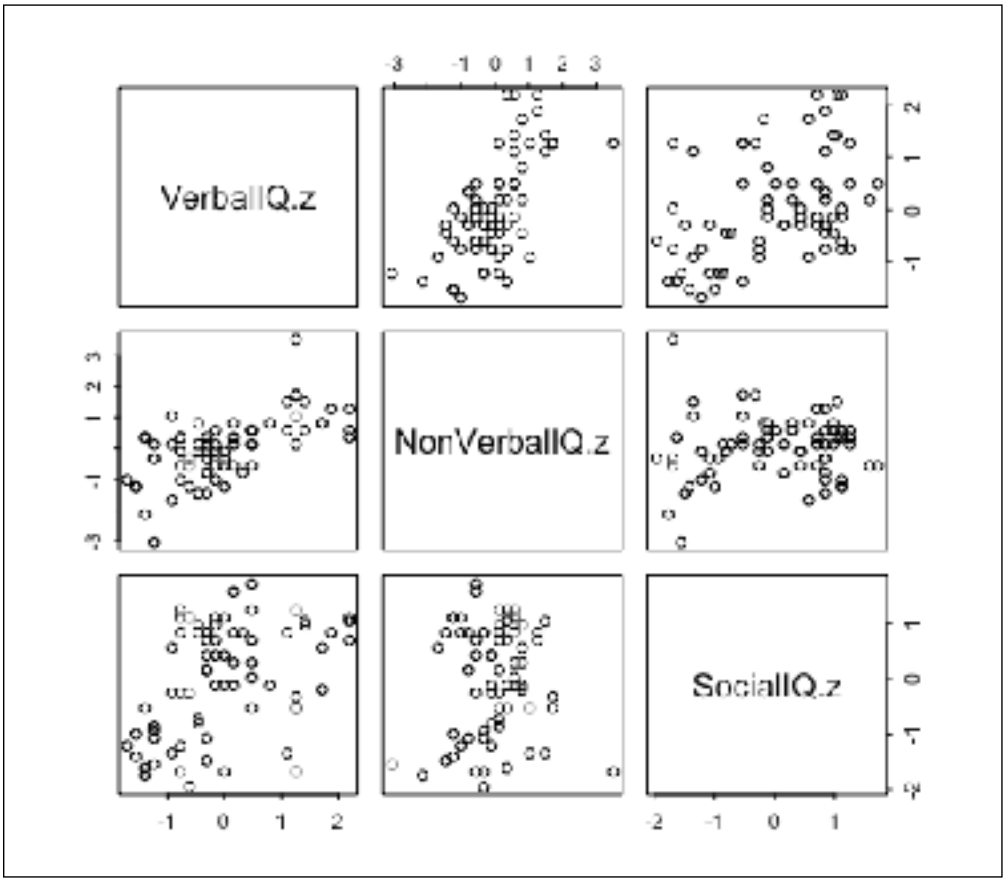
Plot 3.4 - Plot showing the estimates of model 3.2 including the typical developing kids.

3.4

As discussed earlier the IQ measures for the ASD kids are correlated. In table 3.3 the correlation coefficients for kids with and without ASD and all kids are shown. When including the typical developing kids the correlation between social and verbal IQ is weekend and the correlation between social and non-verbal IQ disappears (illustrated in plot 3.5). This is most likely because only very weak, and thereby questionable, correlations are found for these pairs in the typical developing kids and suggests that social IQ is driven by a different underling component than the other IQ measures when considering typical developing kids, but not fully when considering kids with ASD. One interpretation of this could be that one of the symptoms for ASD is an impairment of social intelligence and so the ASD kids rely more on other types of intelligence when doing the social IQ tests. This hypothesis is inline with the model results above and also explains the correlations between IQ measures.

	ASD	TD	All
Non-verbal, Verbal	0.68	0.56	0.62
Social, Non-verbal	0.37	-0.12	0.07
Social, Verbal	0.61	0.17	0.41

Table 3.3 - table of correlation coefficients.



Plot 3.5 - A pairs plot. The observed values of the different IQ measures plotted against each other in pairs. Plotted from the data containing all kids.

Assignment 4

To asses the effect of pitch in schizophrenia two models were made and compared on model contained conservative priors and another contained priors obtained from meta analytical effect sizes of previous studies. A model was made to estimate this meta analytical effect size and because of variation across studies the model included a random intercept per study. The effect size was -0.55 with a standard error of 0.24. All estimates are shown in table 4.1 and visualized in figure 4.1.

	Estimate	St. Dev.	L-95% CI	U-95% CI
Intercept	-0.55	0.24	-1.03	-0.07
sd (Intercept)	0.71	0.22	0.39	1.24

