# Assignment 2: CogSci Knowledge

Deadline for hand-in: 11/3-2020
Github: https://github.com/saraoe/assignment-2-4

# Part 1

**1. What's Riccardo's estimated knowledge of CogSci? What is the probability he knows more than chance (0.5)**

- According to the posterior distribution of Riccardo's (RF's) knowledge of CogSci, it is most probable that Riccardo's knowledge of CogSci is 50% (the peak probability is at 0.5)
- Extracted from the distribution as well, there is a 50 % chance that he knows more than chance (because the area under the curve above 0.5 on the x-axis is 0.5).

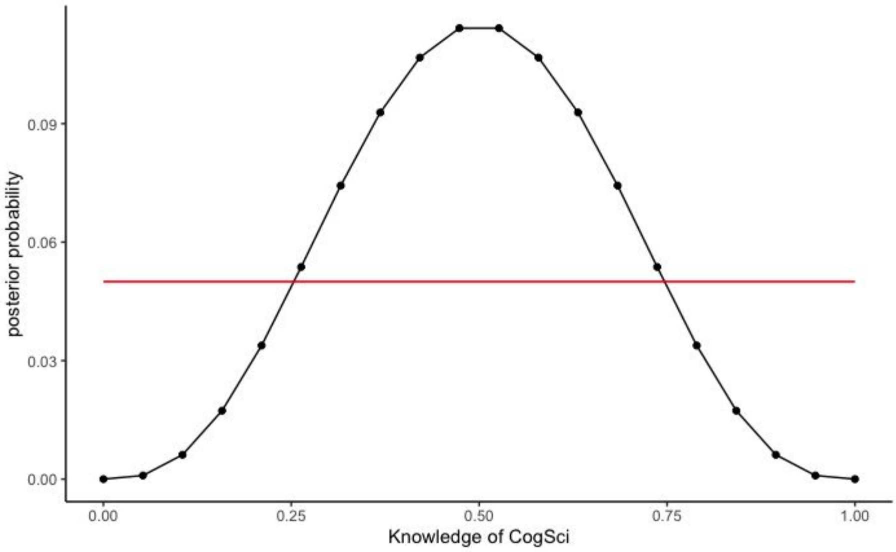



*Figure 1: Prior and posterior distribution of Riccardo's knowledge of CogSci*

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

- KT has the highest peak at 1 on the posterior probability, meaning he has the highest probability of answering all future questions correctly. JS has a slightly lower peak at .81 knowledge of CogSci, however, with a much higher probability (expressed on y-axis in figure 2). Also, there is approx. 0% probability that JS answers below change, while there is approx. 12% probability that KT does (table 1). From sampling the posterior distribution, we have also found that JS has a 52.4 percentage chance of having more knowledge than KT (table 2).
- MW and RF both answered half of their questions correct. However, even though having the same prior, the posterior distribution of the two differs. This is due to MW answering more questions, and the distribution is therefore more narrow around 50%. In other words, fewer

data points means the prior weighs more on the posterior (relative to the data). From sampling the posterior distribution, we have also found that RF has a 49.6 percentage chance of having more knowledge than MW (table 2).

