2023 ■ Vol. 19 ■ no. 3

# Effects of violating the assumptions of equal variance and independent and identically distributed random variables: A case using Bayesian statistical modeling

Tadamasa Sawada [a,b,c,d]

[a]Department of Psychology, Russian-Armenian (Slavonic) University, Yerevan, Armenia
[b]European University of Armenia, Yerevan, Armenia Akian College of Science and Engineering, American University of Armenia, Yerevan, Armenia
[c]Akian College of Science and Engineering, American University of Armenia, Yerevan, Armenia
[d]School of Psychology, HSE University, Moscow, Russia

**Abstract** ■ All statistical methods involve assumptions about the data and the output of the methods can be biased when the assumptions are not supported by the data. One of the common assumptions is equal variance across the conditions. Another common assumption is that variables are independently sampled from identically distributed populations (i.i.d.). The present study describes an example of such a violation of these assumptions and its effect on the results of Bayesian statistical analyses. Yu et al. (2021) developed a Bayesian statistical model that can analyze the same type of data as the one-way repeated-measure ANOVA. Their model assumed equal variance and i.i.d. Unfortunately, these assumptions were not satisfied by their data. In the present study, their model was revised to allow variance to vary with the conditions, and their data was reanalyzed. The results of the analyses using these models were compared with the psychophysical results of Yu et al. (2021). This comparison showed that the violated assumptions biased the results of the analysis. This bias made the results of the analysis appear more supportive of Yu et al.'s (2021) conclusion, but the validity of the analysis's results needs to be re-considered. Note that it is important that one carefully scrutinizes the data and understands the statistical method used to discuss the results of the analysis.

**Keywords** ■ Bayesian statistics, Bayesian modeling, Type-1 error, Unequal variance, Statistical literacy, Independent and identically distributed random variables. **Tools** ■ R.

tada.masa.sawada@gmail.com

10.20982/tqmp.19.3.p281

**Acting Editor** ■ Denis Cousineau (Université d'Ottawa)

**Reviewers**
■ Two anonymous reviewers

## Introduction

All statistical analyses are formulated on the basis of assumptions about the data, and the results of the analysis become unreliable if these assumptions are violated. It is therefore important to know what the assumptions are and how much the results can be affected by the violation of the assumptions.

For example, one of the most common assumptions used in statistical analysis is the normality of a distribution. Namely, the data, or the average of the data, are regarded as samples taken from a normal distribution (Sawada, 2021). It has been shown that the violation of this assumption could affect the results of the analysis (Blanca et al., 2018; Khan & Rayner, 2003; Oberfeld & Franke, 2013; also see Faulkenberry, 2022; Gottardo & Raftery, 2009; Rossell & Rubio, 2018). Note that in practice, the normality assumption is never strictly satisfied (Micceri, 1989). Consequently, the violation of the assumption and its effect on the results of the analysis should be regarded matters of degree. Many other assumptions are also never strictly satisfied and the effects of their violation are quantitative (Blanca et al., 2018; Fagerland, 2012; Kasuya, 2001; van den Bergh et al., 2023).

Assumptions also play critical roles in Bayesian statistical modeling. Bayesian statistical modeling is a descriptive method to formulate a model to analyze the data. The model can be formulated by considering the properties of the data. For example, a model can use a skewed-normal distribution instead of a normal distribution to characterize the data when the skewed-normal distribution characterizes the data better. So, the assumptions used in the analysis depend on the formulations used in the model. If the data violate some assumptions of the model, the model can be re-formulated to better characterize the data. If the model can include some assumptions that are violated by the data, these violated assumptions can affect the results of the analysis that used the model. So, the data should be carefully scrutinized before they are analyzed and used in a Bayesian statistical model to discuss the validity and assumptions of the model. Note that careful observation means not only checking the results of statistical analyses. It also means actually visualizing the data and then inspecting the visualized data carefully.

The flexibility of Bayesian statistical modeling can provide a "researcher degrees of freedom" (Wicherts et al., 2016). This flexibility should be used to characterize data by introducing assumptions about the data into the model or by revising the assumptions. However, this flexibility may also be misused to distort the results of an analyses in favor of a prediction made by a researcher (see Silberzahn et al., 2018; Simmons et al., 2011). Note that this misusage need not be intentional. When the results of an analysis are consistent with the prediction, the researcher may be biased toward accepting the model without paying sufficient attention to the violation of the model's assumptions and the validity of its formulation. Note that the assumptions used in models are often implicit, and it may be difficult for a researcher to recognize them and their potential violation unless he/she brings considerable care to analyzing the data.

