Larrisa Miller
PSYCH 740 - Bayesian Statistics
Professor Jeffrey Starns
December 13th, 2021

Analysis Project

This analysis is based on data collected by UMass Amherst Psychological and Brain Sciences PhD student Kuan-Jung Huang. His experiment investigates grammatical transpositions of open and closed class words within sentences. “C” represents a closed-class (function) word, “O” represents an open-class (content) word, and the numbers 3 and 4 indicate the place of the word in the sentence at which a grammatical mishap occurs. The overarching research questions in this project are:

RQ1: Is the proportion correct for ungrammatical CO3 sentences different than .5?
RQ2: Is there a difference between OC3 and OO3 proportion correct?

For this analysis, I have been provided with multiple datasets for each overarching research question. Each dataset is categorized by research question and version number. For example, RQ1V1 is equivalent to Research Question 1, Version 1. There are four versions of an original and replication dataset for each research question, totalling 16 datasets to be analyzed.

Results

This analysis uses Bayesian t-tests and Bayesian parameter estimation. Bayesian inference is the re-allocation of credibility across possibilities (Krushke 2013; Krushke 2015). Using prior estimates and observed data, Bayesian estimation “reallocate[s] credibility to parameter values that best accommodate the observed data (Krushke 2013 p. 576).” A t-test tells how far a particular value is from a hypothesized population value in units of standard deviation (Starns 2021). In Bayesian analysis, a t-test value is used to calculate the Bayes Factor, or the “relative probability of observed data under alternative theoretical claims (Starns 2021).” Essentially, the Bayes Factor produces the likelihood that the alternative hypothesis is true rather than the null hypothesis. Under the alternative hypothesis, an effect is observed, whereas under the null hypothesis, there is no effect. In Bayesian parameter estimation, prior estimated parameters and observed data are combined to produce posterior distributions representing one’s uncertainty about the true parameter value(s) after seeing the data.

RQ1V1

Prior Distribution and Justification

Based on running through Kuan-Jung’s experiment myself, I felt that it was relatively easy to detect ungrammatical sentences. Thus, I expected there to definitely be an effect, which would indicate a mean different than 0.5. I used a prior distribution of mu = 0.75, indicating that participants would correctly detect the ungrammatical sentence ~ 75% of the time, and a sigma of 0.4 to give me error room. This produces an effect size of 0.625, which is medium strength. I used a Gaussian normal distribution for ease of visualization, with a distribution of effect sizes from -1.5 to 1.5. This prior, mu = 0.75 and sigma = 0.4, is used for all RQ1 versions.

For the replication data set, I used a prior of mu = 0.925 and sigma = 0.228 based on Bayesian parameter estimation using my original priors (mu = 0.75, sigma = 0.4) and the original dataset provided for RQ1V1. In this estimation I weighted the prior values by the posterior probability, calculated using the Bayes Factor. This weighting mechanism is used for all versions of RQ1.

Descriptive Statistics

For all of the RQ1 variants, observed standardized effect size is calculated as (mu - 0.5)/ sigma.

Initial Dataset

Mean	0.705
Standard Deviation	0.206
Observed Standardized Effect Size	0.971

Replication Dataset

Mean	0.605
Standard Deviation	0.231
Observed Standardized Effect Size	0.455

Inferential Analysis

Initial Dataset

	t-value	4.442
	Bayes Factor	496.716
	Posterior Probability of Effect	0.998
Posterior Distribution	**Mean**	0.925
Posterior Distribution	**Standard Deviation**	0.223

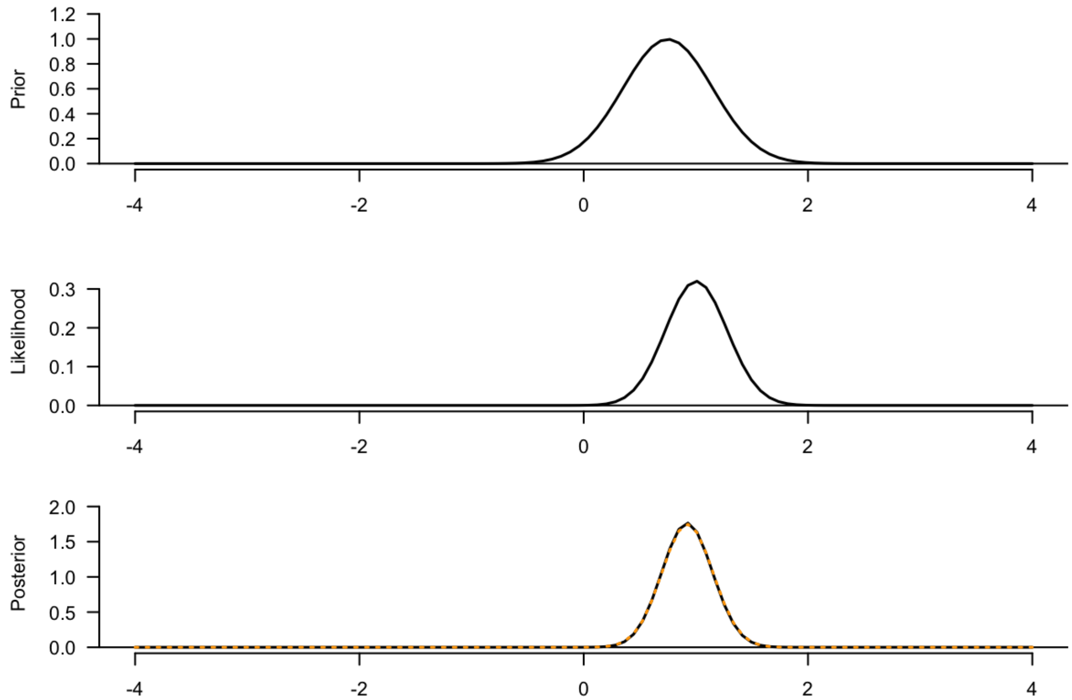

Figure 1: Graph depicting the prior, likelihood, and posterior distributions of the initial dataset for RQ1V1. The x value is effect size, and the y value is probability.

Replication Dataset

	t-value	2.037
	Bayes Factor	1.809
	Posterior Probability of Effect	0.644
Posterior Distribution	**Mean**	0.701
Posterior Distribution	**Standard Deviation**	0.160

