

Power Analysis Simulation

Anonymised for peer review

Table of contents

| | |
|--|----------|
| BRMS Simulation | 1 |
| Simulation method | 1 |
| Modelling | 2 |
| Number of participants best fit to the alternative model by the number of trials . . . | 3 |
| Number of participants with 95% credible estimates on the independence parameter | 3 |
| Number of participants with 89% credible estimates on the independence parameter | 4 |
| Looking at whether people had a preference for source diversity or repetition | 5 |
| References | 6 |

BRMS Simulation

How many trials do we need to be able to detect an effect of source independence at the individual level?

Simulation method

I simulated these data based on the effect sizes of real participant data from [not cited for anonymity] , a study with much fewer trials per participant but a near identical experimental paradigm otherwise. There were two main manipulations in the original study: 2 consensus direction (i.e., whether the members of the consensus were arguing for or against the claim) X 2 consensus independence (i.e., whether the members of the consensus all shared the same primary source, or all shared different sources). This means there were four “cells” in the original experiment experiment. For each cell, I calculated the mean and standard deviation belief update (the difference between their 1-100 belief in the claim before versus after seeing the consensus). Using these means and standard deviations, I simulated either 3, 6, 9, 12, 16, or 20 trials per cell.

In this simulation study we were only interested in independent trials compared to dependent trials, leaving out any comparisons to a contested consensus. This was because we expected contested comparisons to have a much larger effect size, so if there was sufficient power for the independent v dependent comparison, there should be more than enough power to detect a contested v independent comparison.

Modelling

The goal of this simulation was to fit separate Bayesian linear models to each participant. Usually, what we do is fit these models to all of the participants, with participant as a random effect, in order to identify group level consensus effects, but this isn't really able to capture individual differences.

I kept the modelling reasonably simple, with a null model that only considers the consensus direction:

```
post ~ prior
```

Against an alternative model that also considers source independence:

```
post ~ prior + independence
```

One thing to keep in mind is that we applied a transformation on the prior and post scores in the “con” condition, where the tweets were opposing the claim. This is because in those conditions, we would expect the prior score to be greater than the post score, since their belief in the score should be lower. I achieved this by subtracting the prior and post scores from 100 (the maximum score), such that:

```
prior = 100 - original prior
```

```
post = 100 - original post
```

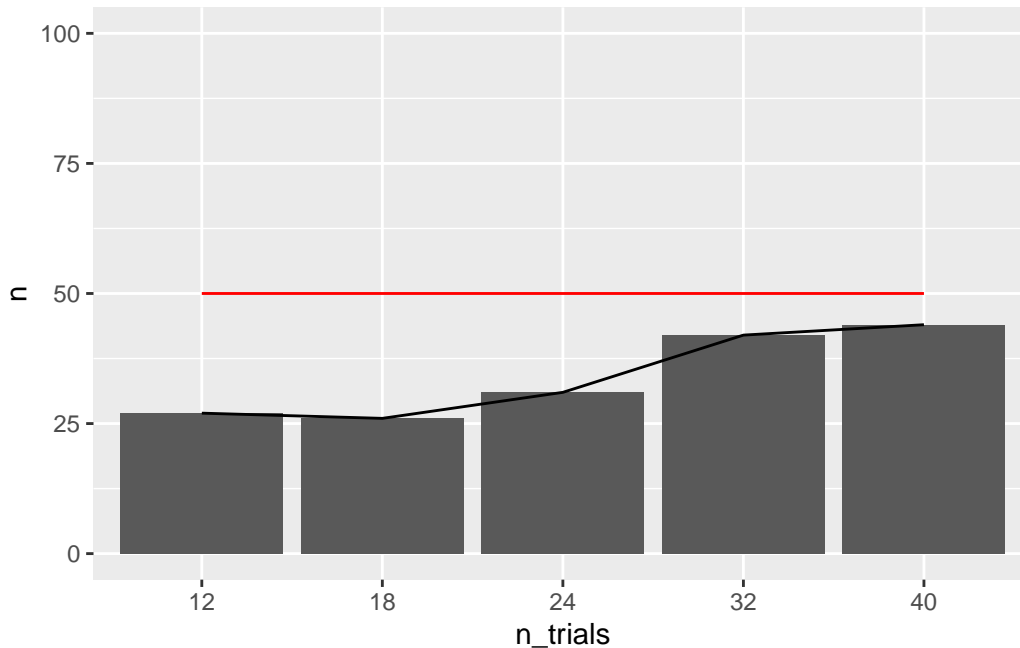
We applied this transformation rather than controlling for consensus direction as a predictor in the model because having fewer variables increases the power of the analysis.

