Estimate you teacher's CogSci knowledge

Link to GitHub: **HERE**

Part 1

1.1 What's Riccardo's estimated knowledge of CogSci? What is the probability he knows more than chance?

Using the grid approximation Riccardo's CogSci knowledge is estimated with the highest posterior probability to be 0.5. The resulting posterior distribution is shown in Figure 1 (top left). Using the quadratic approximation and under the assumption that the distribution is Gaussian the knowledge is estimated to be 0.5 with 89% confidence interval between 0.17 and 0.83. The probability that Riccardo knows more than chance is 0.5.

1.2 Estimate all the teachers' knowledge of CogSci. Who's best?

When the posterior distributions of all teacher's knowledge is plotted we can see that Kristian's estimate of his knowledge is the best with highest posterior probability at 1. However since he answered so few questions the credibility of the estimate is rather doubtful.
Josh's knowledge is the largest after Kristian and with his narrow distribution this estimate is also the most credible one.

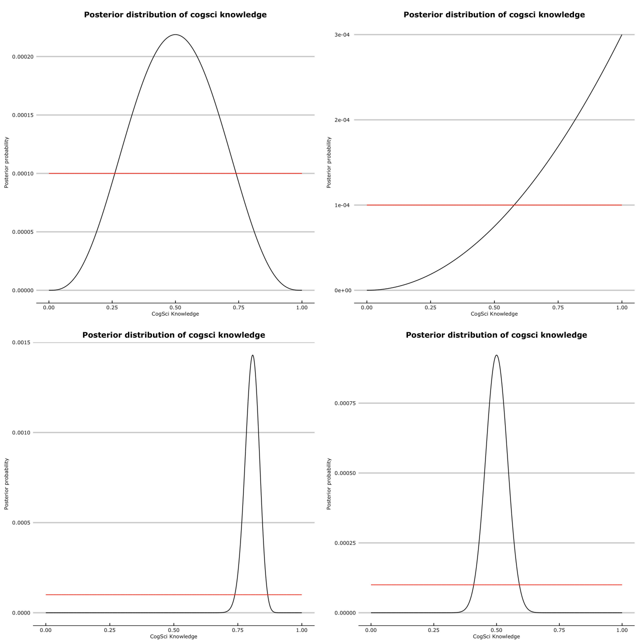

Figure 1: Posterior distributions of all teachers; top-left: Riccardo, top-right: Kristian, down-left: Josh and down-right: Mikkel; with prior plotted in red.

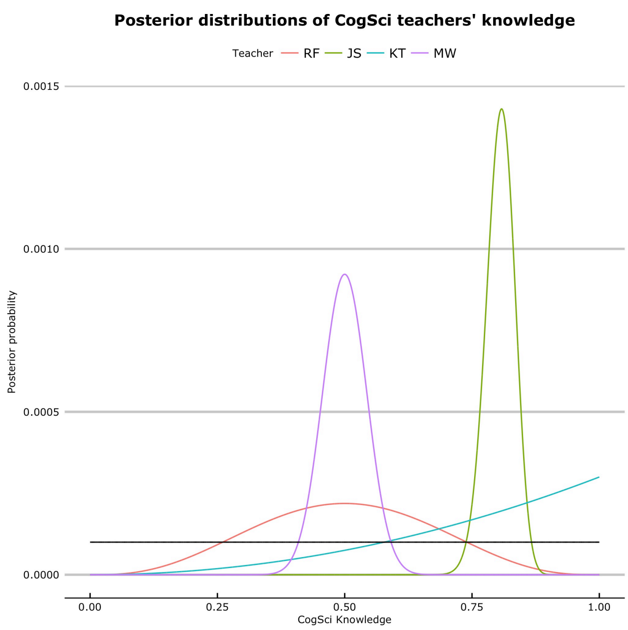

Figure 2: Posterior distributions of all teachers plotted together; red: Riccardo, blue: Kristian, green: Josh, purple: Mikkel and black: prior.

1.2.1 Comment on the posteriors of Riccardo and Mikkel

To compare the posteriors of Riccardo and Mikkel easier they are plotted together in Figure 3. The modes of both distributions are at 0.5. However, since Mikkel answered considerably more questions than Riccardo, his distribution is narrower around the mode which makes that estimate more credible.