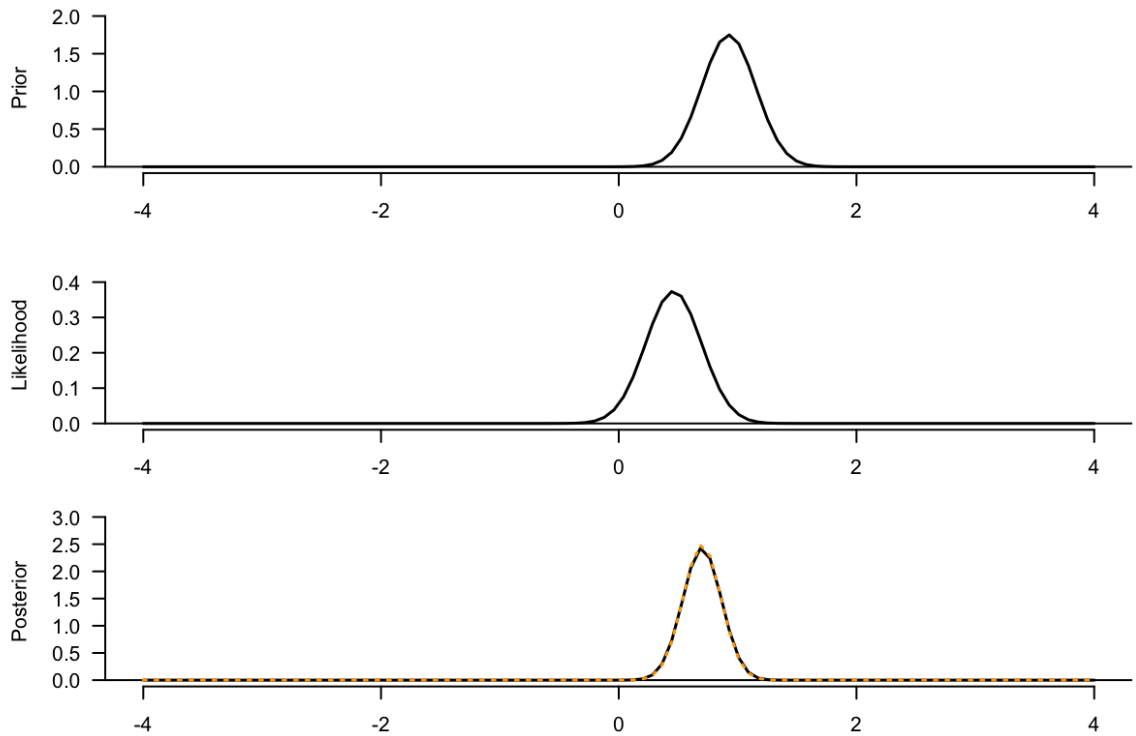

Figure 2: Graph depicting the prior, likelihood, and posterior distributions of the replication dataset for RQ1V2. The x value is effect size, and the y value is probability.

Conclusions

Looking at the original and replication datasets together, I believe that this procedural variant does have an effect. The mean for both datasets themselves as well as the posterior means are all greater than 0.5. Most importantly, the Bayes factor for the initial dataset is 496.716, giving a post probability of 99.8% in favor of the alternative. While the replication dataset has a smaller Bayes factor and lower probability of effect (64%), it is not so small that it undermines the initial dataset's results. The observed standard effect sizes are 0.455 - 0.971. With this combination of effect sizes and Bayes factors, I am concluding that this variant has at least a medium to large level effect.

RQ1V2

Prior Distribution and Justification

For the original dataset, I used my initial prior of mu = 0.75 and sigma = 0.4. For the replication dataset, I used priors of mu = 1.197 and sigma = 0.252 based on Bayesian parameter estimation using the initial priors and the original dataset provided for RQ1V2.

Descriptive Statistics

Initial Dataset

Mean	0.765

Standard Deviation	0.179
Observed Standardized Effect Size	1.480

Replication Dataset

Mean	0.795
Standard Deviation	0.214
Observed Standardized Effect Size	1.378

Inferential Analysis

Initial Dataset

	t-value	6.639
	Bayes Factor	24,338.526
	Posterior Probability of Effect	1.000
Posterior Distribution	**Mean**	1.197
Posterior Distribution	**Standard Deviation**	0.252

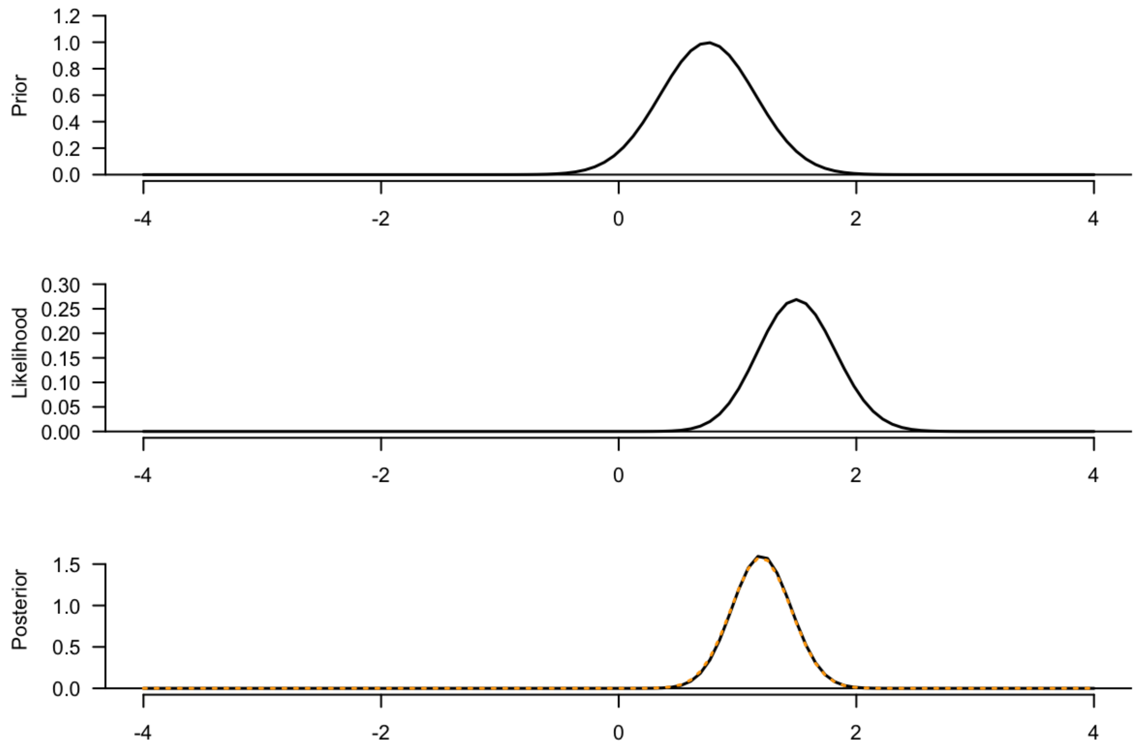

Figure 3: Graph depicting the prior, likelihood, and posterior distributions of the initial dataset for RQ1V2. The x value is effect size, and the y value is probability.

Replication Dataset

	t-value	6.167
	Bayes Factor	29,271.951
	Posterior Probability of Effect	1.00
Posterior Distribution	**Mean**	1.269
Posterior Distribution	**Standard Deviation**	0.196

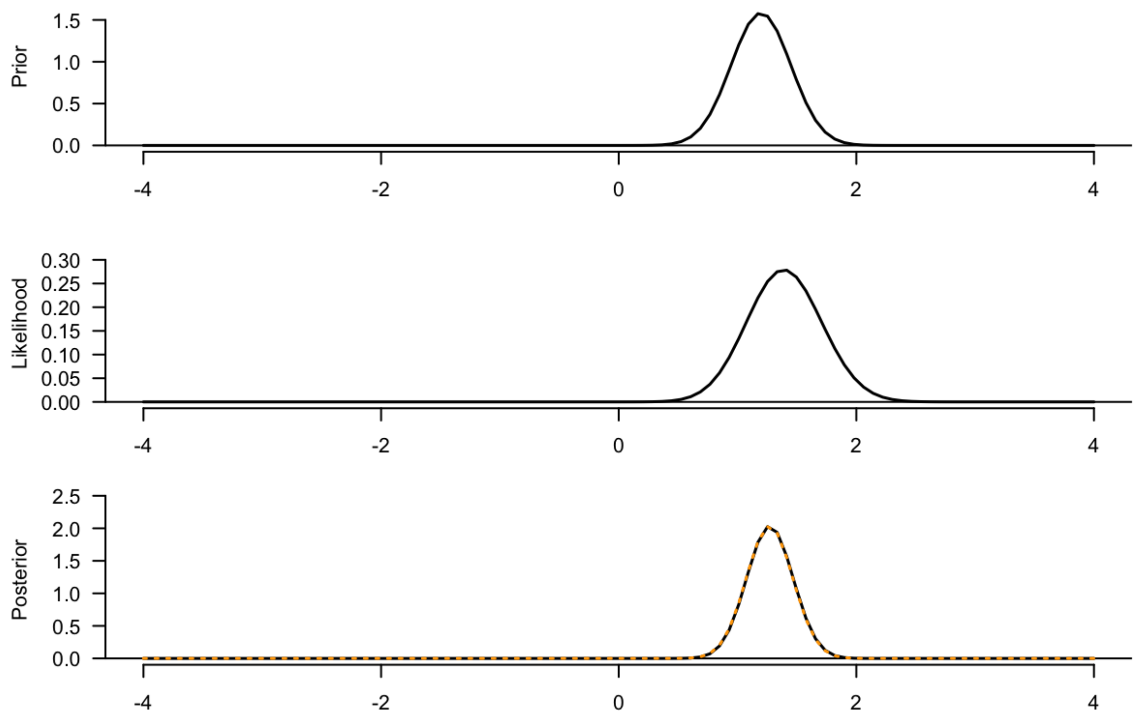

Figure 4: Graph depicting the prior, likelihood, and posterior distributions of the replication dataset for RQ1V2. The x value is effect size, and the y value is probability.