Bayesian models can be formulated in individual studies depending on their data but these models can also have individual problems. So, I believe that discussing individual cases of using Bayesian statistical modeling is important.

I found that Yu et al. (2021, see also Sawada & Pizlo, 2022; Petrov et al., 2022) provide a good example of how a violation of assumptions can distort the results of an analysis. These authors proposed a Bayesian statistical model with some assumptions about their data but the data clearly violated these assumptions. This violation generated a trend in the results of the analysis but this generated trend could be interpreted as an artifact, such as a Type-1 error. I believe that Yu et al. (2021) provides an important example of a problem from which psychologists can learn what they need to be aware of when they perform a Bayesian statistical analysis.

Their model was an analog of a one-way repeated-measures ANOVA that could be used to analyze the data of their psychophysical experiments. Their model was formulated for an experiment that had a single within-subject factor with multiple conditions and multiple measurements were taken from each participant in each condition.

Two of the assumptions used in Yu et al.'s (2021) model were clearly violated by the data in their psychophysical experiment. The first assumption was that the variance of the population of the participants was constant across the conditions. The second assumption was that random variables of a parameter were independently sampled from an identical distribution (i.i.d.). The model characterized the participant's performance in the condition with mean and variance across trials. The mean and variance were regarded as random variables that were sampled from their individual distributions with satisfying the i.i.d. in the model, so, there should not be any systematic relationship between these two variables.

Equal variance across the conditions is one of the common assumptions in statistical analyses, such as Student's unpaired t-test and ANOVAs (Ruscio & Roche, 2012). There are some statistical tests for testing the violation of this assumption (Delacre et al., 2017). The assumption of equal variance is probably another assumption that is never strictly satisfied unless the independent (or quasi-independent) variable of the conditions is completely irrelevant for the dependent variable (Ruscio & Roche, 2012; Grissom, 2000). So, the violation of the assumption and its effect on the results of the analysis should be regarded as matters of degree (Blanca et al., 2018; Fagerland, 2012; Kasuya, 2001) and the violation should be quantitatively evaluated. This evaluation can usually be based on the sample variances or the sample standard deviations of the individual conditions. Note that the sample variance in each condition is attributed to both the between-subject and the within-subject variabilities if the independent variable is within-subject. Then, equal variance should be considered on the basis of the within-subject variability but the violation of the assumption is not always clear from the sample variance. The sample variance should be decomposed into these two types of variabilities (see the section "Reviewing Yu et al. (2021)").

The assumption of i.i.d. is also common in statistical analyses. The Bayesian statistical model of Yu et al. (2021) was formulated under an assumption that the parameters, individually, satisfied the i.i.d. and they are unrelated to one another. Note that the model characterizes the performance of each participant in each condition with the mean and variance across trials in the model. The present study particularly considered the independence of these parameters from one another. There should not be any systematic trend between the means and variances of the participants if the i.i.d. is satisfied.

In the present study, I started by scrutinizing the data obtained in Experiment 1 of Yu et al. (2021) and showed that these data did not satisfy the assumptions of equal variance. Next, their data were analyzed with two Bayesian statistical models. One of the models was used to analyze the data in Yu et al. (2021). This original model assumed that there was equal variance across the conditions. The other model was developed by revising the original model so that the revised model could accept the unequal variance across the conditions. The results of this analysis, which used both the original and revised models, were compared with one another and with the data of Experiment 1 in Yu et al. (2021). This allowed me to discuss how the assumption of equal variance affected the results of the analysis. I also examined whether the data violated i.i.d. and how this violation affected the results of the analysis that used both models. On the basis of these results, I will discuss what one must pay attention to when one uses Bayesian statistical modeling to analyze Psychological data.