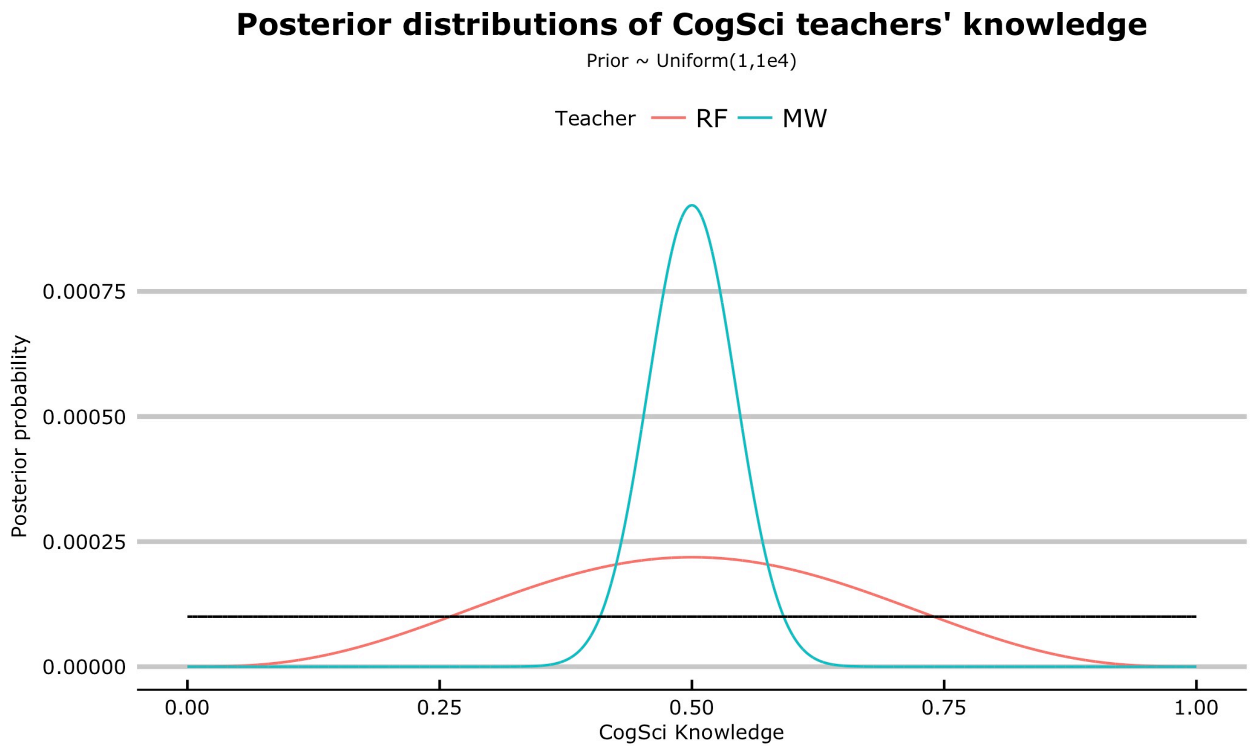

Figure 3: Posterior of Riccardo (red) and Mikkel (blue) plotted together for better comparison.

1.3 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)?