Table 4.1 - Estimates of the meta analytical model

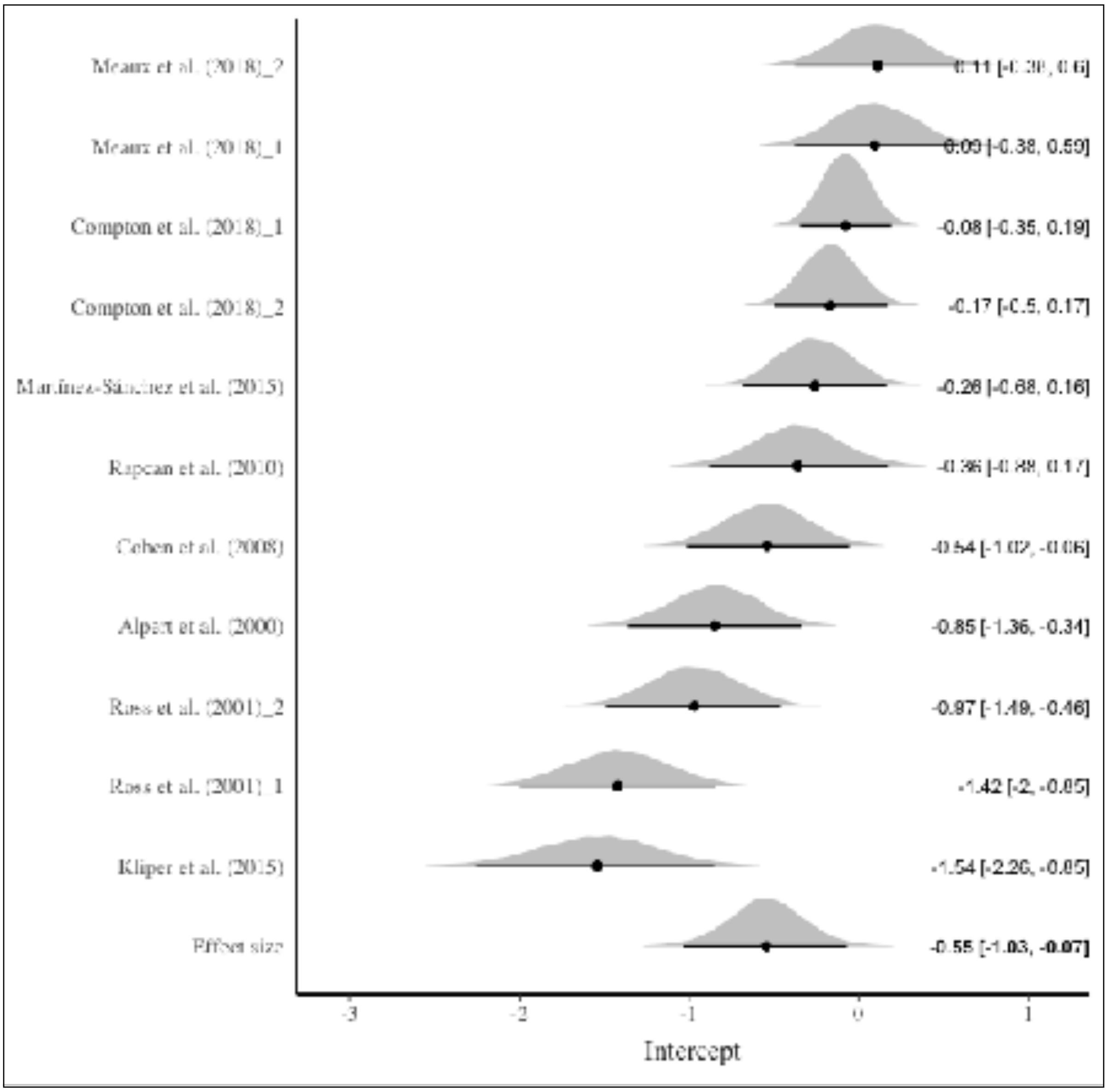


Figure 4.1 - Forest plot showing the model estimates and the effect sizes of individual studies

The two models that was compared are described as model 1 and model 2. The only difference between them was that Model 1 had a skeptical prior for the beta parameter and Model 2 had a prior for the beta parameter reflecting the estimates from the meta analytical model. This later prior was a normal distribution with a mean derived from the estimated effect size of the meta analysis and a standard deviation derived from the standard deviation of the model intercept, which is the estimated standard error of the effect of diagnosis on pitch in schizophrenia since the intercept is a distribution of mean effect sizes. When estimating the parameters of the models, a MCMC technique was used running 4 chains. All models fully converged and had more than a sufficient amount effective samples. The models were compared using WAIC measures and at this stage null model was included. In addition predictive posterior plots were made to further asses the models.

Model 1	**Model 2**
$p_i \sim \text{Normal}(\mu_i, \sigma)$	$p_i \sim \text{Normal}(\mu_i, \sigma)$
$\mu_i = \alpha + \beta d_i$	$\mu_i = \alpha + \beta d_i$
$\alpha \sim \text{Normal}(0{,}1)$	$\alpha \sim \text{Normal}(0{,}1)$
$\beta \sim \text{Normal}(0.1)$	$\beta \sim \text{Normal}(-0.55{,}0.24)$
$\sigma \sim \text{Cauchy}(0{,}2)$	$\sigma \sim \text{Cauchy}(0{,}2)$

Table 4.2 - Table containing model descriptions of model 1 and 2, where p is standard deviations of pitch and d is diagnosis.

For model 1 an effect size of diagnosis was estimated to be -0.08 with a standard deviation of 0.09. For model 2 an effect size of diagnosis was estimated to be -0.38 with a standard deviation of 0.14. All estimates are shown in table 4.3 and visualized in figure 4.2 and 4.3. The WAIC values are shown in table 4.4 and the predictive posterior plots are shown in figure 4.5.

	Estimate	**St. Dev.**	**L-95% CI**	**U-95% CI**
Model 1				
α	0.04	0.09	-0.14	0.22
β	-0.08	0.09	-0.25	0.08
σ	1.00	0.06	0.90	1.13
Model 2				
α	0.19	0.11	-0.01	0.40
β	-0.38	0.14	-0.65	-0.12
σ	1.00	0.06	0.89	1.12
Null model				
α	-0.00	0.08	-0.16	0.16
σ	1.01	0.06	0.90	1.13

Table 4.3 - Table of model estimates.