Conclusions

Looking at the original and replication datasets together, I believe that this procedural variant also has an effect. As in variant 1, all of the reported means are greater than 0.5. More critically, the Bayes factors for both the initial and replication datasets are greater than 24,000, with posterior probability of effect reaching 1. The observed standardized effect sizes are also greater than 1. With Bayes factors and probabilities this large, I am confident that there is a strong effect in this variant.

RQ1V3

Prior Distribution and Justification

For the original dataset, I used my initial prior of mu = 0.75 and sigma = 0.4. For the replication dataset, I used priors of mu = 0.565 and sigma = 0.128 based on Bayesian parameter estimation using the initial priors and the original dataset provided for RQ1V3.

Descriptive Statistics

Initial Dataset

Mean	0.570
Standard Deviation	0.215

Observed Standardized Effect Size	0.326

Replication Dataset

Mean	0.540
Standard Deviation	0.244
Observed Standardized Effect Size	0.164

Inferential Analysis

Initial Dataset

	t-value	1.453
	Bayes Factor	0.917
	Posterior Probability of Effect	0.478
Posterior Distribution	**Mean**	0.436
Posterior Distribution	**Standard Deviation**	0.200

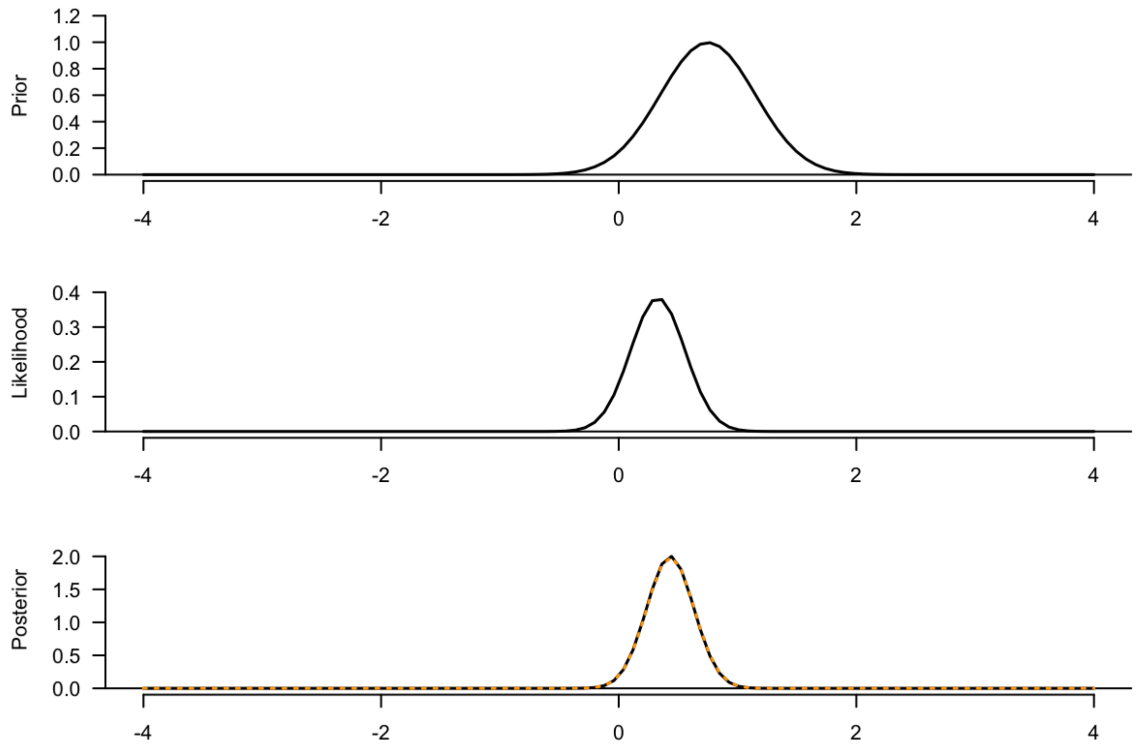

Figure 5: Graph depicting the prior, likelihood, and posterior distributions of the initial dataset for RQ1V3. The x value is effect size, and the y value is probability.

Replication Dataset

	t-value	0.734
	Bayes Factor	0.349
	Posterior Probability of Effect	0.259
Posterior Distribution	**Mean**	0.469
Posterior Distribution	**Standard Deviation**	0.112

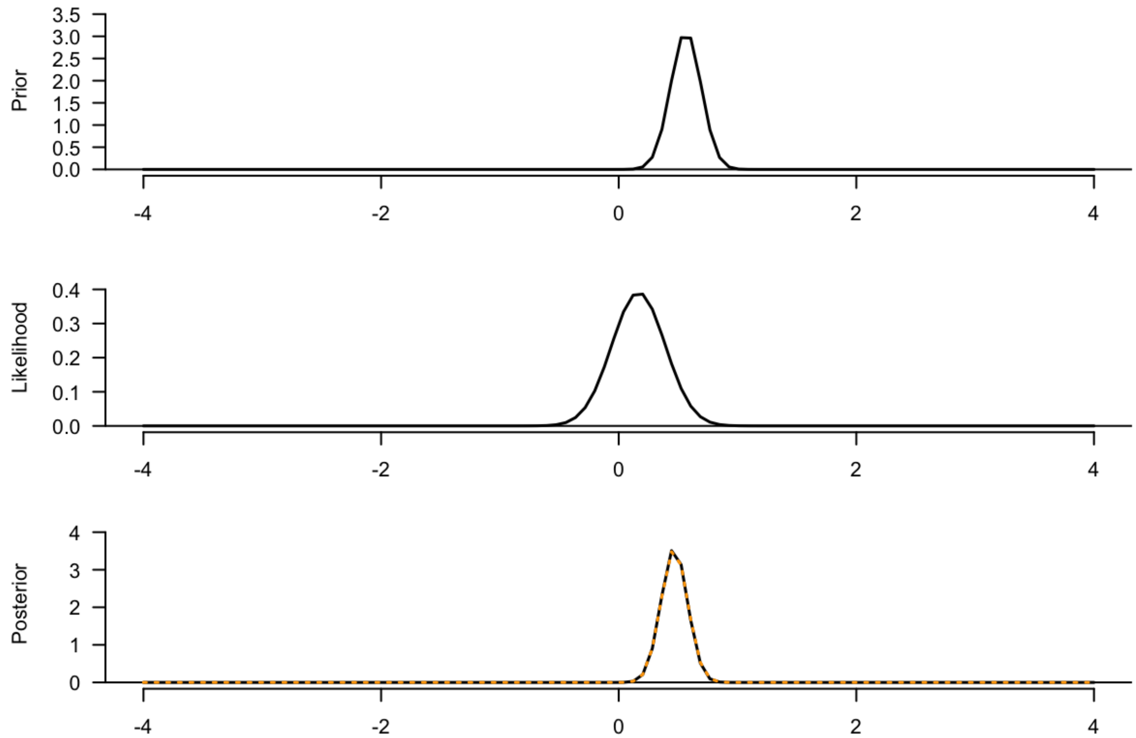

Figure 6: Graph depicting the prior, likelihood, and posterior distributions of the replication dataset for RQ1V3. The x value is effect size, and the y value is probability.