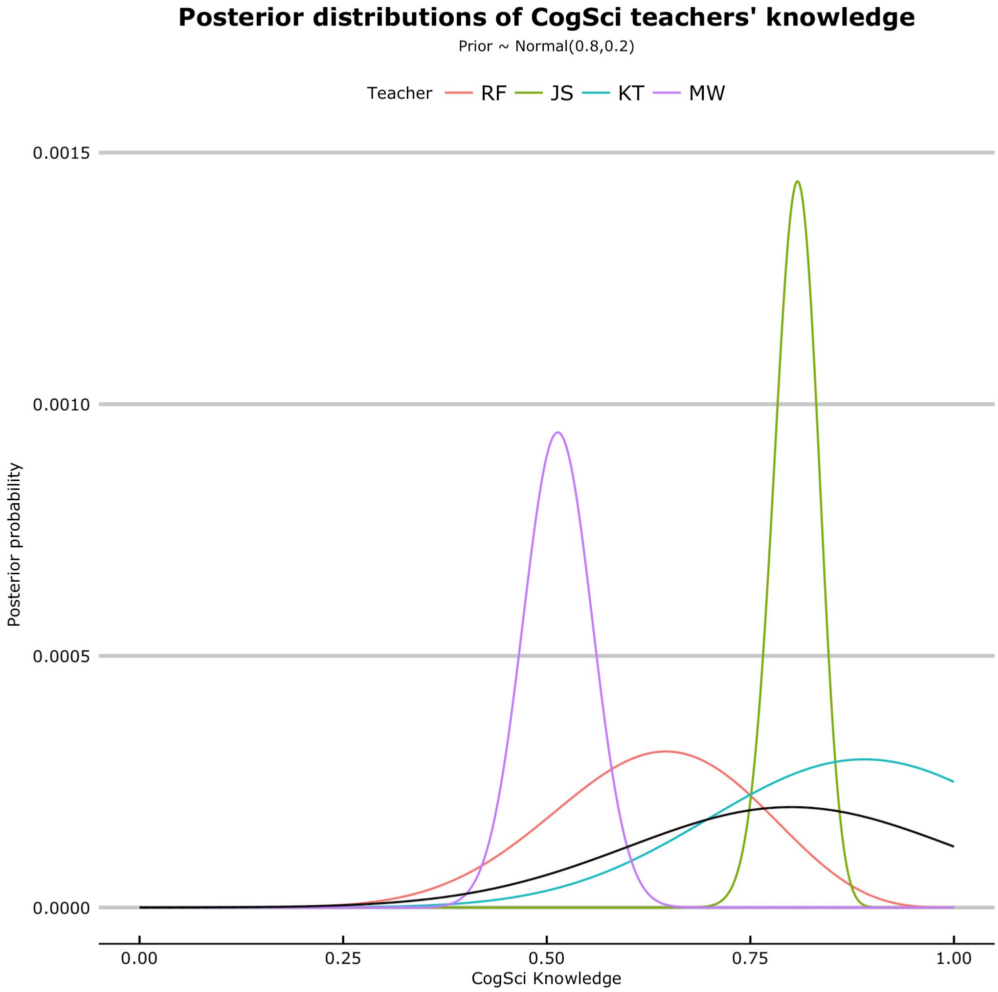

Figure 4: Posterior distributions of all teachers using normal distribution as prior.

The normal prior does change the posteriors.

Riccardo's posterior's mode shifts with the normal prior from 0.5 to 0.7 so more to the mean of the prior. The confidence of the model in this estimate is also higher than with the flat prior. It remains however quite broad distribution.

Josh's knowledge estimate is based on a lot of data and therefore the estimate was quite confident with flat prior. With the normal prior the posterior doesn't change because the mean of the prior is at 0.8 which is similar value as the mean of Josh's posterior with flat prior. With the normal prior we basically confirm our prior and we should increase our belief that Josh is very knowledgable.

Kristian's posterior changed from exponential growth to a more Gaussian distribution. His estimated knowledge is now lower and closer to the mean of prior. He might be still the most knowledgable of the teachers since he got the perfect score.

Mikkel's posterior shifted also from chance level towards the mean of prior. However since the estimate was supported with quite a lot of data the prior does not influence the inference that much and therefore the posterior remains quite narrow.

1.4 You go back to your teachers and collect more data (multiply the previous numbers by 100). Do you still see a difference between the results? Why?

Results are shown in figure 5. With increase of amount of data the prior is affecting the inference only weakly. The posteriors of all teachers are narrower now but the means remained the same. In Kristian's case the posterior skyrocketed since he answered all 200 questions correctly, it seems he really is the best one as was suggested by the previous result with only 2 data-points.