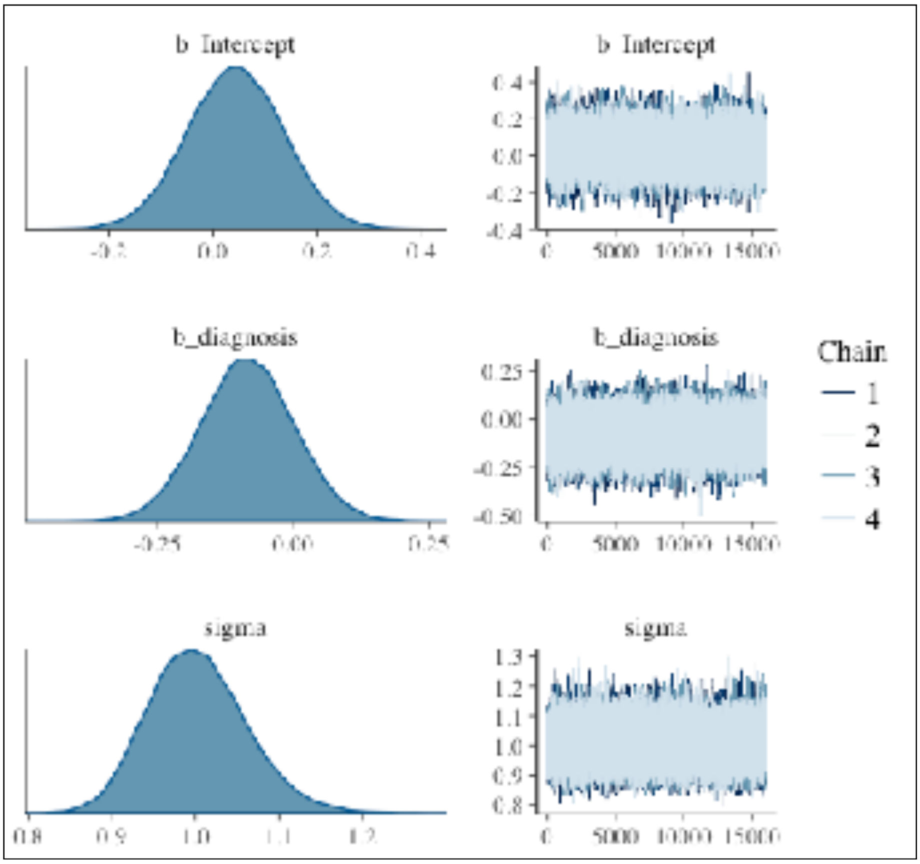


Figure 4.2 - Plots showing the estimated parameter distributions and the chains of the MCMC for model 1.

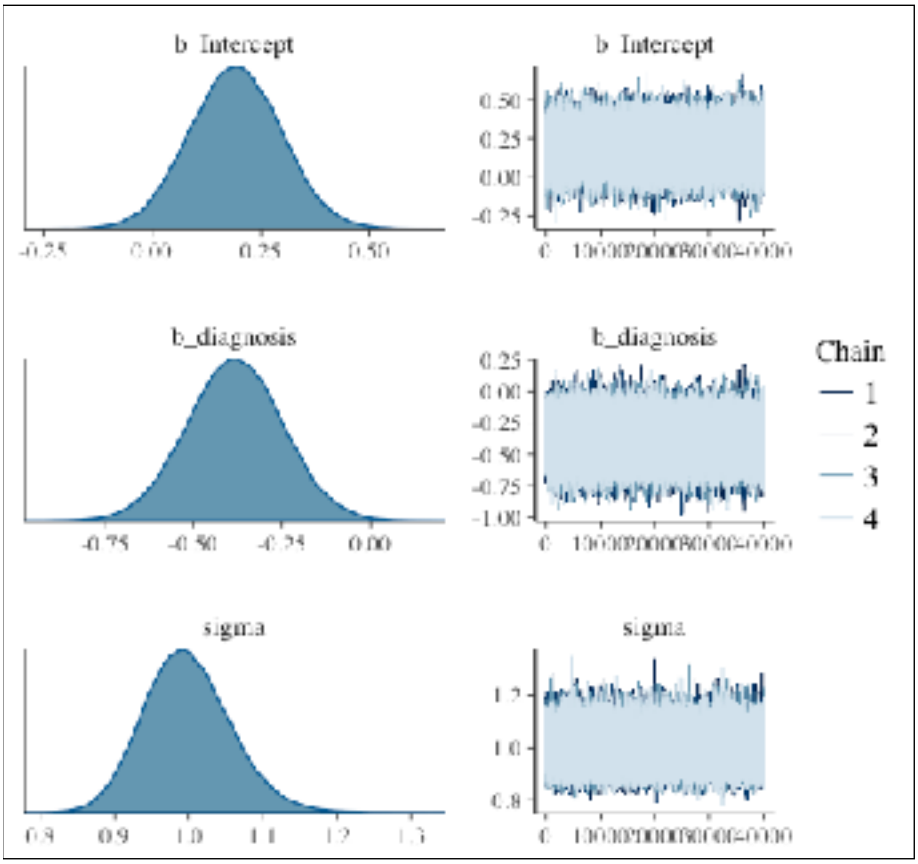


Figure 4.3 - Plots showing the estimated parameter distributions and the chains of the MCMC for model 2.

	WAIC	Standard Error
Model 1	433.30	34.94
Model 2	432.71	36.69
Null model	434.18	34.14
Model 1 - model 2	0.59	3.95
Model 1 - null model	-0.88	1.36
Model 2 - null model	-1.48	5.27

Table 4.4 - Table containing WAIC values and associated standard error for the three models and the difference between them.

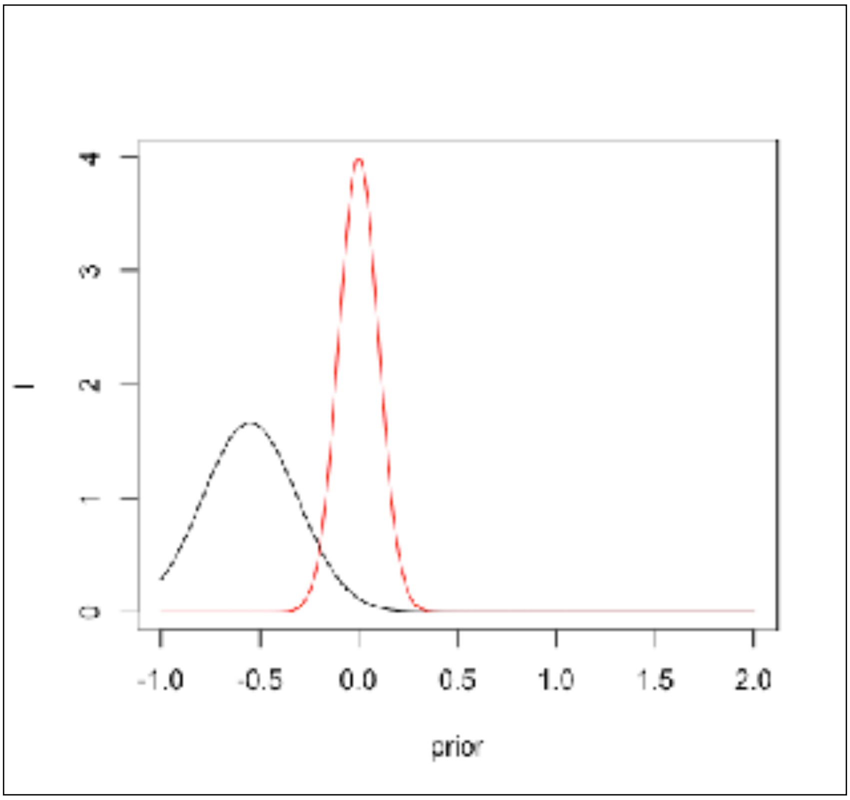


Figure 4.4 - Plot showing the prior distribution for model 1 (red line) and model 2 (black line)