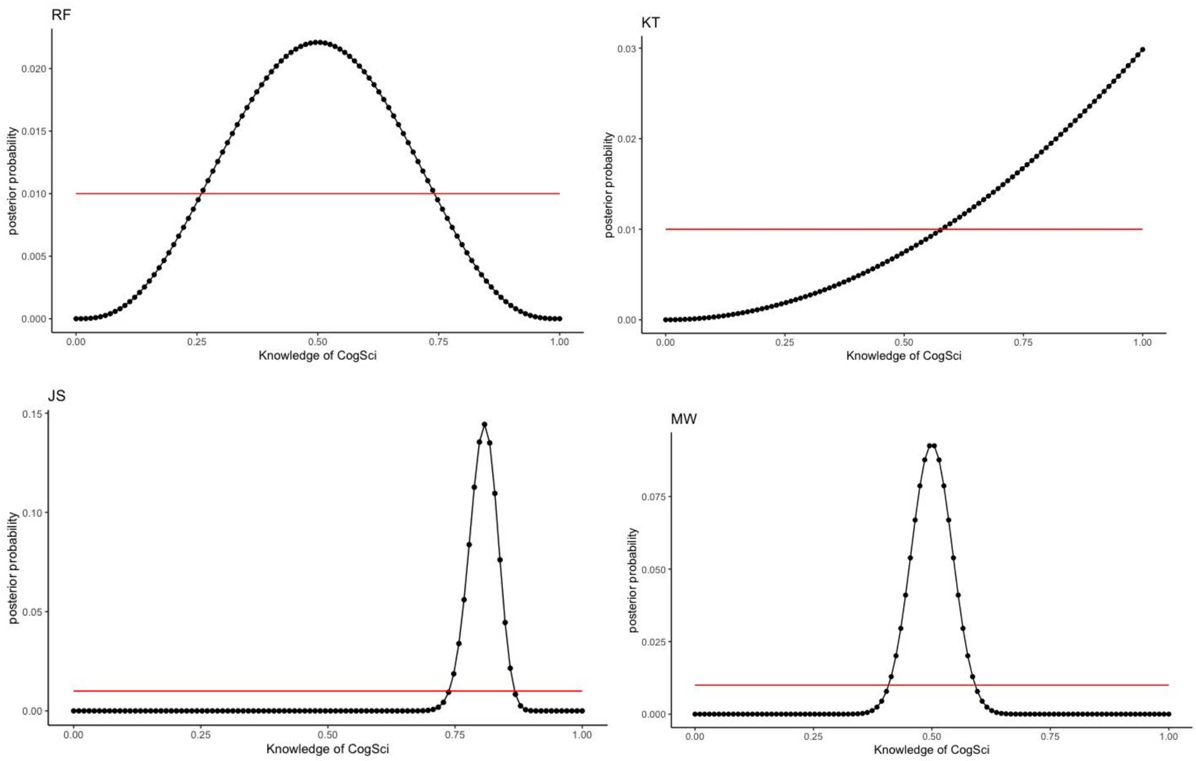



*Figure 2: Prior and posterior distribution of all teacher's knowledge of CogSci with uniform prior*

| **Teacher** | **Knowledge at MAP** | **Probability of knowledge above chance** |
|---|---|---|
| RF | 0.5 | 0.5 |
| KT | 1 | 0.8769 |
| JS | 0.81 | Approx. 1 |
| MW | 0.5 | 0.5 |

*Table 1: Teacher's knowledge of CogSci from posterior calculated from a flat prior*

| | **KT>** | **JS>** | **RF>** | **MW>** |
|---|---|---|---|---|
| **KT** | na | 52.4 | 16.52 | 13.06 |
| **JS** | 47.57 | na | 3.28 | Approx. 0 |
| **RF** | 83.48 | 96.7 | na | 50.37 |
| **MW** | 86.93 | 100 | 49.62 | na |

*Table 2: Matrix of the comparison of teacher's knowledge given the flat prior (e.g. JS has a 52.4 percentage chance of having more knowledge than KT)*

**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)?**

- The posterior distributions have changed as we have changed the prior probability distribution. By changing the prior, the posterior changes according to the new prior.
- This is especially clear when looking at Mikkel Wallentin and Riccardo. Mikkel is hardly influenced (the peak on the posterior distribution is still around .5), whereas Riccardo is highly influenced with a posterior distribution skewed towards .8 (and therefore looks closer to the prior).
- To also compare the results regarding who has the highest probability of being the smartest teacher, we reran the previous comparisons with the new prior.
  In table 4, we can see the probability of the different teachers having a true knowledge of CogSci higher than the others. KT and JS both have very high probabilities of being the smartest. An example of change is that the probability of KT being smarter than JS is now >50, whereas it was <50, with the flat prior.