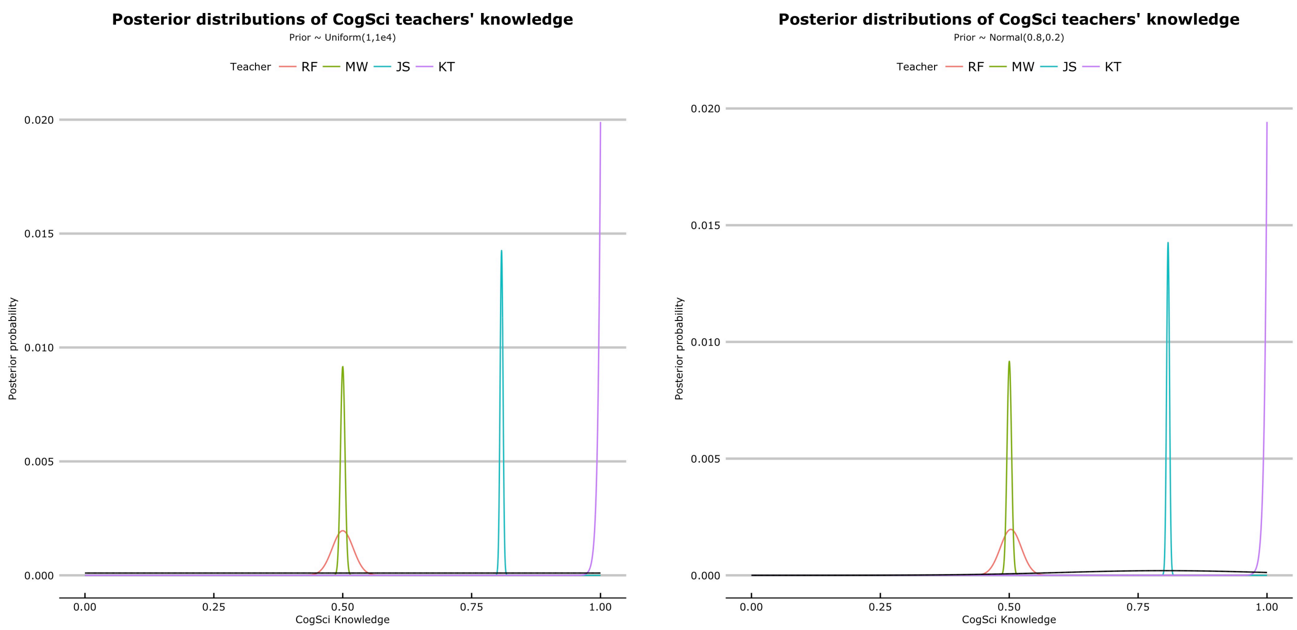

Figure 5: Posteriors resulting from more collected data. Left: a uniform prior; right: a normal prior (mean=0.8, sd=0.2)

1.5 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?

A normal distribution with mean at chance level and very small standard deviation.
prior ~ Normal(0.5,0.01).
The results using this skeptic prior are plotted below (Figure 6).

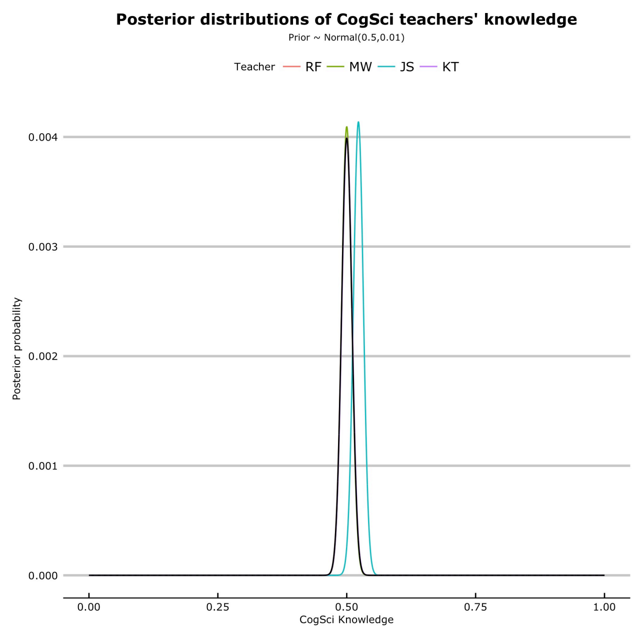

Figure 6: Posterior using skeptic prior

Part 2 Focus on prediction

2.1 Prediction performance in Bayesian VS. frequentist models

As in many comparisons bayesian versus frequentist techniques the main difference is in how the uncertainty around the inference is handled. When some data are predicted using frequentist model we end up with the same number of predictions as the cases in the data, the uncertainty of each prediction is lost. The performance is then measured by looking at difference between the predictions and the actual values i.e computing root mean square error.
When we predict data using the bayesian approach the model returns not only the most likely value but offers all possible outcomes and how probable each of them is in the world of the model. We can then compare not only by how much the prediction is off but also how certain the model is about its prediction and how unlikely the real data is in light of model's knowledge.