Conclusions

Looking at the original and replication datasets together, I believe that this procedural variant does *not* have an effect. All of the reported means are very close to 0.5, and both the initial and replication Bayes factors are under 1, with probabilities for there being an effect both less than 50%. If there were to be an effect, it would be quite small, given the low observed standardized effect sizes.

RQ1V4

Prior Distribution and Justification

For the original dataset, I used my initial prior of mu = 0.75 and sigma = 0.4. For the replication dataset, I used priors of mu = 0.501 and sigma = 0.128 based on Bayesian parameter estimation using the initial priors and the original dataset provided for RQ1V4.

Descriptive Statistics

Initial Dataset

Mean	0.540
Standard Deviation	0.258
Observed Standardized Effect Size	0.155

Replication Dataset

Mean	0.460
Standard Deviation	0.193
Observed Standardized Effect Size	-0.207

Inferential Analysis

Initial Dataset

	t-value	0.693
	Bayes Factor	0.271
	Posterior Probability of Effect	0.213
Posterior Distribution	**Mean**	0.300
Posterior Distribution	**Standard Deviation**	0.196

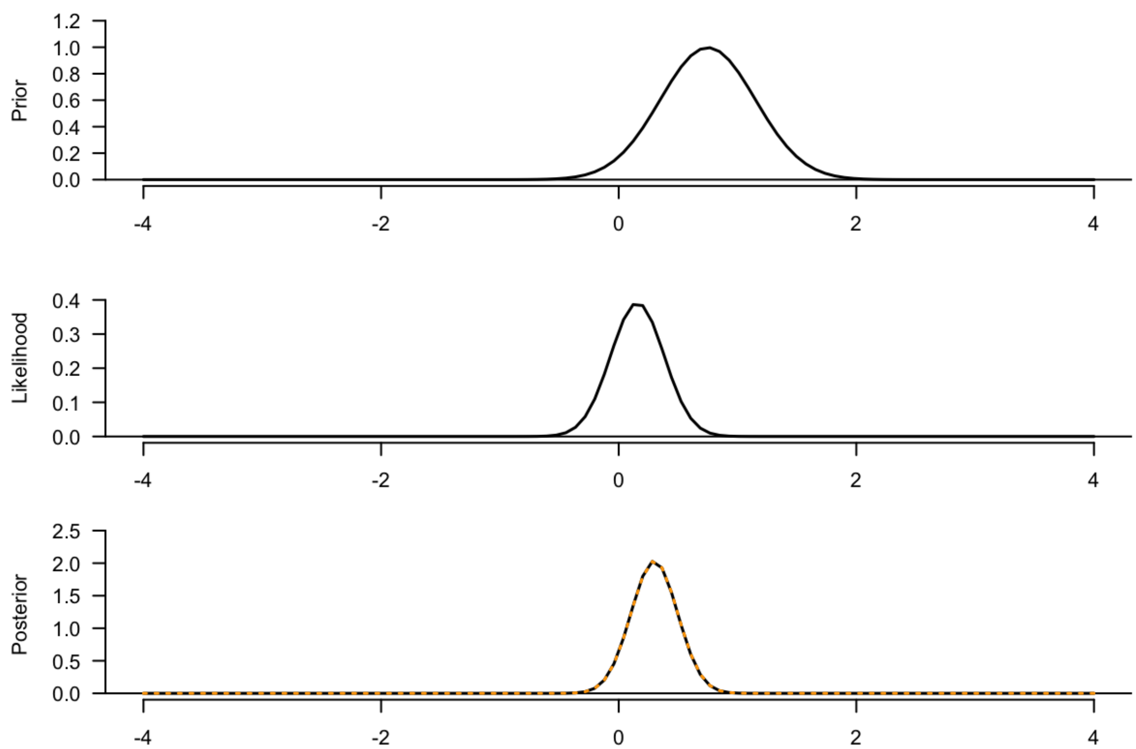

Figure 7: Graph depicting the prior, likelihood, and posterior distributions of the initial dataset for RQ1V4. The x value is effect size, and the y value is probability.

Replication Dataset

	t-value	-0.927
	Bayes Factor	0.032
	Posterior Probability of Effect	0.031
Posterior Distribution	**Mean**	0.332
Posterior Distribution	**Standard Deviation**	0.108

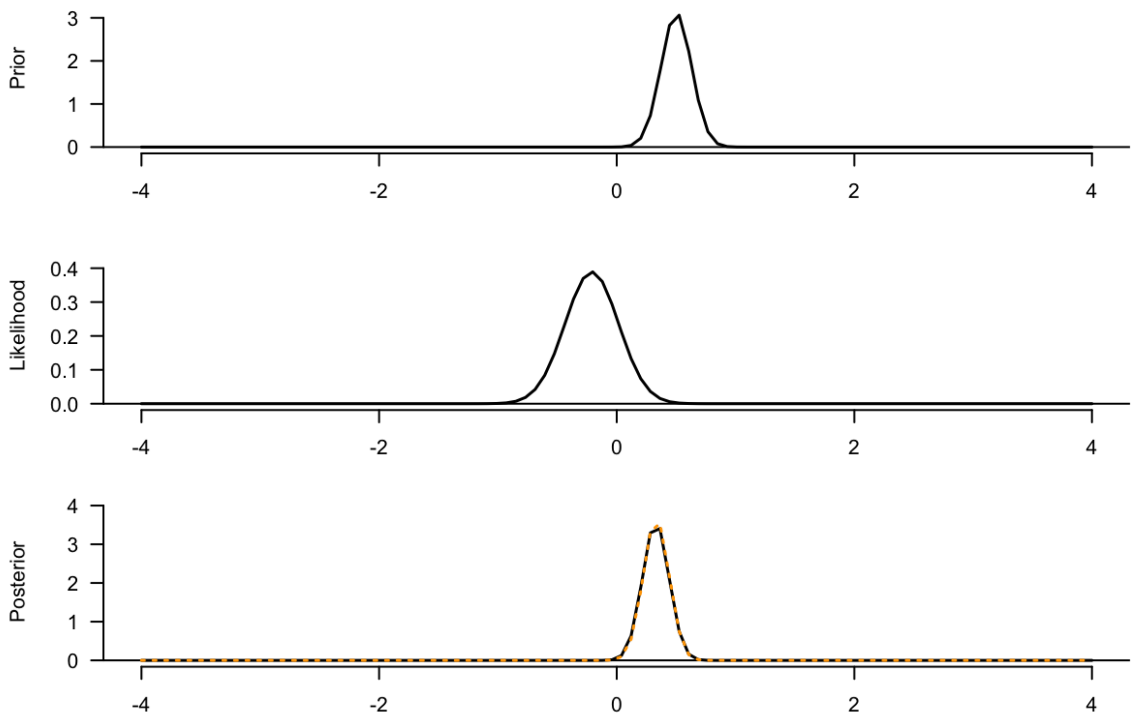

Figure 8: Graph depicting the prior, likelihood, and posterior distributions of the replication dataset for RQ1V4. The x value is effect size, and the y value is probability.