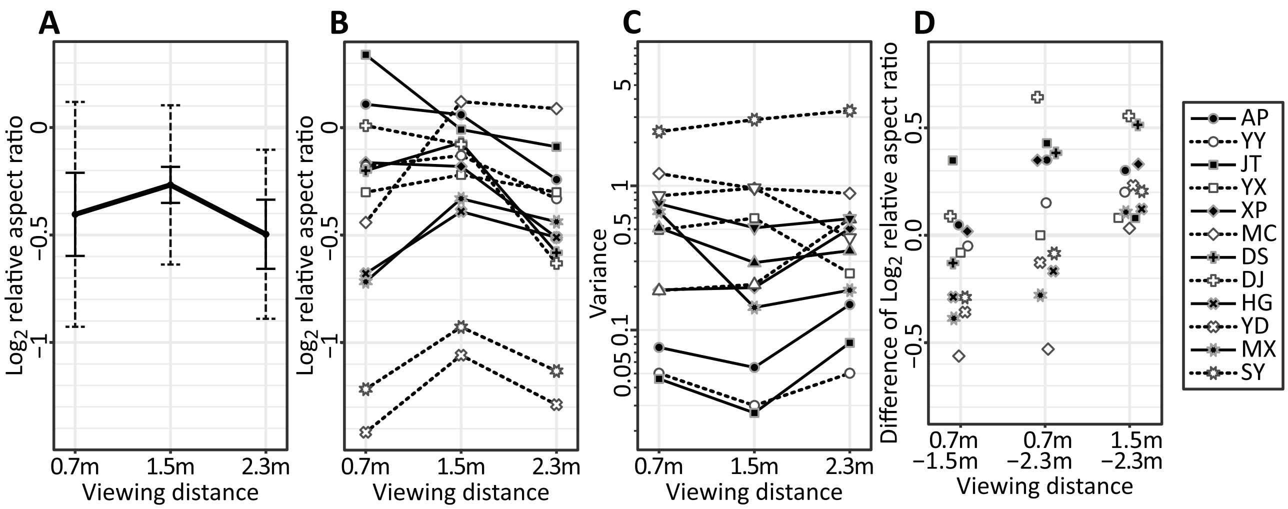
**Reviewing Yu et al. (2021)**

The original model of Yu et al. (2021) had a hierarchical structure with two levels for each participant and for a group of the participants (Appendix, see Yu et al., 2021 for details). The responses of each participant in each condition were represented as samples from a normal distribution in the model. The mean of this distribution was $\mu_{ik}$ and the variance of the distribution was ${\sigma_{ik}}^2$ where $i$ represented the participant and $k$ represented the condition. These two parameters were the most relevant to the sample mean and the sample variance that were computed from the responses of the participant $i$ in condition $k$. Both the mean $\mu_{ik}$ and the variance ${\sigma_{ik}}^2$ of the participant could vary depending on the viewing distance conditions. The mean of the distribution $\mu_{ik}$ was also modeled as a sum of random samples $\beta_i$ and $\theta_{ik}$ from two normal distributions: $\mu_{ik} = \beta_i + \theta_{ik}$. The sample $\beta_i$ was taken from the normal distribution whose mean and variance were $\mu^\beta$ and $(\sigma^\beta)^2$. These two parameters were constant across the participants and the conditions. The other sample $\theta_{ik}$ was taken from the normal distribution whose mean and variance were $\mu^{\theta_k}$ and $(\sigma^\theta)^2$. The mean $\mu^{\theta_k}$ varied depending on the conditions but the variance $(\sigma^\theta)^2$ was constant across the conditions. The sum of $\mu^\beta$ and $\mu^{\theta_k}$ was most relevant to the sample mean of the participants' group in condition $k$. The sum of $(\sigma^\beta)^2$ and $(\sigma^\theta)^2$ was most relevant to the sample variance of the participants' group. The parameters $(\sigma^\beta)^2$ and $(\sigma^\theta)^2$ are independent of the viewing distance conditions. Namely, the model assumes that the variances of the participants' group were equal across the viewing distance conditions. All of these parameters, which are modeled as random variables, satisfy i.i.d. The mean $\mu_{ik}$ and the variance ${\sigma_{ik}}^2$ of the participants are modeled on the basis of these variables, but no variable is used in common to model $\mu_{ik}$ and ${\sigma_{ik}}^2$, so there should not be any systematic trend between $\mu_{ik}$ and ${\sigma_{ik}}^2$.