2.2 Predicting performance of all teachers

There are two ways of assessing the new data in light of last year's data. Both solutions are described below for all teachers. Solution no. 1 is to use last year's posterior distribution as a prior and update it with the new data. Solution no. 2 is to draw samples from last year's posterior and using these construct a predictive posterior and comparing the new data with the predicted probability of these data.

2.2.1 Riccardo Fusaroli

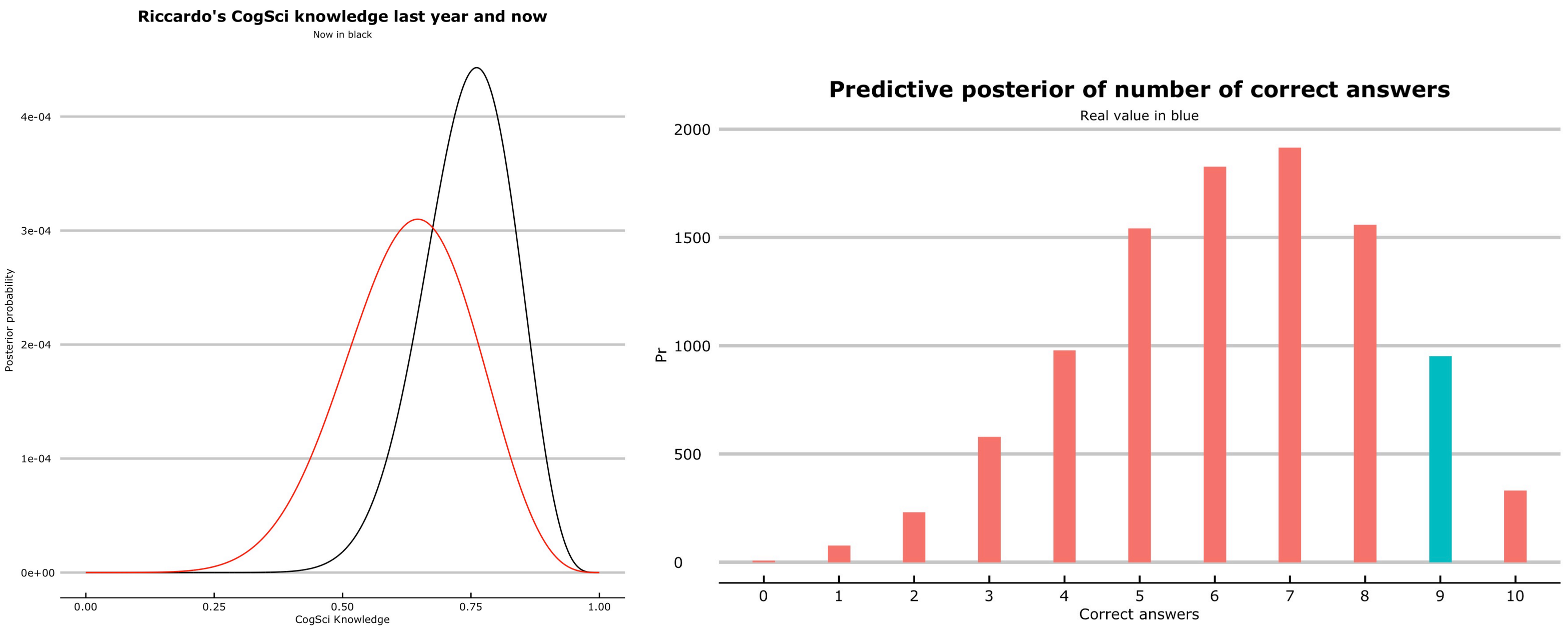

Figure 1: Predicting performance of Riccardo; left: solution no. 1, right: solution no.2

Solution no. 1:
The mode of the posterior distribution shifted from 0.5 to 0.75, the distribution is more narrow and the highest probability is higher than the mode of the prior. That suggests that the old model's prediction would underestimate Riccardo's performance and that the old model was supported with less data than the new one.