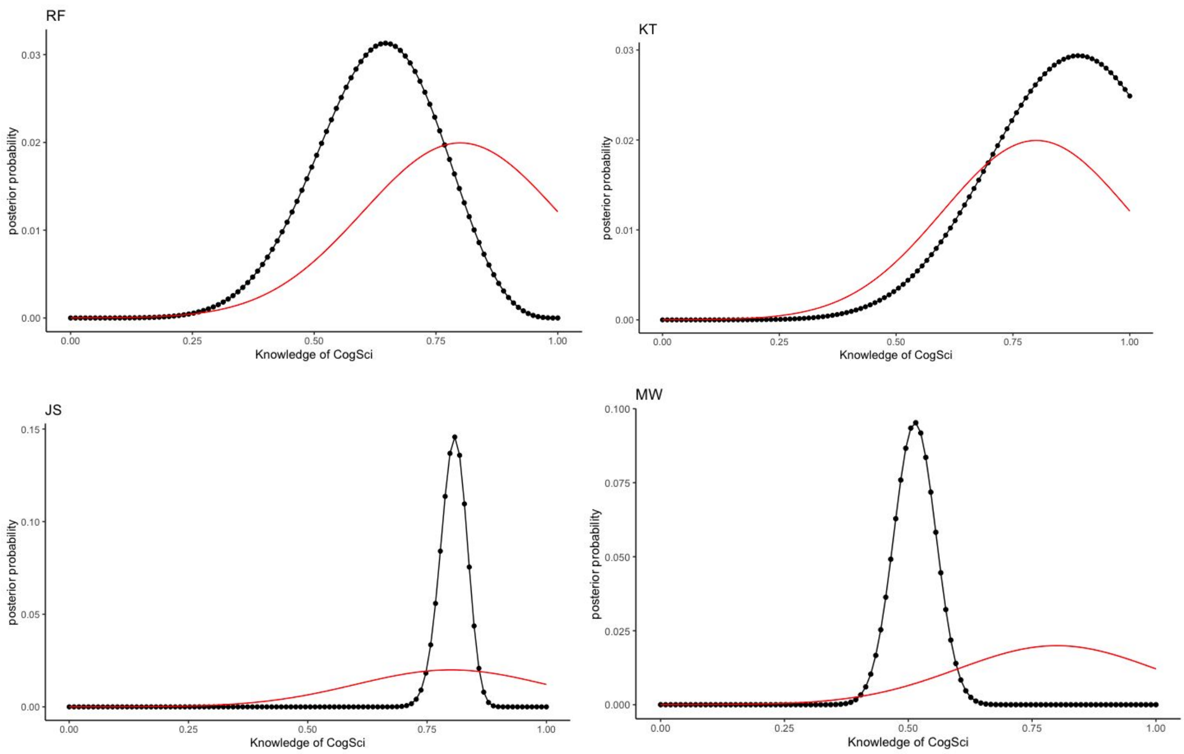



*Figure 3: Prior and posterior distribution of all teacher's knowledge of CogSci with normally distributed prior*

| Teacher | Knowledge at MAP | Probability of knowledge above chance |
|---|---|---|
| RF | 0.65 | 0.8418 |
| KT | 0.89 | 0.9761 |
| JS | 0.81 | Approx. 1 |
| MW | 0.52 | 0.62 |

*Table 3: Teacher's knowledge of CogSci from posterior calculated from a normally distributed prior*

| | KT> | JS> | RF> | MW> |
|---|---|---|---|---|
| **KT** | na | 45.4 | 17.11 | 3.2 |
| **JS** | 54.38 | na | 7.84 | Approx. 0 |
| **RF** | 82.74 | 91.66 | na | 20 |
| **MW** | 96.6 | 100 | 80.2 | na |

*Table 4: Matrix of the comparison of teacher's knowledge given the new prior (0.8 mean and 0.2 SD) e.g. JS has a 45.4 percentage chance of having more knowledge than KT*

**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?**

- In the results above we found that differences in the prior affected the posterior distribution heavily. However, after adding extra data (by multiplying the previous numbers by 100) the differences between the posterior distributions with different priors decreases as the data increases.