Conclusions

Looking at the original and replication datasets together, I believe that this procedural variant does *not* have an effect. All of the reported means are close to or below 0.5. The Bayes factors are both far below 0, and the replication post probability effect is just 3.1% in favor of the alternative. If there were to be an effect, it would be quite small, given the low observed standardized effect sizes.

RQ2V1

Prior Distribution and Justification

Based on running through Kuan-Jung's experiment myself, I felt that it was quite difficult to differentiate between OO3 and OC3. It seems that OO3 may be slightly easier to detect, so I set my mean to reflect OO3-OC3, or in favor of OO3. However, I was quite uncertain of the differences between these conditions, so I set my mu to be quite small, at 0.2. Based on this uncertainty, I set my prior sigma to be 0.5. This gives an effect size of 0.4. I used a Gaussian normal distribution for ease of visualization with the effect size distribution set from -2 to 2. This prior (mu and sigma) is used for all RQ2 versions. ***All calculations for RQ2 are in the order of OO3-OC3.***

For the replication data set, I used a prior of mu = 0.501 and sigma = 0.192 based on Bayesian parameter estimation using my original priors (mu = 0.2, sigma = 0.5) and the original dataset provided for RQ2V1. In this estimation I weighted the prior values by the posterior probability, calculated using the Bayes Factor. This weighting mechanism is used for all versions of RQ2.

Descriptive Statistics

Initial Dataset

Mean	0.140
Standard Deviation	0.196
Observed Standardized Effect Size	0.714

Replication Dataset

Mean	0.120
Standard Deviation	0.208
Observed Standardized Effect Size	0.577

Inferential Analysis

Initial Dataset

	t-value	3.573
	Bayes Factor	12.001
	Post Probability of Effect	0.923
Posterior Distribution	**Mean**	0.501
Posterior Distribution	**Standard Deviation**	0.192

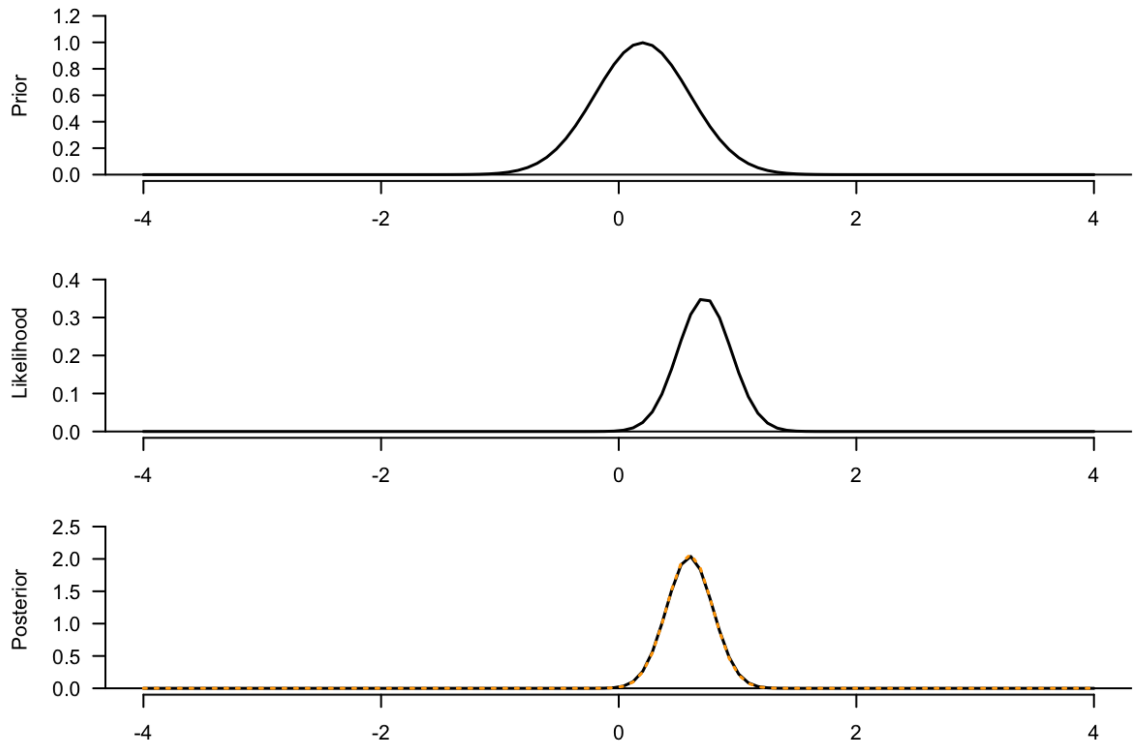

Figure 9: Graph depicting the prior, likelihood, and posterior distributions of the original dataset for RQ2V1. The x value is effect size, and the y value is probability.

Replication Dataset

	t-value	2.882
	Bayes Factor	27.3076
	Posterior Probability of Effect	0.965
Posterior Distribution	**Mean**	0.533
Posterior Distribution	**Standard Deviation**	0.140

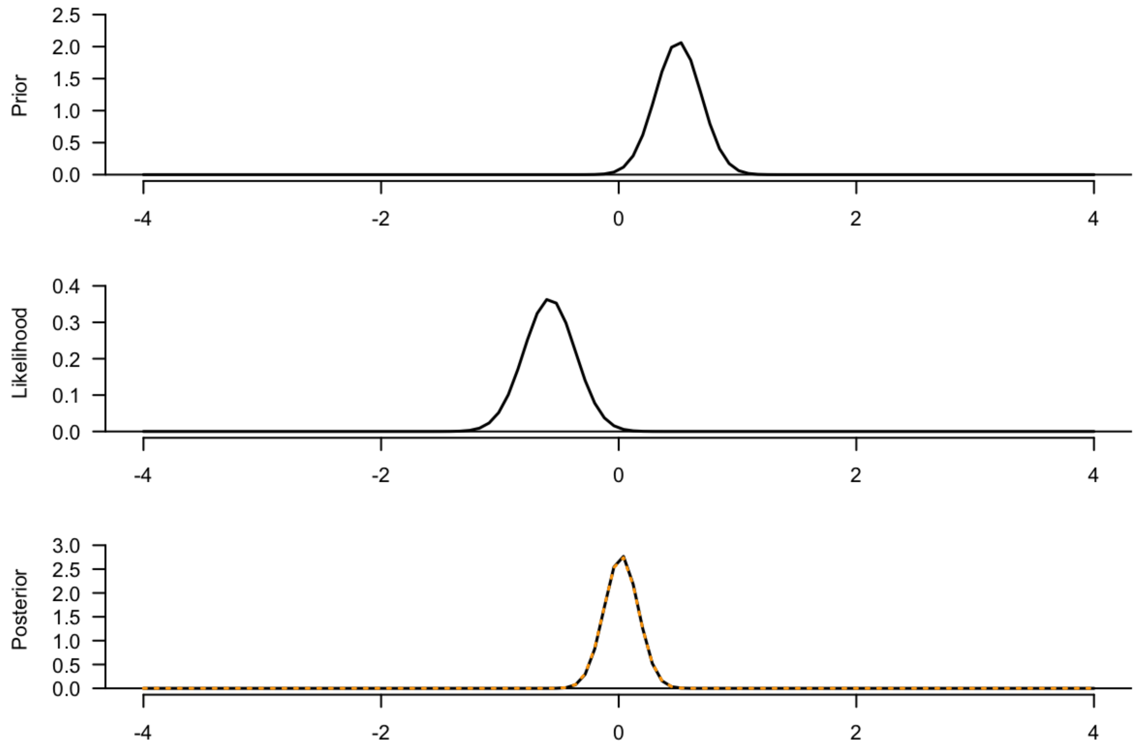

Figure 10: Graph depicting the prior, likelihood, and posterior distributions of the replication dataset for RQ2V1. The x value is effect size, and the y value is probability.