Solution no. 2:

By constructing a predictive posterior from the old model we can see that the model would predict 7 correct answers as the most likely performance whereas the real performance is predicted to be approximately only half as likely.

2.2.2 Kristian Tylén

Solution no. 1:
The mode of the posterior distribution shifted from 0.9 to 0.7, the distribution is more narrow and the highest probability is higher than the mode of the prior. That means that the old model would overestimate Kristian's performance and that the old model was supported with less data than the new one.

Solution no. 2:
By constructing a predictive posterior from the old model we can see that the model would predict 12 correct answers as the most likely performance whereas the real performance is predicted to be approximately only half as likely. This happens because in the old data Kristian answered all questions correctly unfortunately there was only two of them.

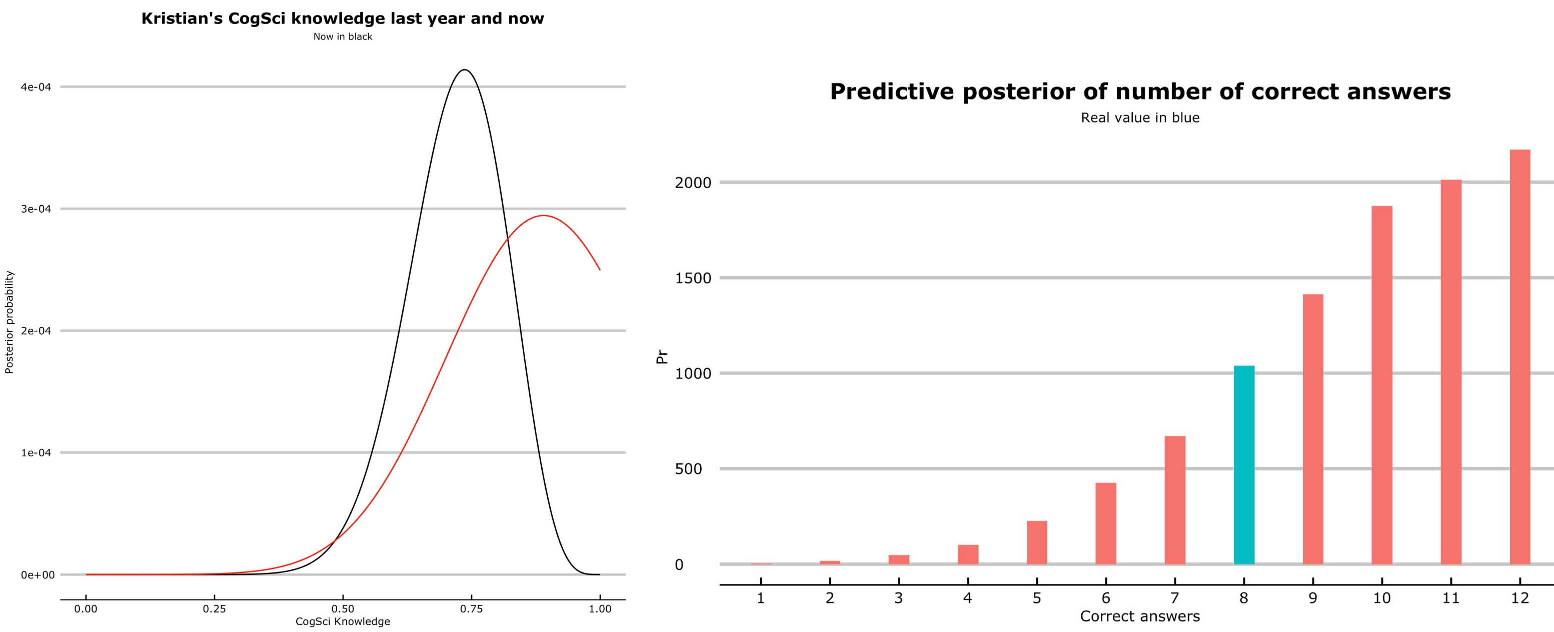

Figure 2: Predicting performance of Kristian; left: solution no. 1, right: solution no.2

2.2.3 Joshua Skewes

Solution no. 1:
The mode of the posterior distribution shifted from 0.85 to 0.9, the distribution is even more narrow and the highest probability is higher than the mode of the prior. That means that the old model would underestimate Josh's performance. The shift of the posterior is smaller than in case of the others because the old model was trained on a lot of data and now we are updating it again with almost the same amount of data.