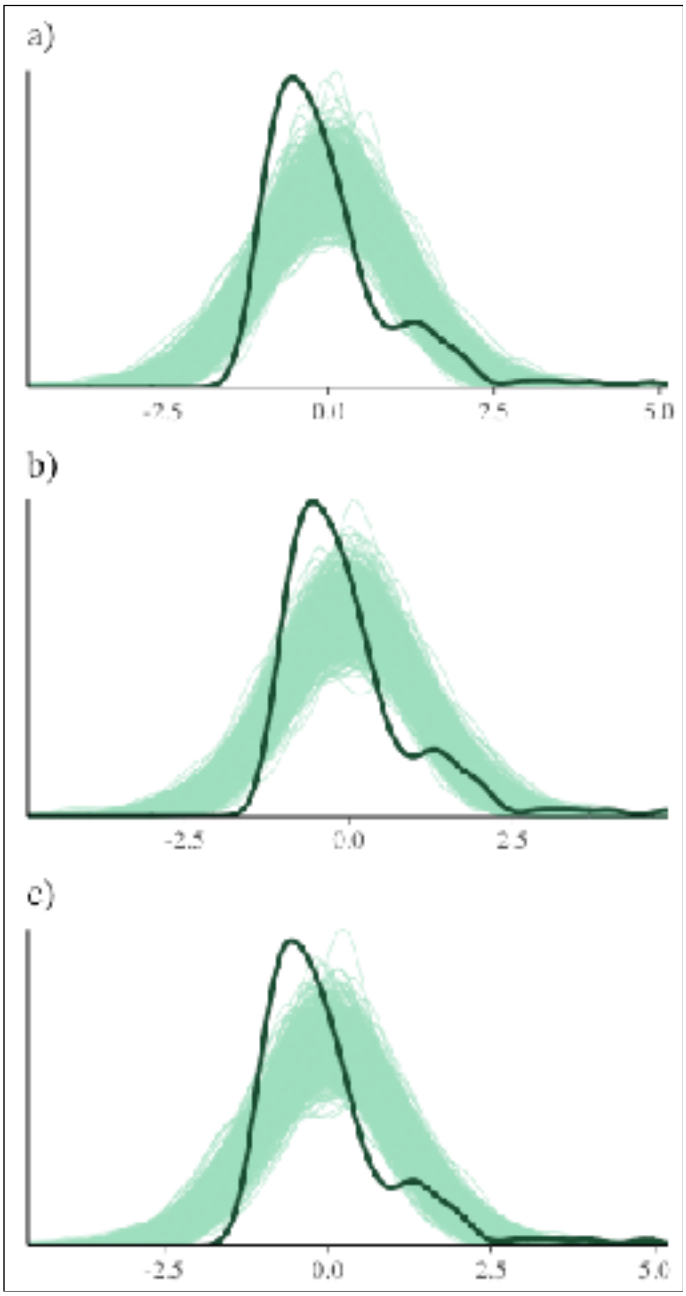


Figure 4.5 - Density plot showing the observed data (dark line) and 1000 simulated predictive posterior distributions (light green lines) for the different model 1 (a), model 2 (b) and the null model (c)

Model 2 shows to give the lowest WAIC value in comparison with the others and Model 1 is still better than the null model. Model 2 also has a larger effect size for the beta than model 1 and the credibility interval doesn't overlap with zero contrasting the credibility interval of the beta in model 1. This means that the effect of model 2 is more reliable than the effect of model 1. In figure 4.4 the priors of the beta parameters are visualized. From this plot it is evident that the two models assumes quite different prior states of the world, but as can be seen in figure 4.5, the predictive posteriors are hard to tell apart which suggests that the difference in size of the effects are not that big in terms of prediction.

Conceptually it is important for the progression of science to be able to incorporate meta analytical data in statistical analysis of empirical studies. The reason for this is that in order to get closer to any "true" effect one has to take as many observations in as possible. Using meta analytical priors is a way to incorporate the existing knowledge of a field and thereby a way to get closer estimations of "true" effects.
This study shows how this can be done and also provides evidence that a meta analytical prior can improve a model. As discussed above the meta analytical prior did not improve the model much, but it still serves to show the better performance of a model incorporating meta analytical priors. However the use of meta analytical priors might not bee exclusively better than conservative priors of the type used in this study. The reason for this is that scientific practice should be one of skepticism and using conservative priors is a way to model this. Studies within a field might be biased in many unpredictable ways, and these biases will be build into models with meta analytical priors if not scientists are not careful. In conclusion meta analytical priors is a good tool for model improvement but should not dominate in the bigger picture of scientific practice.

Christoffer Lundbak Olesen 25/05/2018

Assignment 5

Group members:
Mads Birch Sørensen (201606024)
Christoffer Lundbak Olesen (201608188)

Link to GitHub:
https://github.com/MadsBirch/socult.git

Introduction

To explore the relationship between resource asymmetry and cooperation we build a simulation model and coded it in the coding language "R" (R Core Team, 2017). In our simulation a total of 144 actors divided into 8 groups (18 actors in each group) are playing a trading game with the goal of maximizing their level of happiness. Each simulation they go through a sequence of 1000 rounds and the number of outgroup trades was recorded for each round. We had a total of 16 conditions with 4 values of asymmetry and 4 values of isolation. A total of 40 simulations were run.

Measures

In our analysis we used the amount of out-group trades as a measure of intergroup cooperation. Further, the variables for asymmetry, isolation, and time reflect the 16 conditions and their experimental effect over time.
To improve model convergence, the variables asymmetry and isolation were mean centered while the round variable was scaled. As the out-group trade variable is a count variable with a very large n and a low rate of success (out-group trades), we decided to implement a Poisson outcome distribution with a logarithmic link function. Due to time constraints, a subset of the data including 10 simulations and 200 rounds, was used for the analysis.