**Uniform prior data:**

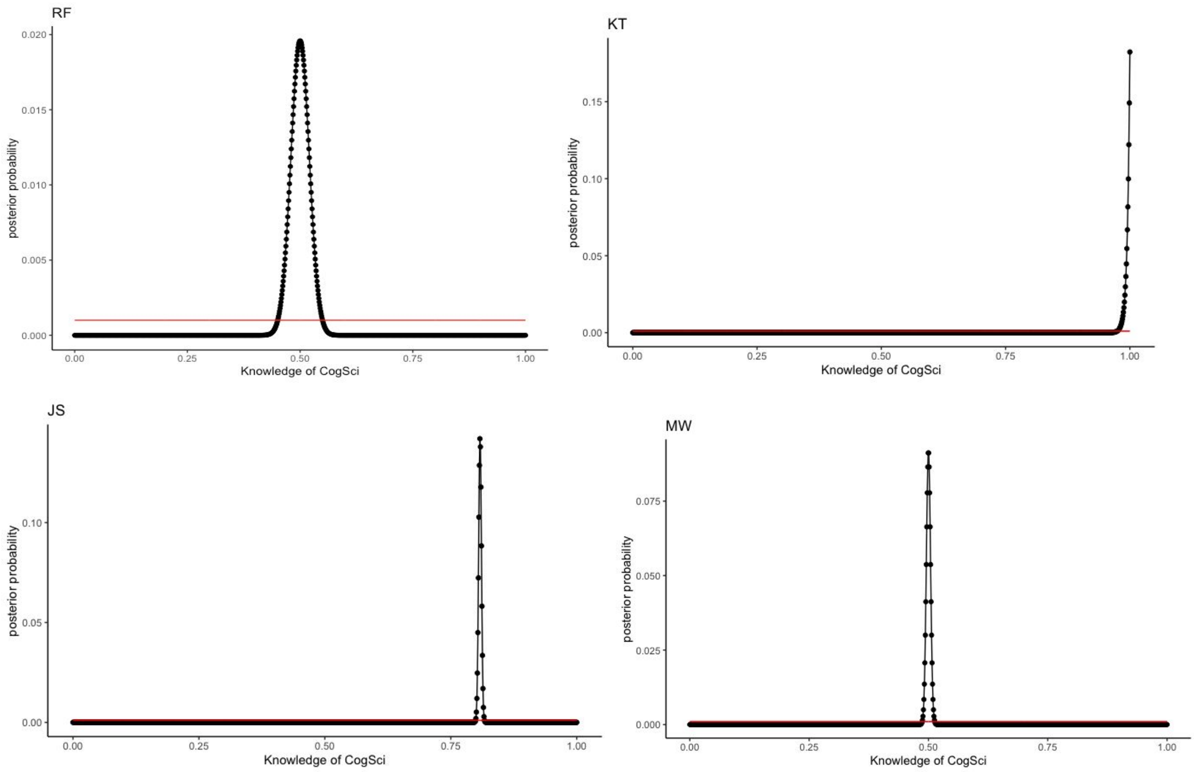



*Figure 4: Prior and posterior distribution of all teacher's knowledge of CogSci with uniform prior and new data*

| Teacher | Knowledge at MAP | Probability of knowledge above chance |
|---|---|---|
| RF | 0.5 | 0.5 |

| | | |
|---|---|---|
| KT | 1 | Approx. 1 |
| JS | 0.808 | Approx. 1 |
| MW | 0.5 | 0.5 |

*Table 4: Teacher's knowledge of CogSci from posterior calculated from a flat prior with the new data*