Conclusions

Looking at the original and replication datasets together, I believe that this procedural variant has an effect. The Bayes factor for both initial and replication datasets are sizable, with posterior probability of effect at 92.3% and 96.5%. The observed standardized effect sizes are between 0.5 and 0.7. Given these values, I believe that there is a medium to large effect towards the OO3 condition in this variant.

RQ2V2

Prior Distribution and Justification

For the original dataset, I used my initial prior of mu = 0.2 and sigma = 0.5. For the replication dataset, I used priors of mu = 0.637 and sigma = 0.204 based on Bayesian parameter estimation using the initial priors and the original dataset provided for RQ2V2.

Descriptive Statistics

Initial Dataset

Mean	0.164
Standard Deviation	0.178
Observed Standardized Effect Size	0.921

Replication Dataset

Mean	0.164
Standard Deviation	0.152
Observed Standardized Effect Size	1.079

Inferential Analysis

Initial Dataset

	t-value	4.615
	Bayes Factor	62.902
	Posterior Probability of Effect	0.984
Posterior Distribution	**Mean**	0.637
Posterior Distribution	**Standard Deviation**	0.204

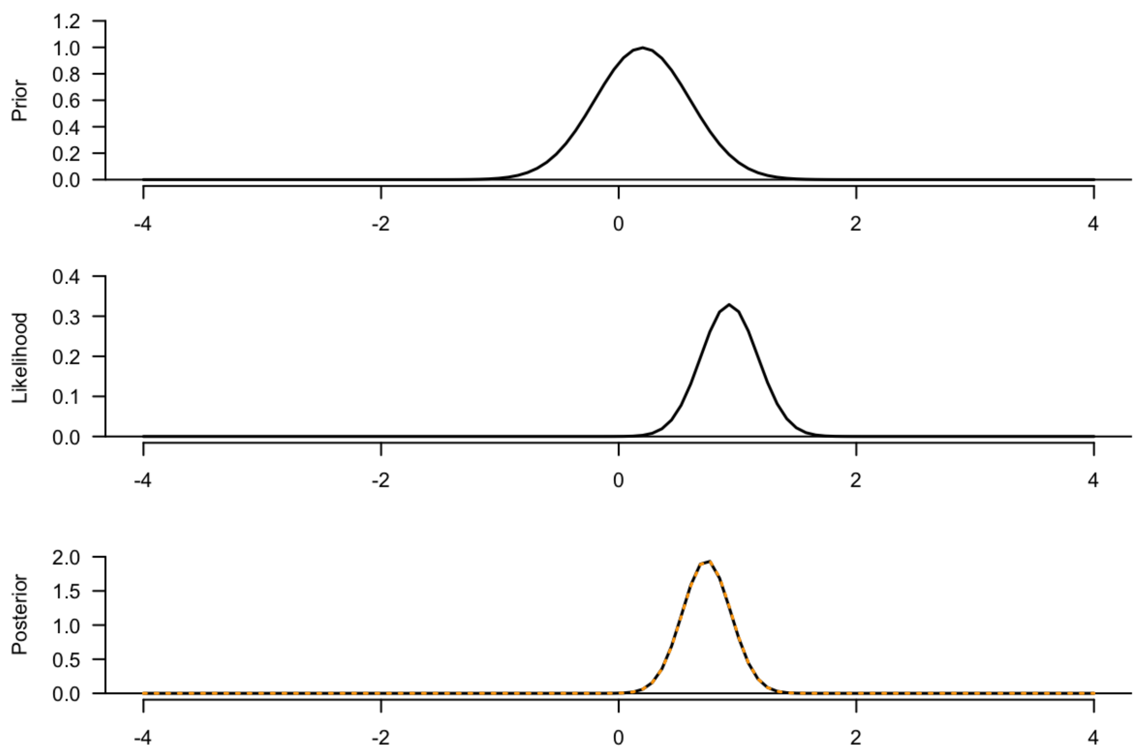

Figure 11: Graph depicting the prior, likelihood, and posterior distributions of the initial dataset for RQ2V2. The x value is effect size, and the y value is probability.

Replication Dataset

	t-value	5.380
	Bayes Factor	4,663.350
	Posterior Probability of Effect	0.100
Posterior Distribution	**Mean**	0.813
Posterior Distribution	**Standard Deviation**	0.156

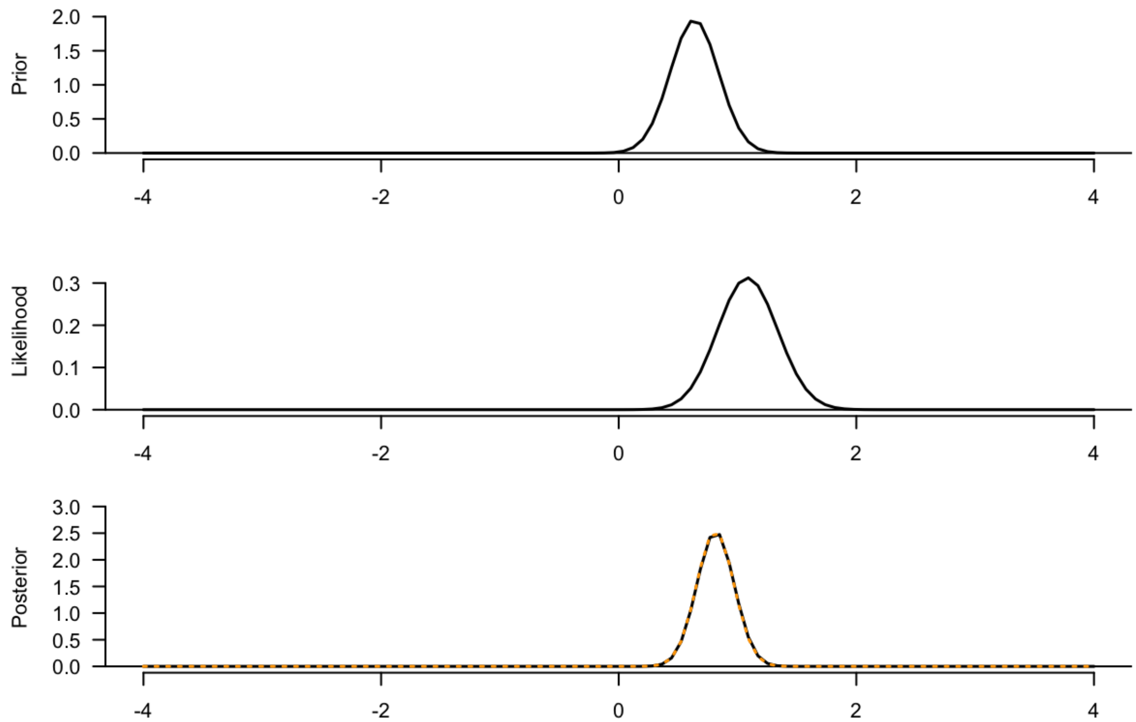

Figure 12: Graph depicting the prior, likelihood, and posterior distributions of the replication dataset for RQ2V2. The x value is effect size, and the y value is probability.

Conclusions

Looking at the original and replication datasets together, I believe that this procedural variant has an effect. The Bayes factor for both initial and replication datasets are quite sizable, with posterior probability of effect at 98% to 100%. The observed standardized effect sizes are near 1. Given these values, I believe that there is a large effect towards the OO3 condition in this variant.