The models

We implemented six multilevel models using the brms package in the coding language "R" (Bürkner, 2017; R Core Team, 2017).
We employed four different multilevel linear regression models all with the variable out-group trade as dependent variable, round as a predictor and a random component allowing for random intercept for each simulation and a random slope for the round variable. In addition, a null model was constructed to be included in a model quality comparison procedure.
The full model included a three-way interaction between the variables asymmetry, isolation, and round as the effect of asymmetry varies depending on the level of asymmetry and vice versa. Additionally, the round variable, accounting for the cumulative effect of time, interacts with both asymmetry and isolation as the agents learn from their previous choices.

Below the full model is notated, where S denotes the simulation variable, A denotes the asymmetry variable, I denotes the isolation variables and R denotes the round variable. Priors were set with the intent to improve model sampling, as the large amount of data ends up overwhelming the set priors regardless.

$$\text{Outgroup Trade} \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \alpha_{S[i]} + \beta_A A_i + \beta_I I_i + \beta_{R|S[i]} R_i + \beta_{AIR} A_i I_i R_i + \beta_{AI} A_i I_i + \beta_{AR} A_i R_i + \beta_{IR} I_i R_i$$

$$\begin{bmatrix} \alpha_S \\ \beta_S \end{bmatrix} \sim MVNormal\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}, S\right)$$

$$S = \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} R \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix}$$

$$\alpha \sim Normal(0,5)$$

$$\beta_A \sim Normal(0,2)$$

$$\beta_I \sim Normal(0,2)$$

$$\beta_R \sim Normal(0,1)$$

$$\beta_{AI} \sim Normal(0,5)$$

$$\beta_{AR} \sim Normal(0,1)$$

$$\beta_{AIR} \sim Normal(0,2)$$

$$\beta_{IR} \sim Normal(0,1)$$

$$\sigma_\alpha \sim Cauchy(0,2)$$

$$\sigma_\beta \sim Cauchy(0,2)$$

$$R \sim LKJcorr(2)$$

Model 5.1 - The full model

Results

The models were run for a total of 2000 iterations over 2 chains, of which 1000 iterations were warmup samples. The full model converged nicely, which is indicated by the fact that rhat equals 1 for all model estimates and the number of effective samples is high relative to the set number of samples for most parameters. Additionally, the model estimates are both well-mixing and stationary, as illustrated by the trace-plots in Figure 1a and 1b.

The model estimated the effect of the interaction between the variables asymmetry, isolation, and round to be 1.41, 95% CI [1.12, 1.77], meaning that the interaction between the three variables are estimated to affect out-group trade (Table 1).

However, from these estimates it is not evident in which direction this interaction is predicted to affect out-group trade. For that reason, the marginal effects are illustrated in Figure 4 and 5. Figure 4 illustrates the effect of asymmetry when isolation and round are kept constant while Figure 5 shows the effect of isolation when asymmetry and round are kept constant. Additionally,

Figure 3 illustrates that when isolation is low the effect of resource asymmetry is higher than if isolation is high and vice versa.

Figure 3 illustrates the fact that the effect of round is not meaningful in isolation, but the variable contains valuable information about the development in the other variables, when included in the interaction. The inclusion of the round variable is justified for two reasons. The first reason is conceptually rooted in the design of the simulation where agents play the trading game over a number of rounds and the effect of resource asymmetry and isolation is measured over time in the amount of out-group trades happening. Second, the inclusion of the round variable allows us to model a random slope for the cumulative effect of time in each simulation.

Parameters:	Estimate	Est.Error	l-95% CI	u-95% CI	Eff.Sample	Rhat
α	1.98	0.01	1.97	2.00	572	1.00
Asymmetry	-1.23	0.02	0.02	-1.19	2000	1.00
Isolation	-2.68	0.02	-2.72	-2.65	2000	1.00
Round	0.00	0.00	-0.01	0.00	1122	1.00
Asymmetry*Isolation	3.30	0.16	2.97	3.61	2000	1.00
Asymmetry*Round	-0.53	0.02	-0.56	-0.49	2000	1.00
Isolation*Round	0.00	0.02	-0.03	0.04	2000	1.00
Asymmetry*Isolation*Round	1.44	0.17	1.12	1.77	2000	1.00

Table 1 - Main effects of full model

Parameters:	Estimate	Est.Error	l-95% CI	u-95% CI	Eff.Sample	Rhat
σ_α	0.02	0.01	0.01	0.04	667	1.00
σ_β	0.01	0.00	0.00	0.02	583	1.00
R	0.34	0.32	-0.35	0.84	1287	1.00

Tabel 2 - Random effects of full model

Model quality

Comparing the five models using the widely applicable information criterion (WAIC) showed that the full model is predicted to outperform all other models in this analysis (Table 3). In addition, Figure 2 illustrates that the predictive posterior of the model fits the data well, reflecting the distribution of the data closely.

Model:	WAIC	SE
Full model	183266.81	477.38
No interaction model	184726.88	488.62
Isolation model	190215.92	531.68
Asymmetry model	207243.83	636.51
Null model	213054.47	531.68
Difference between full model and:	WAIC	SE
No interaction model	-1460.07	109.39
Isolation model	-5489.03	244.42
Asymmetry model	-23977.02	440.65
Null model	-29787.66	847.84