*Normal prior data:*

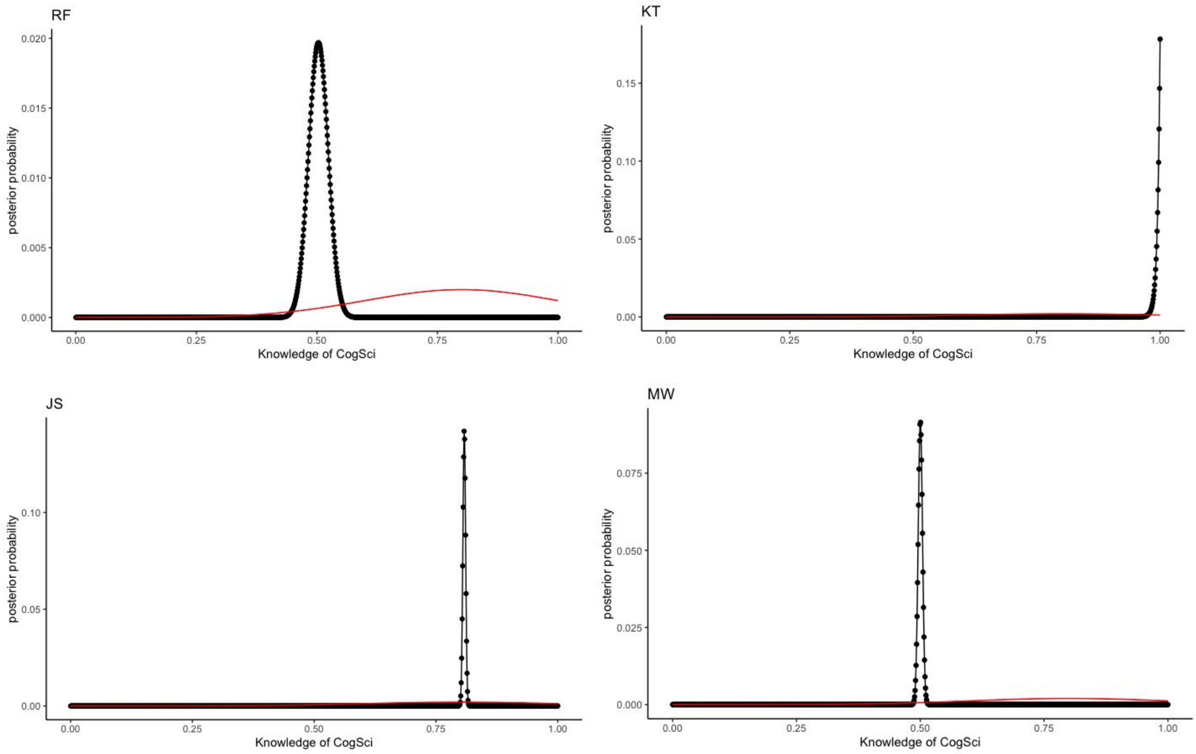



*Figure 5: Prior and posterior distribution of all teacher's knowledge of CogSci with a normally distributed prior and the new data*

| Teacher | Knowledge at MAP | Probability of knowledge above chance |
|---|---|---|
| RF | 0.504 | 0.5604 |
| KT | 1 | Approx. 1 |
| JS | 0.808 | Approx. 1 |
| MW | 0.501 | 0.5130 |

*Table 5: Teacher's knowledge of CogSci from posterior calculated from a normally distributed prior with the new data*

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

- Since it does not change the facts about the data we have to change the prior (our own assumption)

- As we are skeptical, we believe that the mean will be .5 (chance). Moreover, we have made a small SD (0.05), as we are fairly certain that all teachers will answer close to half of the questions correct by change.