The general goal here was to see at which point increasing more trials does not appear to increase the number of participants best fit to the alternative model. Once that plateau begins to occur, then we should have a good idea of how many trials per participants we will need.

Number of participants best fit to the alternative model by the number of trials

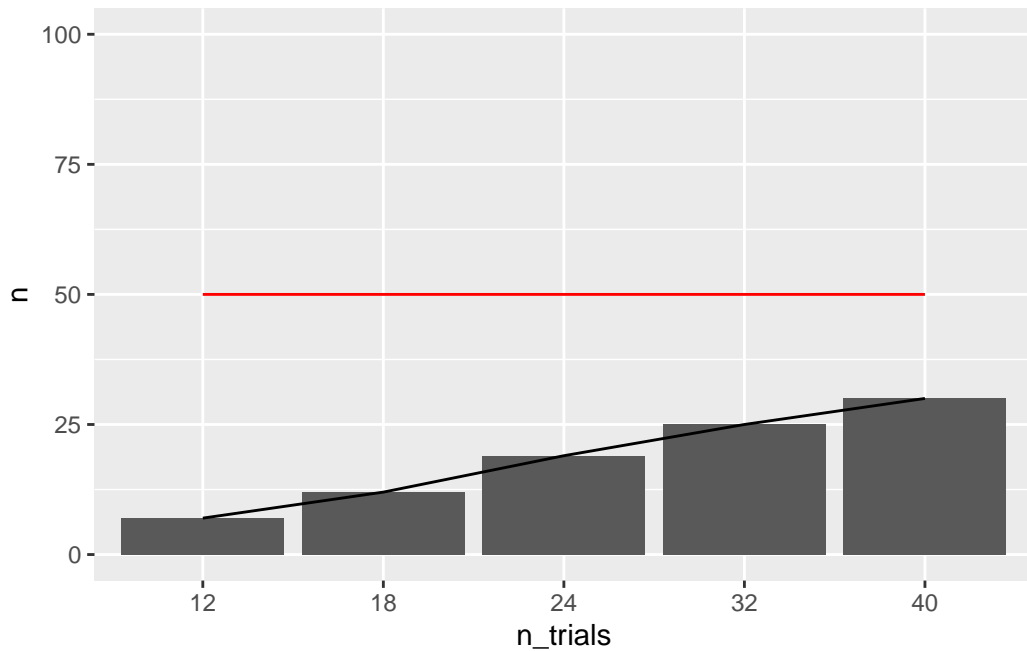
Using LOOIC, this analysis compared which participants were best fit to the alternative model versus the null model.

The red line indicates a very rough estimate for the number of participants we thought might be best fit by an alternative model. We calculated this by participants who, in the [anonymous] et al. study, consistently had responses in one source dependence condition higher than their median response (so the number of participants who had $> 50\%$ of their responses for one kind of source larger than the median score).



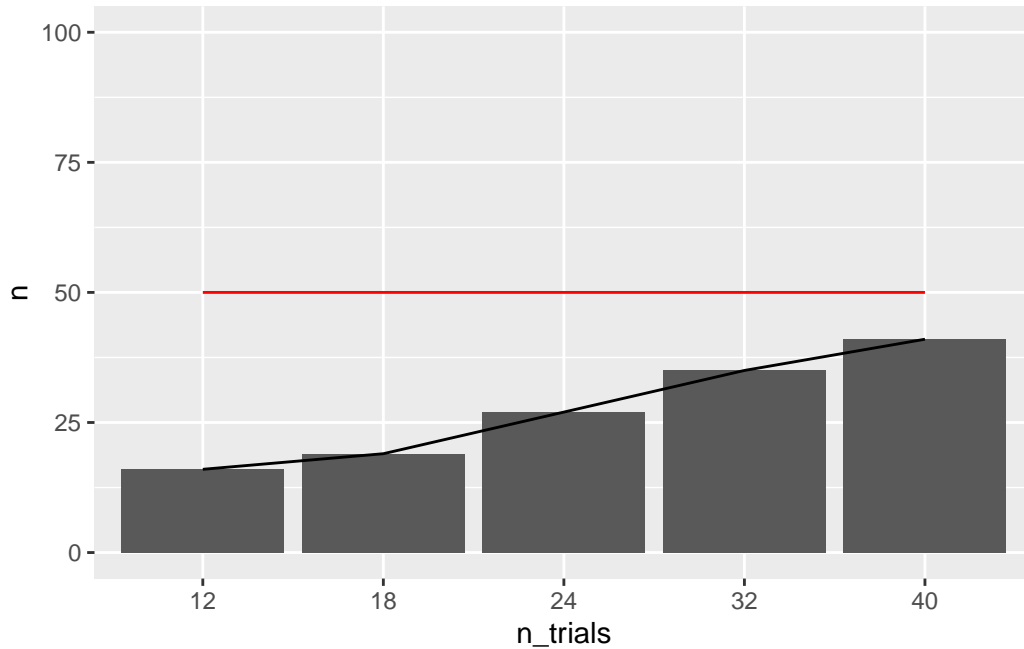
Number of participants with 95% credible estimates on the independence parameter