Tabel 3 - Model comparison using WAIC

Figures

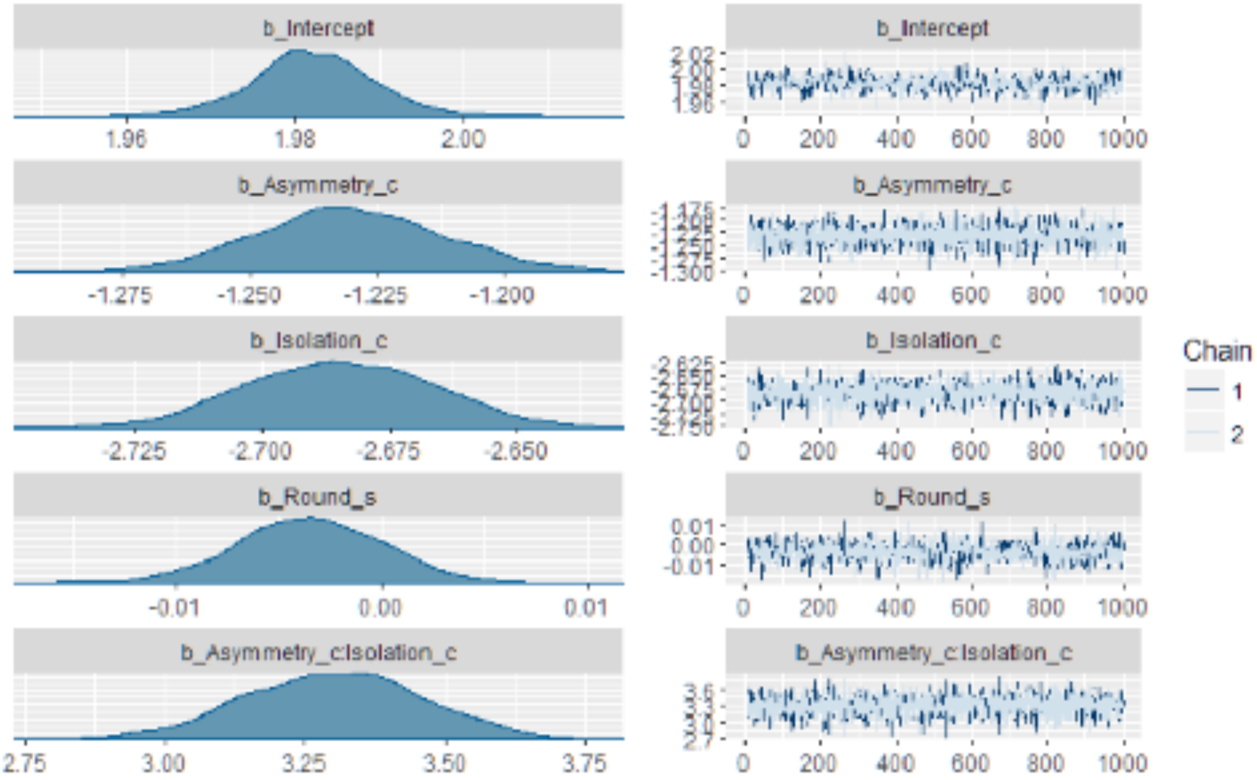


Figure 1a - Plot of model beta distributions and trace plots

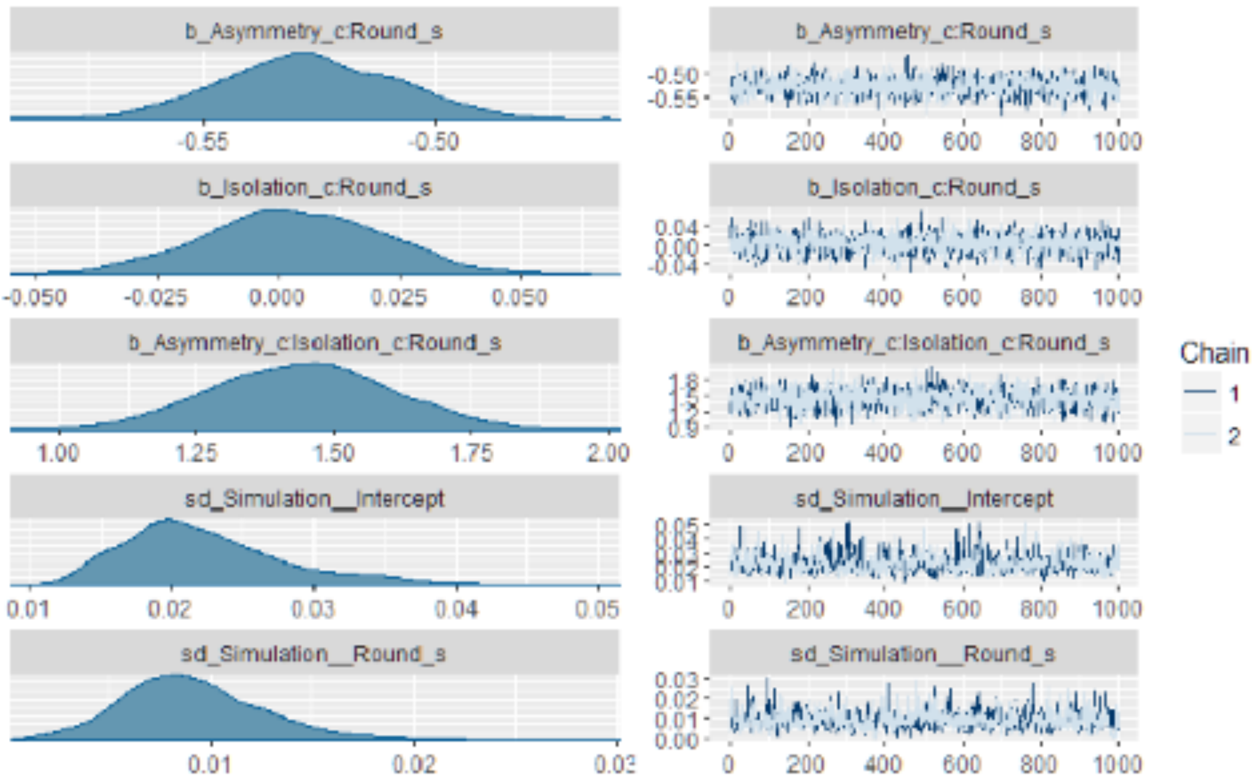


Figure 1b - Plot of model beta distributions and trace plots (cont.)