Solution no. 2:
By constructing a predictive posterior from the old model we can see that the model would predict 140 correct answers as the most likely performance whereas the real performance (148) is predicted to be approximately only half as likely. From the first solution we saw that the model

does not change that much after updating it with the new data but as we are trying to predict big set of questions the spread of the predictions is larger and therefore the prediction is more off.

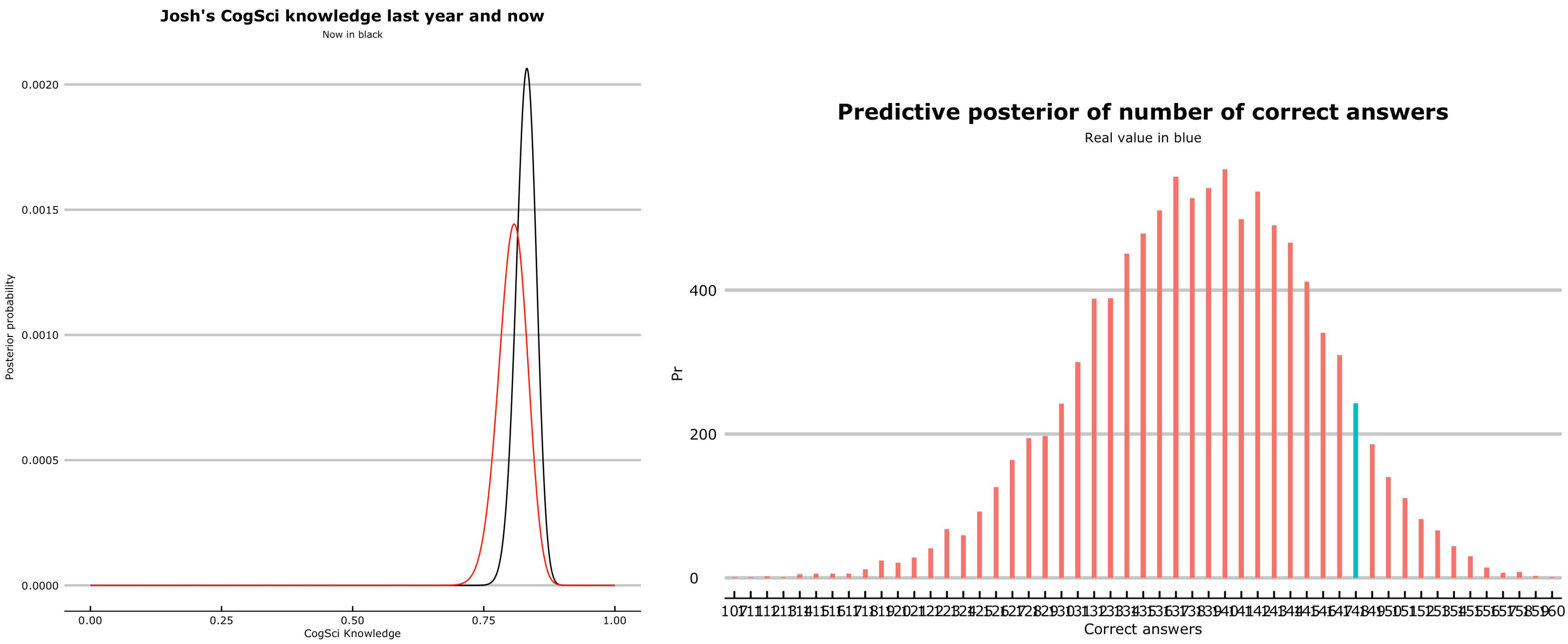

Figure 3: Predicting performance of Josh; left: solution no. 1, right: solution no.2

2.2.4 Mikkel Wallentin

Solution no. 1:
The mode of the posterior distribution did not move at all, the posterior got narrower and the highest probability is higher than the mode of the prior. That means that the new data fit into the old model and made the estimate more confident.