Another way to evaluate consensus independence is to look at the estimate and associated credible intervals of the independence parameter.



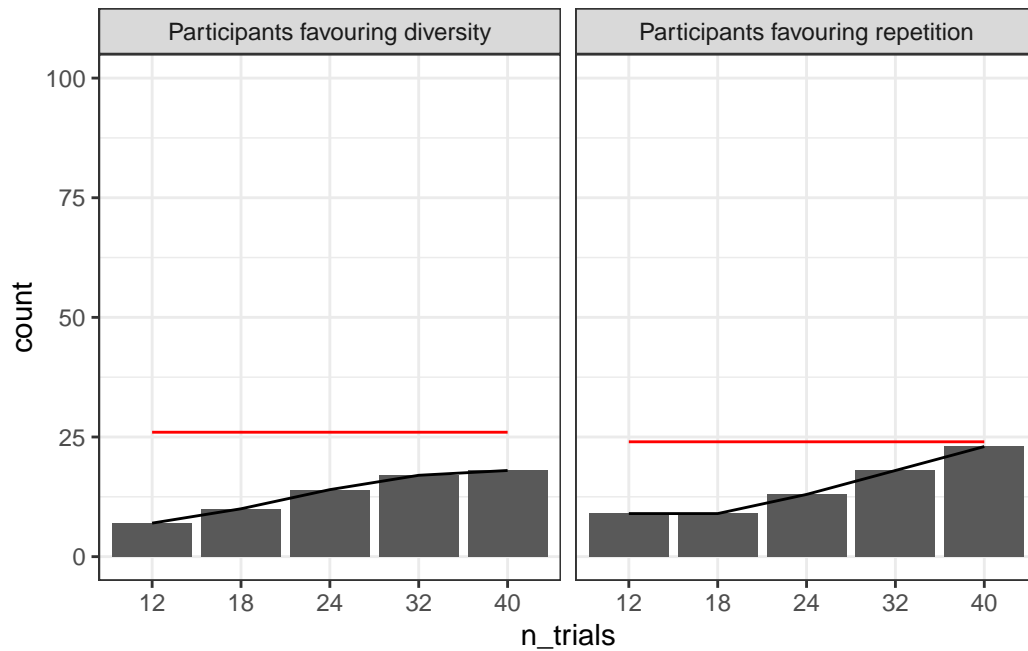
Number of participants with 89% credible estimates on the independence parameter

It has been argued that a 95% credible interval is not actually the most appropriate for Bayesian credible intervals, since unless you have an effective sample size (ESS) of at least 10,000, a 95% credible interval is not very stable (Kruschke 2014). An alternate interval is the 89% credible interval, which is much more stable for the ESS that we will have in our study which is around 4000 (Kruschke (2014)).



Looking at whether people had a preference for source diversity or repetition

A credible estimate of independence, or a model comparison favoring the alternative model only tells us whether there was an effect of consensus independence, but it does not tell us whether people preferred independent (diverse sources) or dependent (repeated) sources. People could reasonably have a preference for either, since diversity indicates more of the hypothesis spaces has been searched, whereas repetition might suggest that the source is more reliable, since it has been endorsed by more people. We can determine whether people had a preference for either by looking at the sign of the independence parameter. If it is positive, that person had a preference for diversity. If it was negative that person had a preference for repetition.



It looks like with 40 trials (20 trials per cell), we get pretty close to the maximum estimate for each condition. It's a bit worse for the independence condition, but based on the trend it doesn't really look like adding more trials will result in more people having credible estimates.

References

Kruschke, John. 2014. *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Academic Press.