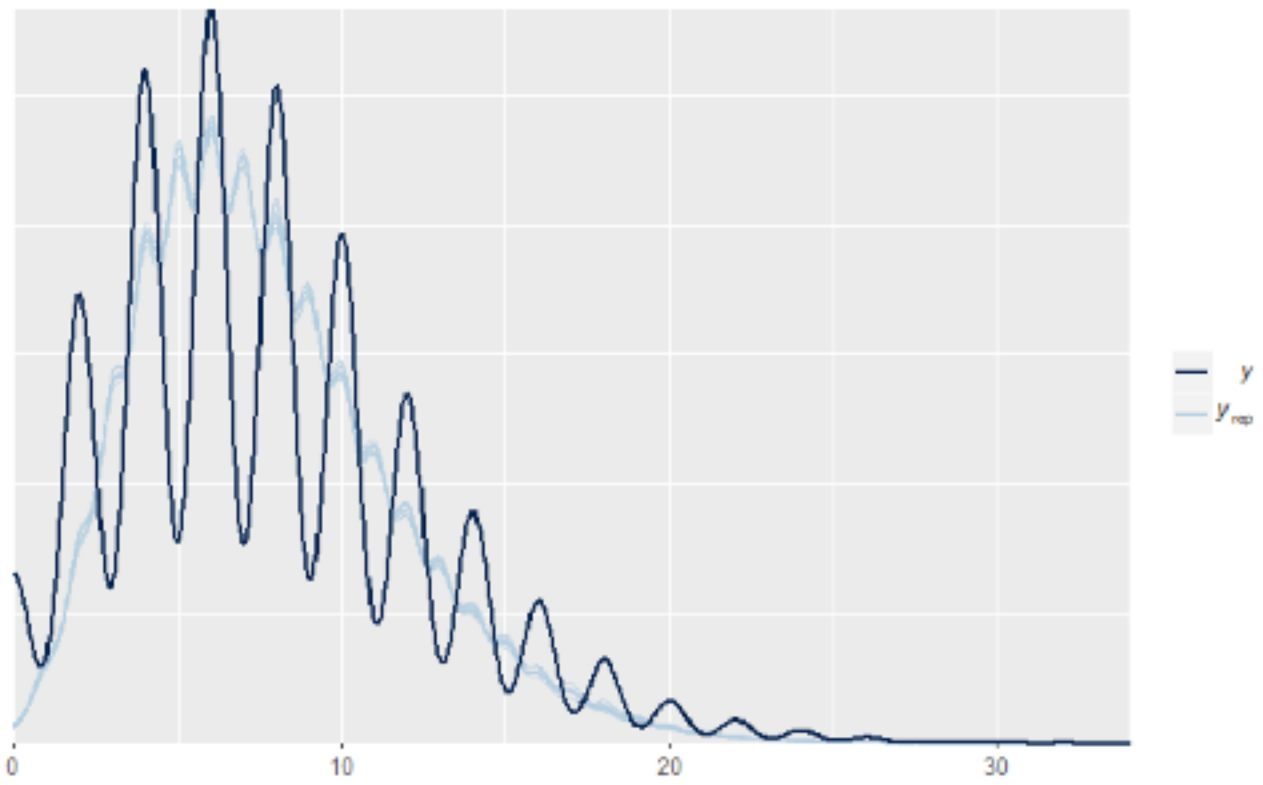


Figure 2 - Predictive posterior of full model

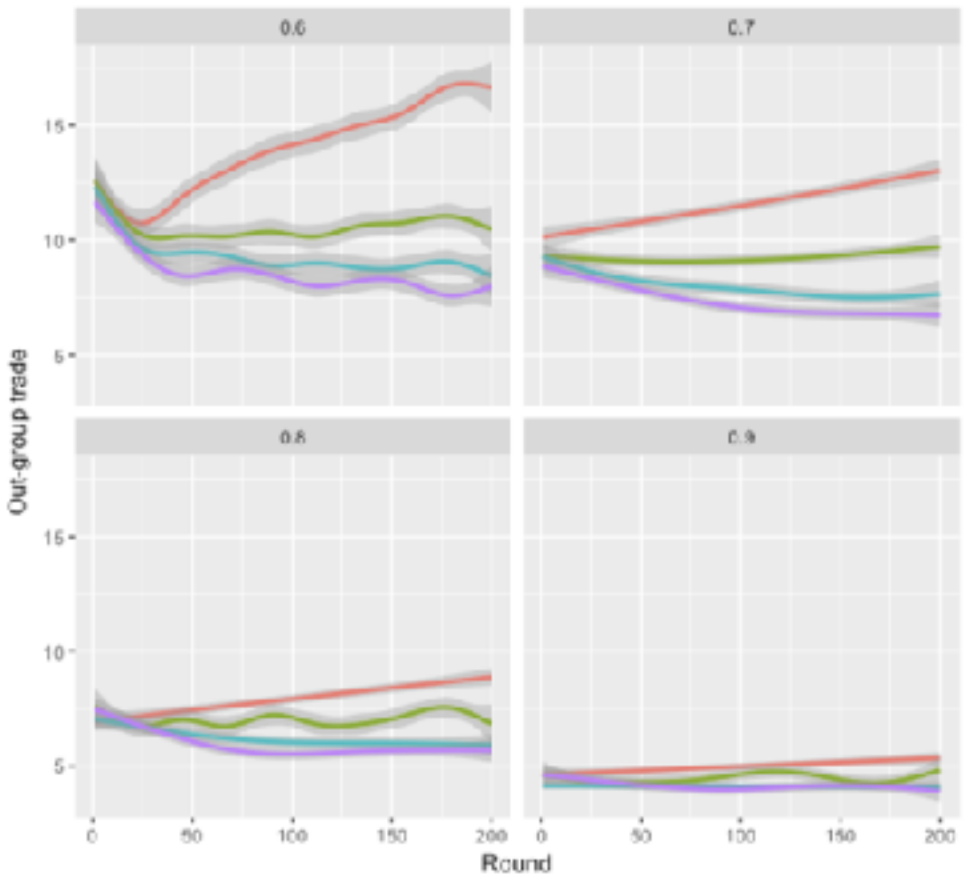


Figure 3 - Four plots showing the average development of outgroup trade over 200 rounds. Each plot represents the values of the isolation conditions and the colors represents the values of the asymmetry conditions. Red = 0.25, green = 0.35, blue = 0.45 and purple = 0.55. Note that a lower number is a more asymmetric distribution of resources.