Consider Experiment 1 of Yu et al. (2021). They tested the effect of an object's distance on the perceived depth-interval of the object's shape based on binocular disparity under two conditions of the object's size (the fixed-physical-size condition and the fixed-projected-size condition, see also Sawada & Pizlo, 2022; Petrov et al., 2022). There were three different viewing distances (0.7 m, 1.5 m, and 2.3 m) in the experiment. In the fixed-physical-size condition, the size of the object in the scene was kept constant while the retinal image size of the object was changed as a function of the viewing distance. In the fixed-projected-size condition, the retinal image size of the object was kept constant by changing the object's size in the scene as a function of the viewing distance. These two conditions about object size (fixed-physical-size and fixed-projected-size) were tested in separate blocks. The three viewing distance conditions (0.7 m, 1.5 m, and 2.3 m) were randomized within each block. Note that Experiment 1 was a two-factor within-subject design, but note that their model was formulated for a single-factor within-subject design. The results of the two object size conditions were analyzed separately in Yu et al. (2021). In my study, the fixed-projected-size condition of Experiment 1 was particularly interesting because the assumption of equal variance of the participants was not well satisfied in this condition (see below).

The results of the fixed-projected-size condition are shown in Figure 1 (replotted from Figures 4 and 5 in Yu et al., 2021). Figure 1A shows averaged results from all 12 participants. Figures 1B and 1C show the sample means and the sample variances of the individual participants. The abscissas represent the viewing distance of the object. The ordinates of Figures 1A and 1B represent the binary logarithm of the scaling factor of the depth-interval (see Yu et al., 2021, for details). A larger value of the log scaling factor in Figures 1A and 1B means that the depth-interval was more extended. The error bars with dotted lines in Figure 1A represent the conventional standard deviation. The error bars with solid lines in Figure 1A represent the within-subject standard deviation, which is the standard deviation

**Figure 1** ■ The results of the fixed-projected-size condition of Experiment 1 in Yu et al. (2021). (A) The averaged results of all 12 participants (was replotted from Yu et al., 2021, Figure 4). The error bars with dotted lines represent the conventional standard deviation. The error bars with solid lines represent the within-subject standard deviation, which is the standard deviation estimated after removing the between-subject variability from the data (Cousineau, 2005). (B) The sample means of the individual participants was replotted from Figure 5 in Yu et al. (2021). (C) The sample variances of the individual participants. (D) The difference of the sample means between the pair of the viewing distances.




estimated after removing the between-subject variability from the data (Cousineau, 2005). The symbols in Figures 1B and 1C represent the individual participants. The average factor is largest at 1.5 m viewing distance and it is similar between 0.7 m and 2.3 m. Yu et al. (2021) analyzed their results by using their Bayesian statistical model and claimed that the log scaling factor became smaller as the viewing distance became larger (see the sub-section "Analysis").

Note that the original model of Yu et al. (2021) assumes that the variance of the population is equal across the conditions of viewing distance, but the empirical results of the experiment do not support this assumption. The standard deviations in Figure 1 (dotted error bars) are 0.52, 0.37, and 0.39 at 0.7 m, 1.5 m, and 2.3 m viewing distance. The within-subject standard deviations in 1 (solid error bars) are 0.19, 0.084, and 0.16. The within-subject standard deviation at 0.7 m viewing distance is more than twice as large as the standard deviation at 1.5 m. The original model assumed the variances of the individual participants could vary depending on the condition. 1 does not show any clear trend across the participants.

Experiment 1 in Yu et al. (2021) used a repeated-measure design and each participant was tested in all of the viewing distance conditions. The dotted error bars in Figure 1A reflect both the between-subject and the within-subject variabilities. The between-subject variability represents the variability of the positions of the curves in Figure 1B. The variability of the curves' positions can be regarded as the main effect of the participants and it is characterized by the parameter $(\sigma^{\beta})^2$ in the Yu et al. (2021) model. The within-subject variability represents the variability of the shapes of the curves in Figure 1B. The variability of the curves' shapes can be regarded as the interaction between the participants and the viewing distance and it is characterized by the parameter $(\sigma^{\theta})^2$ in the Yu et al. (2021) model. The parameter $(\sigma^{\theta})^2$ is constant across the conditions of the viewing distance. This means that Yu et al. (2021) assumed that the within-subject variability was constant across the conditions. Note that they assumed that the variances of the individual participants could vary depending on the condition. The variances of the individual participants in the individual conditions (Figure 1C) are characterized by the parameter $(\sigma_{ik})^2$ where $i$ represented the participant and $k$ represented the condition. The parameter $(\sigma_{ik})^2$ varied depending on the conditions.

The variability of the shapes of the curves can be evaluated on the basis of the difference in the log scaling factor between each pair of viewing distance conditions. The difference in the log scaling factor is plotted in Figure 1D. The ordinate represents the difference in the log scaling factor and the abscissa represents the pair of the viewing distances. The symbols represent the individual participants.