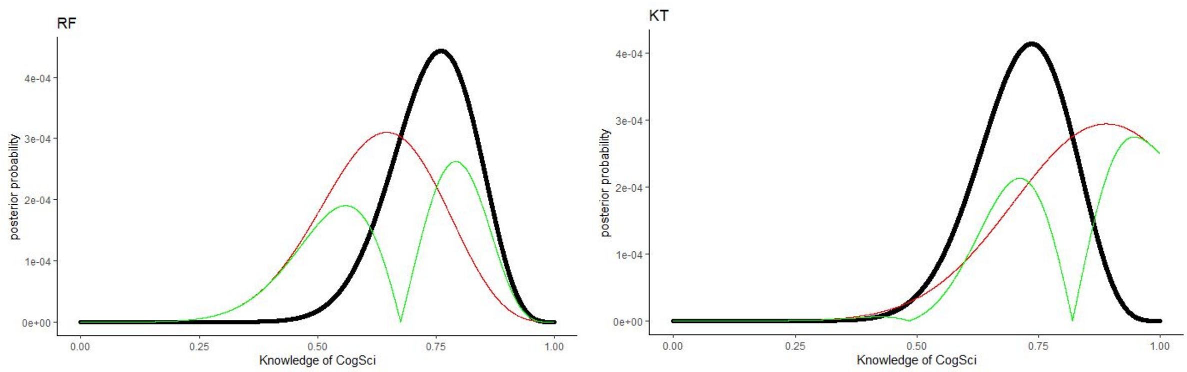
# Part 2 - Focusing on predictions

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

**In a frequentist framework,** we make a null hypothesis to test our assumptions. Our null hypothesis would be that there is no difference in the teacher's knowledge of CogSci from the first to the second year. We make a t-test and investigate this hypothesis, and if the p-value is below 0.05 we can reject the null hypothesis.
This type of model can lead to overconfident estimations of what is right and wrong, because the conclusion is binomial - either we reject the null hypothesis or we don't.

**In a Bayesian framework**, we would not test the difference on an arbitrary cutoff. Rather, we would just explain the differences in the prior, and the posterior. This we could do in a number of different ways, e.g. plotting prior and posterior, point estimate (e.g. MAP), intervals (e.g. PI or HPDI) or subtracting the prior from the posterior.

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



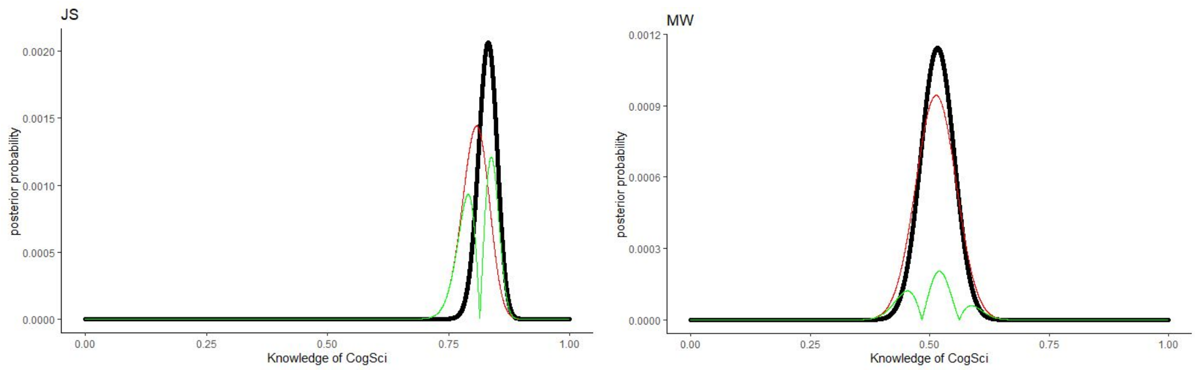



*Fig. 6: Plot of prior (red), posterior (black) and the difference between the two in absolute numbers (green)*

We have calculated a new posterior from the new data, using the posterior from the first year as the prior. The distributions are shown in figure 6.
The results from our model on the data from the first year seems to predict the data from the second year fairly well. The predictions are especially good for MW, where the difference between the posterior from the first year and the posterior from the second is very small (see green line on fig. 6). The overlap of the old posterior (now prior) and the posterior distributions are, however, quite large for all teachers, indicating that it makes relatively good predictions about the teachers knowledge of CogSci.

We have also calculated the difference in maximum a posteriori (MAP) for the two posterior distributions (old data and new data) of each teacher (table 5). This too shows that the prediction is good for MW. However, there is a somewhat large difference in MAP for KT and RF. Though, when examining the plots of the distributions (fig. 6) there is a great overlap.

| teacher | MAP_posterior | MAP_prior | MAP_diff |
|---|---|---|---|
| JS | 0.8321 | 0.8079 | 0.0242 |
| KT | 0.7364 | 0.8899 | 0.1535 |
| MW | 0.5166 | 0.5136 | 0.0030 |
| RF | 0.7611 | 0.6464 | 0.1147 |

*Table 5: MAP for both posterior (map_posterior) and prior (map_prior) distributions for each teacher and the difference between those (map_diff)*

| | **Prior HPDI (50%)** | **Posterior HPDI (50%)** |
|---|---|---|
| **RF** | 0.5658566 - 0.7353735 | 0.6969697 - 0.8187819 |
| **KT** | 0.8137814 - 0.9838984 | 0.6679668 - 0.7981798 |
| **JS** | 0.7868787 - 0.8243824 | 0.8192819 - 0.8449845 |
| **MW** | 0.4826483 - 0.5391539 | 0.4944494 - 0.5417542 |