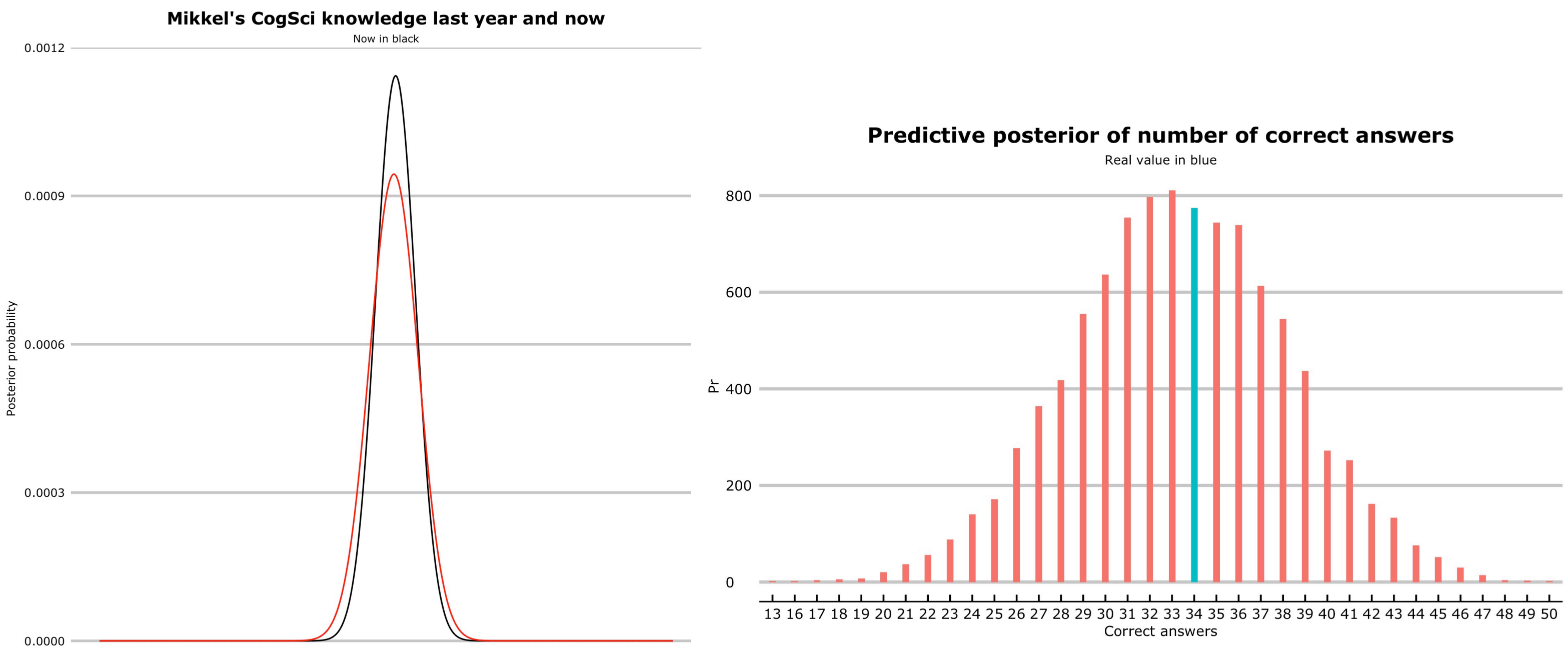

Figure 4: Predicting performance of Mikkel; left: solution no. 1, right: solution no.2

Solution no. 2:
By constructing a predictive posterior from the old model we can see that the model would predict 33 correct answers as the most likely performance but the real value is 34 but the probabilities of these two outcomes are very similar. This can be explained either by the fact that both are generated from very similar parameter values that are within the standard deviation of the posterior. Second explanation is the sampling variance which turns out to be correct because when the predictive posterior is constructed repeatedly the most likely outcome varies between values 33 to 34. The model therefore makes almost perfect prediction.