The points are more narrowly distributed for the pair of 1.5 m and 2.3 m ($M = 0.23$, $SD = 0.17$, 95% CI [0.12, 0.34]) than the pair of 0.7 m and 1.5 m ($M = -0.14$, $SD = 0.25$, 95% CI [$-0.30$, 0.022]) and the pair of 0.7 m and 2.3 m ($M = 0.093$, $SD = 0.35$, 95% CI [$-0.13$, 0.31]). The homogeneity of the variance of the differences between the pairs of the viewing distances was statistically rejected when Mauchly's Test for Sphericity was performed ($W = 0.43$, $p = 0.015$, see Cousineau et al., 2021). The 95% confidence interval only suggests that the difference between 1.5 m and 2.3 m viewing distance is non-null. This trend between 1.5 m and 2.3 m viewing distance is consistent with the claim in Yu et al. (2021) that the depth-interval became more compressed as the viewing distance became larger (Cohen's $d = 1.4$).

The differences between 0.7 m and 2.3 m viewing distance (Cohen's $d = 0.27$) and between 0.7 m and 1.5 m viewing distance (Cohen's $d = -0.56$) were not supported by their confidence intervals. The magnitude of the difference between 0.7 m and 2.3 m was smaller than the magnitude of the difference between 0.7 m and 1.5 m. But, the Bayesian statistical analysis in Yu et al. (2021) supported the difference between 0.7 m and 2.3 m. This supported difference was the basis of their claim that the depth-interval became more compressed as the viewing distance became larger (Yu et al., 2021, pp. 7–8). On the other hand, the sign of the difference between 0.7 m and 1.5 m was not consistent with their claim and was not supported by their Bayesian statistical analysis. So, this difference was not discussed in Yu et al. (2021). I replicated their analysis and the difference in the results between the Bayesian statistical analysis and the conventional analysis, i.e., confidence intervals based on a t-distribution. This difference will be discussed in the following section.

## Revising the Model of Yu et al. (2021)

The original model of Yu et al. (2021) was revised in this study to incorporate the variability of the shapes of the curves (Appendix). It was revised by introducing a difference in the variance that depended on the viewing distance conditions for the group of the participants. Note that the variance in the group's level was characterized by $(\sigma^{\beta})^2$ and $(\sigma^{\theta})^2$ in the original model. These parameters were constant across the conditions. The parameter $(\sigma^{\theta})^2$ was a variance of a normal distribution whose mean $\mu^{\theta_k}$ varied, depending on the conditions in the original model. In the revised model, the variance of this normal distribution was $(\sigma^{\theta_k})^2$ where $k$ represented the condition. The parameter $(\sigma^{\theta_k})^2$ varied depending on the conditions. Once this was done, this normal distribution characterized the difference of the mean and the variance across the conditions in the group level of the revised model. Note that the other parameter $(\sigma^{\beta})^2$, which characterizes the group's level variance, was a variance of a normal distribution whose mean $\mu^{\beta}$ was constant across the conditions in both the revised and original models. This distribution characterized the base of the mean and the variance that were constant across the conditions in the group's level.

### *Analysis*

The revised model in my study and in the original model of Yu et al. (2021) were used to analyze the empirical results obtained in Experiment 1 of Yu et al. (2021). This analysis used the Markov Chain Monte Carlo (MCMC) method to estimate the parameters of the models, and the results of this analysis were used to compare the revised model and the original model.