RQ2V3

Prior Distribution and Justification

For the original dataset, I used my initial prior of mu = 0.2 and sigma = 0.5. For the replication dataset, I used priors of mu = 0.396 and sigma = 0.116 based on Bayesian parameter estimation using the initial priors and the original dataset provided for RQ2V3.

Descriptive Statistics

Initial Dataset

Mean	0.096
Standard Deviation	0.181
Observed Standardized Effect Size	0.530

Replication Dataset

Mean	0.104
Standard Deviation	0.124
Observed Standardized Effect Size	0.839

Inferential Analysis

Initial Dataset

	t-value	2.646
	Bayes Factor	2.914
	Posterior Probability of Effect	0.745
Posterior Distribution	**Mean**	0.372
Posterior Distribution	**Standard Deviation**	0.188

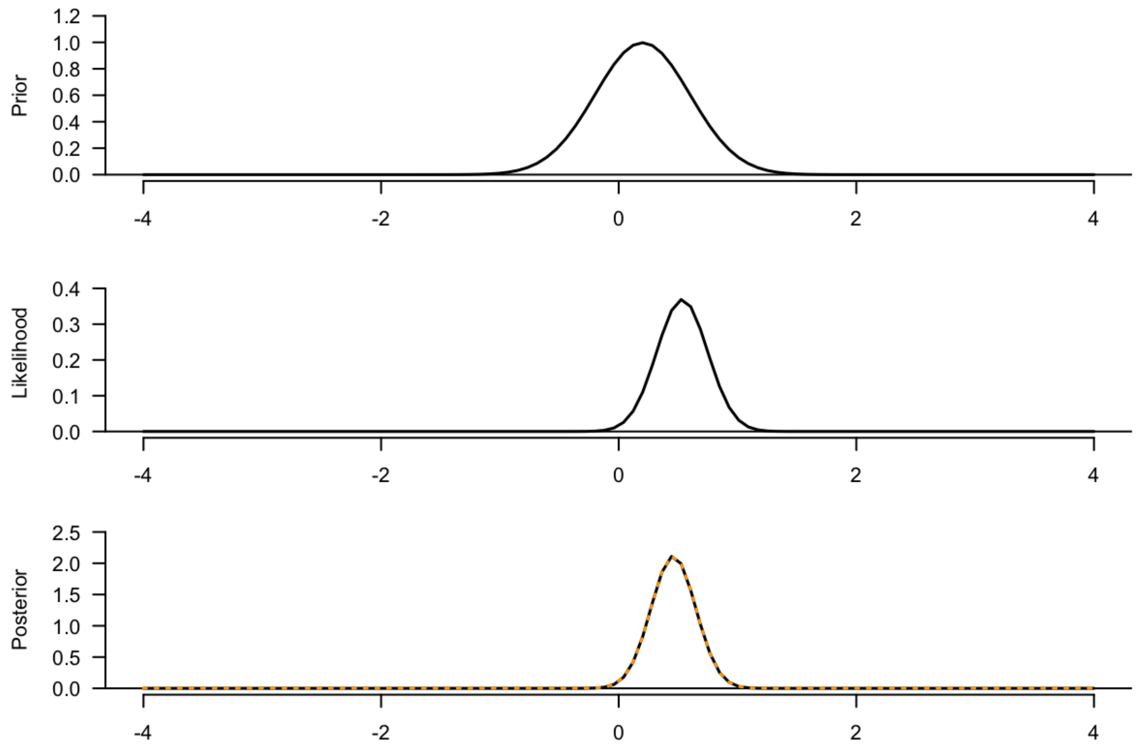

Figure 13: Graph depicting the prior, likelihood, and posterior distributions of the initial dataset for RQ2V3. The x value is effect size, and the y value is probability.

Replication Dataset

	t-value	4.190
	Bayes Factor	164.425
	Posterior Probability of Effect	0.994
Posterior Distribution	**Mean**	0.485
Posterior Distribution	**Standard Deviation**	0.104

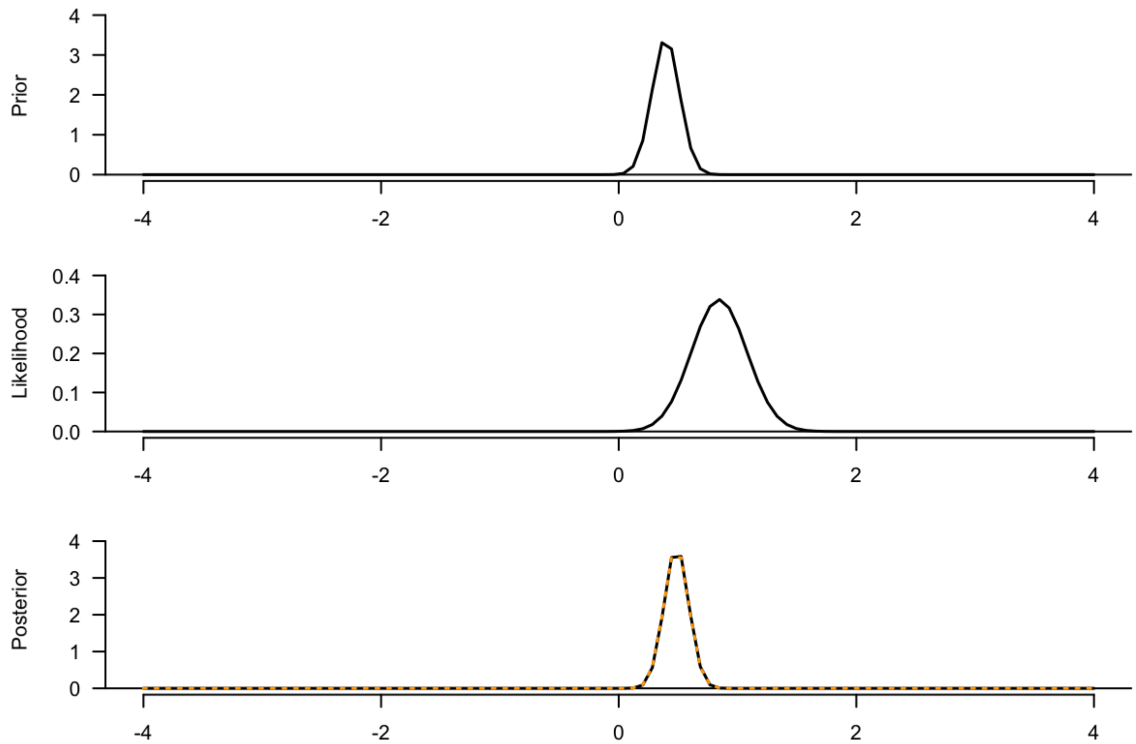

Figure 14: Graph depicting the prior, likelihood, and posterior distributions of the replication dataset for RQ2V3. The x value is effect size, and the y value is probability.

Conclusions

Looking at the original and replication datasets together, I believe that this procedural variant has an effect. The Bayes factor for both initial and replication datasets are sizable, with posterior probability of effect between 74% and 99%. The observed standardized effect sizes are between 0.5 and 0.8. Given these values, I believe that there is a medium to large effect towards the OO3 condition in this variant.

RQ2V4

Prior Distribution and Justification

For the original dataset, I used my initial prior of mu = 0.2 and sigma = 0.5. For the replication dataset, I used priors of mu = 0.228 and sigma = 0.116 based on Bayesian parameter estimation using the initial priors and the original dataset provided for RQ2V2.