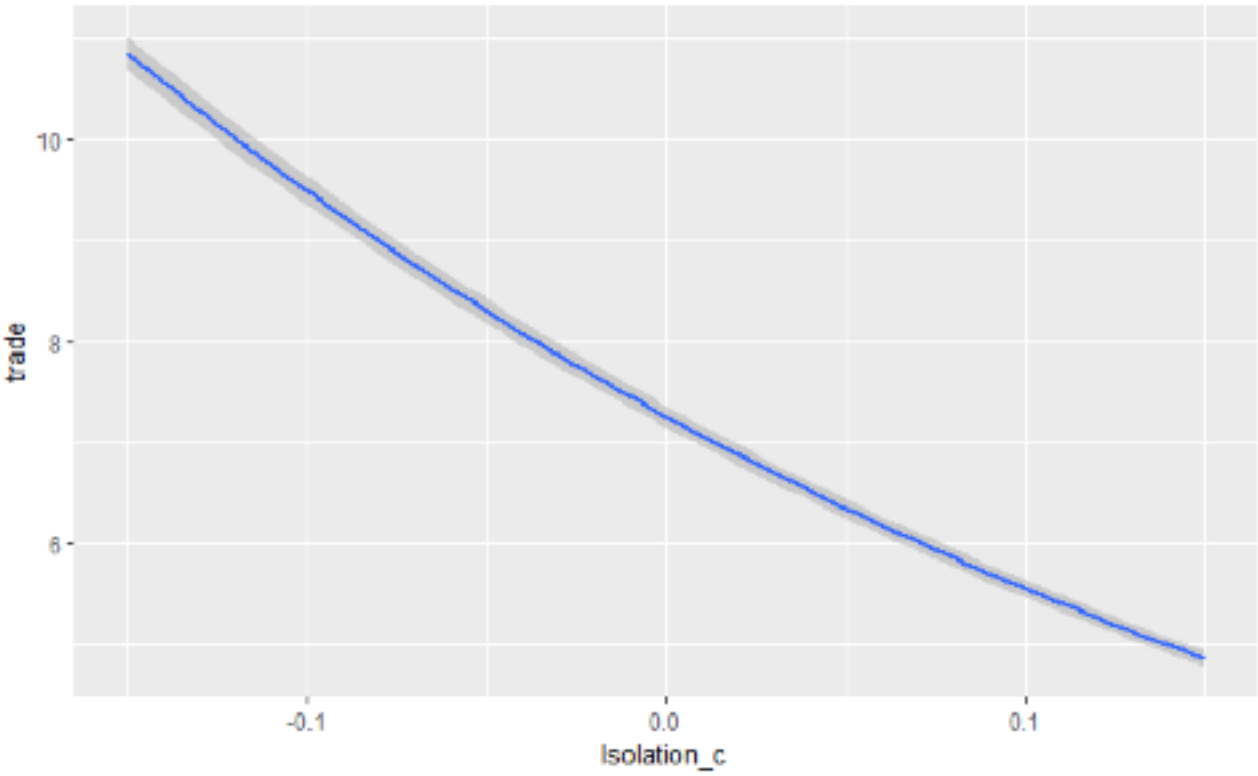


Figure 4 - Marginal effects for the asymmetry variable

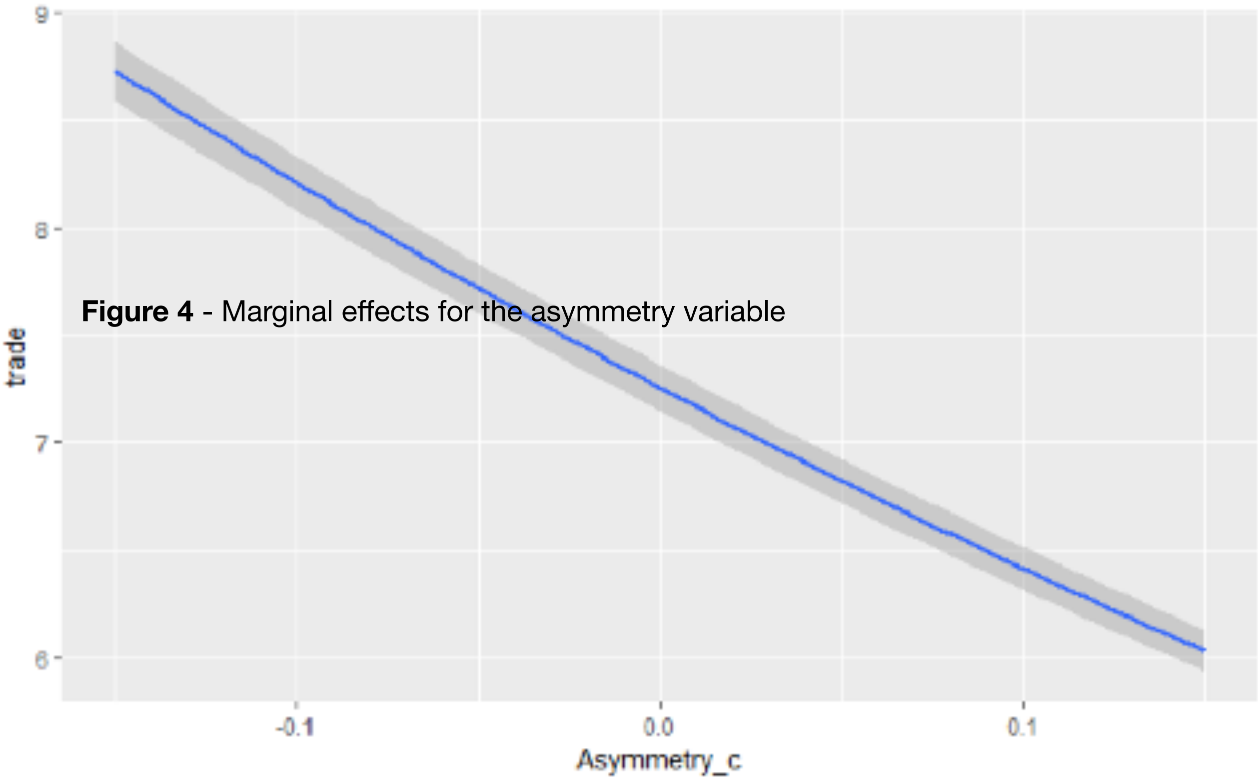


Figure 5 - Marginal effects for the isolation variable

References

Bürkner, P.-C. (2017). **brms** : An *R* Package for Bayesian Multilevel Models Using *Stan*. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from www.R-project.org