### *Methods*

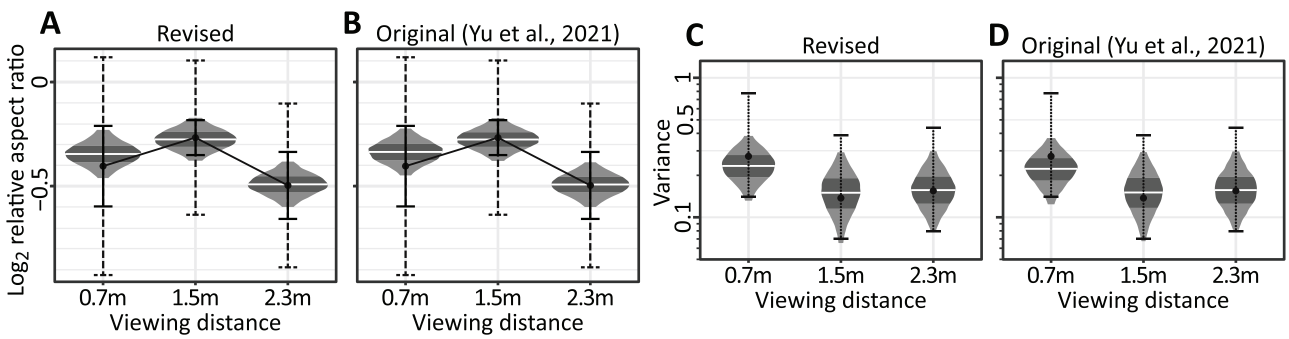
The analysis was conducted by using JAGS (*Just Another Gibbs Sampler*, mcmc-jags.sourceforge.io/, Plummer, 2003) interfaced to R via a library `rjags` and R-studio. Yu et al. (2021) implemented the original model in JAGS. This implementation of the original model was provided by the authors of Yu et al. (2021) and it was used for the analyses in my study. The revised model of this study was also implemented in JAGS based on the implementation of the original model. The JAGS implementation of the revised model and the R scripts for the analysis in my study were uploaded to osf.io/4tu5h/.

The data of Experiment 1 in Yu et al. (2021) were recovered from the graphs that showed the results of these studies (Figure 5 in Yu et al., 2021). The graphs were captured as image files and the data were captured by tracing the plots in the graphs by using PlotDigitizer (plotdigitizer.sourceforge.net). The recovered data were the means and their standard errors of the log scaling factor of the depth-interval for each participant in the three viewing distance conditions. Their standard deviations were computed by multiplying $N_J^{0.5}$ with the recovered standard errors where $N_J$ was the number of trials in each condition, which was 30.

Both the revised model and the original model were formulated to take data from individual trials as their inputs, but these data could not be provided by the authors of Yu et al. (2021), so I had to generate my own appropriate data. I did this by generating synthetic data of individual trials by using a function rnorm in R language that generates pseudo-random numbers following a standard normal distribution. For each condition of each participant, NT random numbers were generated and were transformed so that my synthetic data matched the recovered data from Yu et al.'s (2021) Figure 5 in descriptive statistics:

$$x_{ijk} = \left(\hat{x}_{ijk} - \hat{\mu}_{ik}\right)\frac{\bar{\sigma}_{ik}}{\hat{\sigma}_{ik}} + \bar{\mu}_{ik} \tag{1}$$

**Figure 2** ■ The Bayesian posterior predictive distributions (95% HDIs) of (A, B) the means and (C, D) the variance of the log scaling factor across the participants from the revised model and the original model. The empirical results of the fixed-projected-size condition in Experiment 1 of Yu et al. (2021) were superimposed on these graphs. The error bars in (A, B) are plotted in the same way as Figure 1A. The error bars in (C, D) represent the 95% confidence interval of the variance on the basis of a $\chi^2$ distribution.




where $\hat{x}_{ijk}$ was a random number from the generator and $x_{ijk}$ was the synthetic data of the $j$-th trial of participant $i$ in condition $k$ transformed from $\hat{x}_{ijk}$. The parameters $\hat{\mu}_{ik}$ and $\hat{\sigma}_{ik}$ are the sample mean and the sample standard deviation of the random variables before the transformation. The sample mean and the sample standard deviation of the synthetic data became identical to the sample mean $\bar{\mu}_{ik}$ and the sample standard deviation $\bar{\sigma}_{ik}$ of the recovered data after the transformation.

The procedure for the analyses of the revised model of my study and of the original model of Yu et al. (2021) followed the procedure used in the analysis in Yu et al. (2021). These two models were applied to the synthetic data using the Markov chain Monte Carlo (MCMC) method. Results of each model were obtained from two chains of the MCMC sampling. Each chain started with 1000 samples of a burn-in period that was followed by 9000 samples. Convergence of the chains was visually confirmed on the basis of the trace and Gelman-Rubin-Brooks plots of all the parameters in the group's level of the models.

Analysis sessions were repeated 100 times to check the robustness of the results of the analysis. Note that the analyses were based on the synthetic data that were randomly generated in each session. This randomness of the synthetic data affected the results of the analyses but the effect was small. The results of the analyses were discussed on the basis of their trends that were robust across the 100 sessions unless specified.