*Table 7: HPDI intervals for each individual, prior (which was the posterior for the old data) and posterior for the new data*

Now, we assess how the model we have generated from the old data will predict the new data. We sample 10.000 times from the posterior distribution of each teacher and asses the change of obtaining the specific number of correct answers given the model we have from the old data (+/- 3%).

| JS | RF | MW | KT |
|---|---|---|---|
| 28.24% | 9.6% | 24.11% | 10.72% |

*Table 8: Prediction of new data given old model with +/- 3% interval from actual data*

The old model predicts the new data for JS and MW well - above 20% of the times the model would estimate a guess within the 3% interval of the actual guess in the new data (table 8). However, the old model does not predict RF and KT as well. This can also be seen in the plot of prediction errors (figure 7), where the errors of JS and MW have a higher density around zero (no prediction error), while this is not the case for RF and KT.

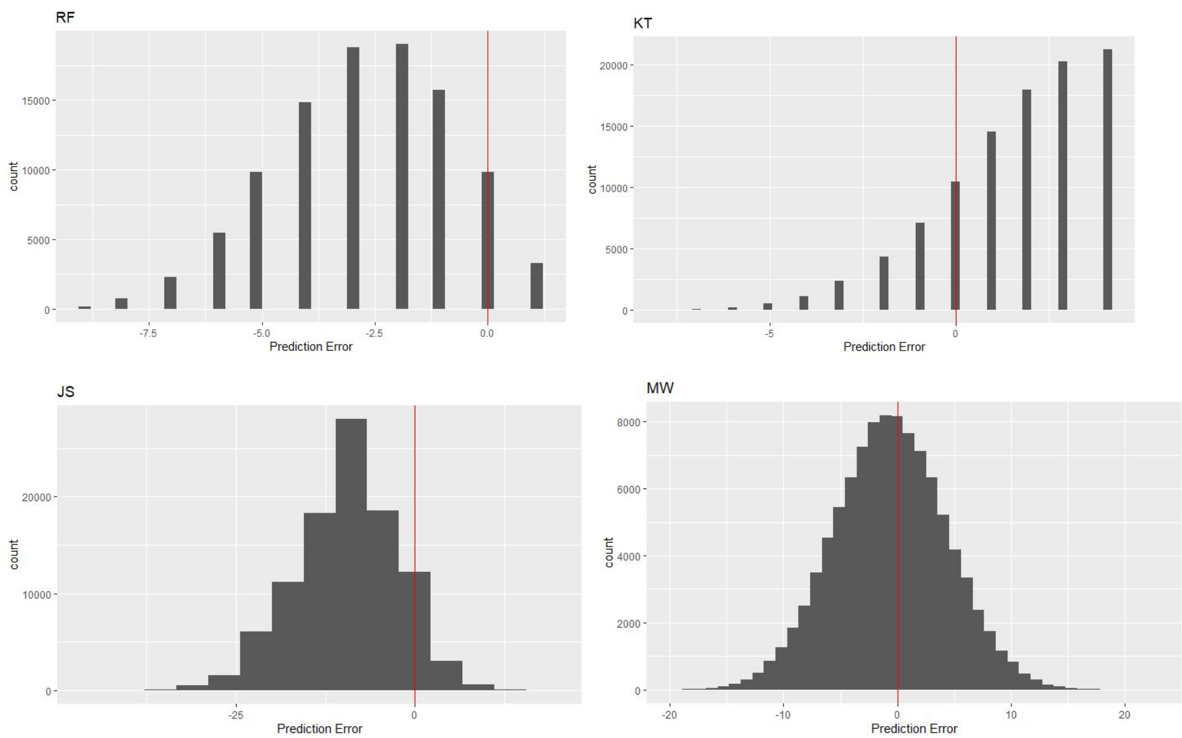



*Figure 7: Prediction errors for each teacher, when using the old posterior to predict the new data. The red line is zero prediction error.*