Descriptive Statistics

Initial Dataset

Mean	0.004
Standard Deviation	0.219
Observed Standardized Effect Size	0.018

Replication Dataset

Mean	0.016
Standard Deviation	0.193
Observed Standardized Effect Size	0.083

Inferential Analysis

Initial Dataset

	t-value	0.091
	Bayes Factor	0.399
	Posterior Probability of Effect	0.285
Posterior Distribution	**Mean**	-0.028
Posterior Distribution	**Standard Deviation**	0.180

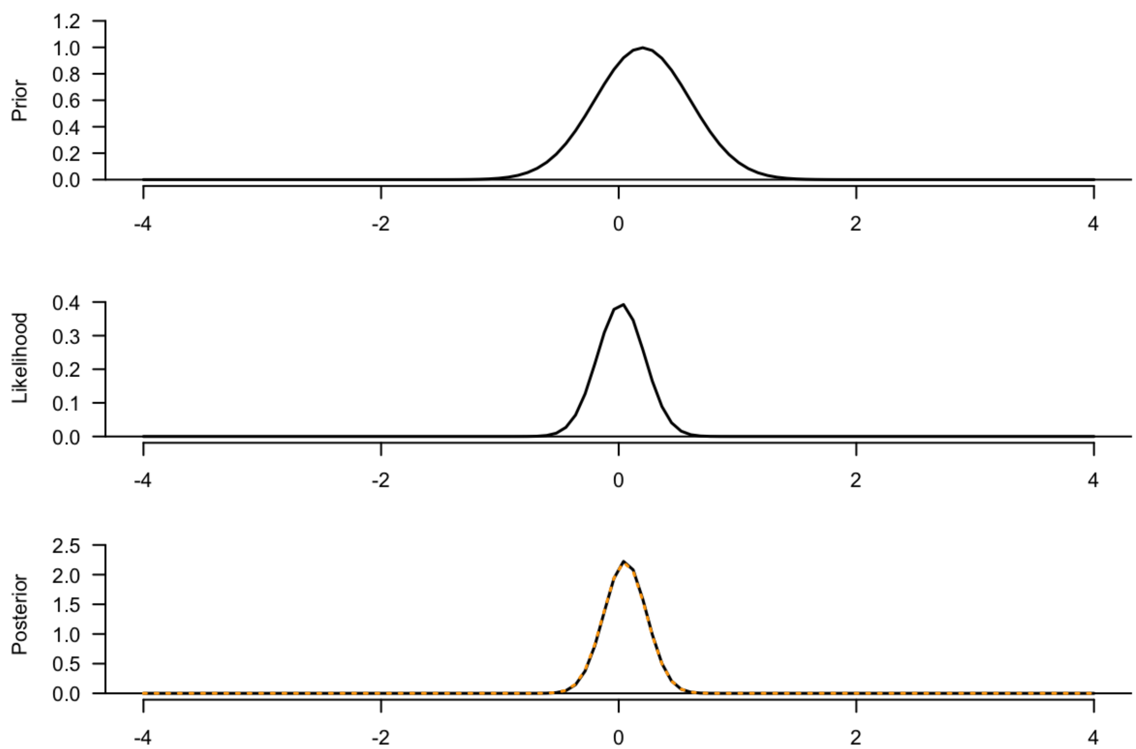

Figure 15: Graph depicting the prior, likelihood, and posterior distributions of the original dataset for RQ2V4. The x value is effect size, and the y value is probability.

Replication Dataset

	t-value	0.415
	Bayes Factor	0.778
	Posterior Probability of Effect	0.438
Posterior Distribution	**Mean**	0.188
Posterior Distribution	**Standard Deviation**	0.100

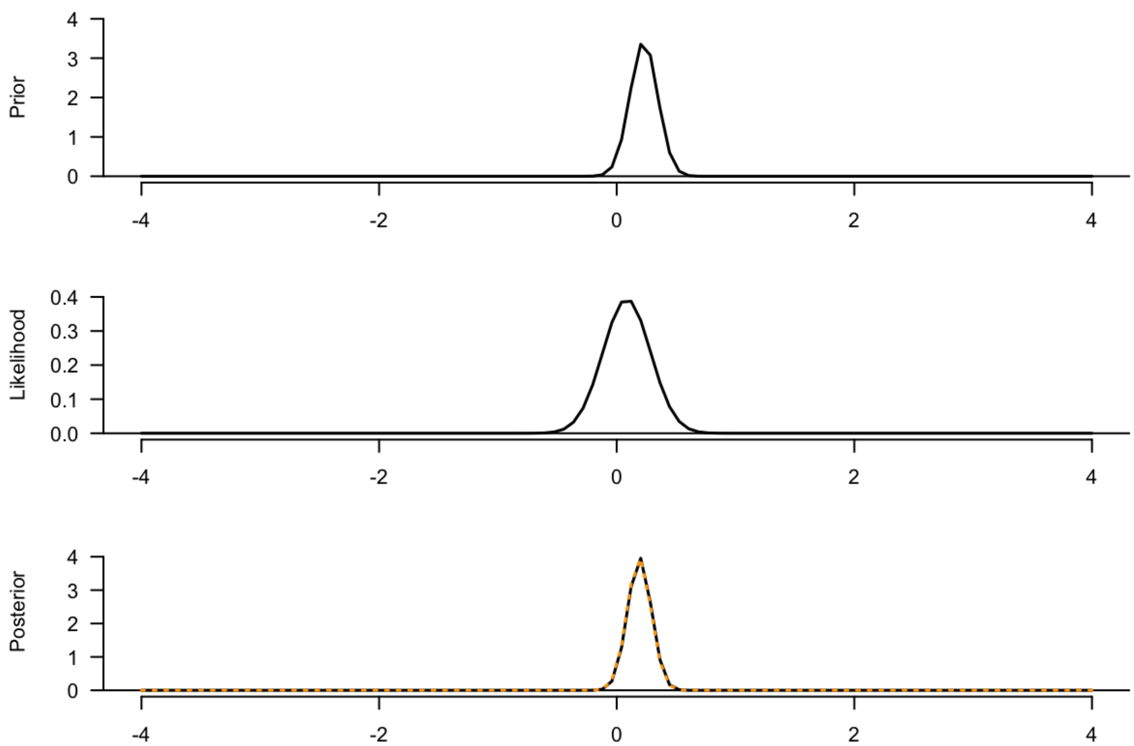

Figure 16: Graph depicting the prior, likelihood, and posterior distributions of the replication dataset for RQ2V4. The x value is effect size, and the y value is probability.

Conclusions

Looking at the original and replication datasets together, I believe that this procedural variant does *not* have an effect. The Bayes factors are both below 1, and the post probability effects are both under 50%. If there were to be an effect, it would be quite small, given the very low observed standardized effect sizes.

Conclusion

RQ1

Overall, I believe that I received two datasets in which there was an effect (RQ1V1 and RQ1V2) and two datasets in which there was not an effect (RQ1V3 and RQ1V4). For those that had an effect, it appeared to be medium to large, with large Bayes Factors and probabilities of effects. Based on this analysis, it appears that the proportion correct for ungrammatical CO3 sentences is indeed different than 0.5.

RQ2

I believe that I received three datasets in which there was an effect (RQ2V1, RQ2V2, and RQ2V3) and one in which there was not an effect (RQ2V4). For those that had an effect, it appeared to be a medium to large sized effect towards the OO3 condition. Based on this analysis, there does appear to be a difference between OC3 and OO3 proportion correct.

References

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. Journal of Experimental Psychology: General, 142(2), 573.

Krushke, J. K. (2015). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan.* (2nd Ed.) Academic Press / Elsevier. ISBN: 978-0-12-405888-0.

Starns, J. (2021, Oct. 18). *Bayesian t-tests in the Bayes Factor Package.* [PowerPoint slides]. Department of Psychological and Brain Sciences, University of Massachusetts Amherst.