*Results*

The revised model and the original model were first compared on the basis of their Bayesian posterior predictive distributions of the binary logarithm of the depth scaling factor in the fixed-projected-size condition.

Figures 2A and 2B show the Bayesian posterior predictive distributions of the means of the log scaling factor across the participants. Figure 2A shows the results of the revised model and Figure 2B shows the results of the original model. The abscissas represent the viewing distance of the object. The ordinates represent the log scaling factor. The shaded areas are violin plots that represent the shapes of the posterior probability distributions. The violin plots are trimmed with the 95% highest density interval (HDI). The darker regions of the violin plots represent the 25th and the 75th percentiles of the distributions. The white line segments in the darker regions represent the medians of the distributions. The sample means and the sample standard deviations of the fixed-projected-size condition in Experiment 1 of Yu et al. (2021) were superimposed on these graphs (see Figure 1A for the same empirical results). Both of the models fit the empirical results almost equally well but they tend to overestimate the mean of the log scaling factor somewhat at the 0.7 m viewing distance relative to the empirical results.

Figures 2C and 2D show the Bayesian posterior predictive distributions of the variance of the log scaling factor across the participants. Both of the models fit the empirical results almost equally well but tend to underestimate the variance of the log scaling factor somewhat at the 0.7 m viewing distance compared with the empirical results.

Now, consider the Bayesian posterior estimates of parameters in the models. The posterior estimates of the parameters are summarized in Table 1.

Figure 3 shows the Bayesian posterior distributions of $d_k$ of the revised model (Figure 3A) and the original model (Figure 3B) in the same format as Figure 2. These graphs

**Table 1** ■ The parameters of the revised and original models estimated from the 100 analysis sessions. The columns represent the 25th, the 50th (median), and the 75th percentiles of the HDIs of the Bayesian posterior distributions of the parameters for the revised and original models.

| | Revised | | | Original | | |
|---|---|---|---|---|---|---|
| | 25% | 50% | 75% | 25% | 50% | 75% |
| $\mu^{\theta_1}$ | −0.0067 | 0.026 | 0.058 | 0.00090 | 0.028 | 0.055 |
| $\mu^{\theta_2}$ | 0.058 | 0.086 | 0.11 | 0.061 | 0.088 | 0.12 |
| $\mu^{\theta_3}$ | −0.14 | −0.11 | −0.082 | −0.14 | −0.12 | −0.090 |
| $(\sigma^{\theta_1})^2$ | 0.018 | 0.028 | 0.042 | 0.013 | 0.018 | 0.024 |
| $(\sigma^{\theta_2})^2$ | 0.012 | 0.018 | 0.027 | 0.013 | 0.018 | 0.024 |
| $(\sigma^{\theta_3})^2$ | 0.014 | 0.021 | 0.031 | 0.013 | 0.018 | 0.024 |
| $\mu^{\beta}$ | −0.45 | −0.37 | −0.29 | −0.45 | −0.37 | −0.28 |
| $(\sigma^{\beta})^2$ | 0.12 | 0.16 | 0.23 | 0.12 | 0.17 | 0.24 |
| $\mu^{\lambda_1}$ | −0.65 | −0.52 | −0.39 | −0.65 | −0.52 | −0.39 |
| $\mu^{\lambda_2}$ | −0.80 | −0.67 | −0.54 | −0.80 | −0.67 | −0.54 |
| $\mu^{\lambda_3}$ | −0.65 | −0.52 | −0.39 | −0.64 | −0.52 | −0.39 |
| $(\sigma^{\lambda})^2$ | 0.35 | 0.42 | 0.51 | 0.35 | 0.42 | 0.50 |

of the revised model and of the original model are roughly similar to one another. The effect size is largest at 1.5 m viewing distance and smallest at 2.3 m. The HDIs do not overlap between 1.5 m and 2.3 m of viewing distance in both models. The HDIs do overlap one another between 0.7 m and 1.5 m and between 0.7 m and 2.3 m in both models. The overlaps between 0.7 m and 1.5 m are substantial for both models. The overlap between 0.7 m and 2.3 m is larger for the revised model than for the original model. This overlap is small for the original model but it is stably observed in all the 100 sessions of the simulation. The Bayesian posterior distributions of $\mu^{\theta_k}$ show the same trends as the posterior distributions of $d_k$ for both models (Figures 3C, 3D).

*Independent and identically distributed random variables*

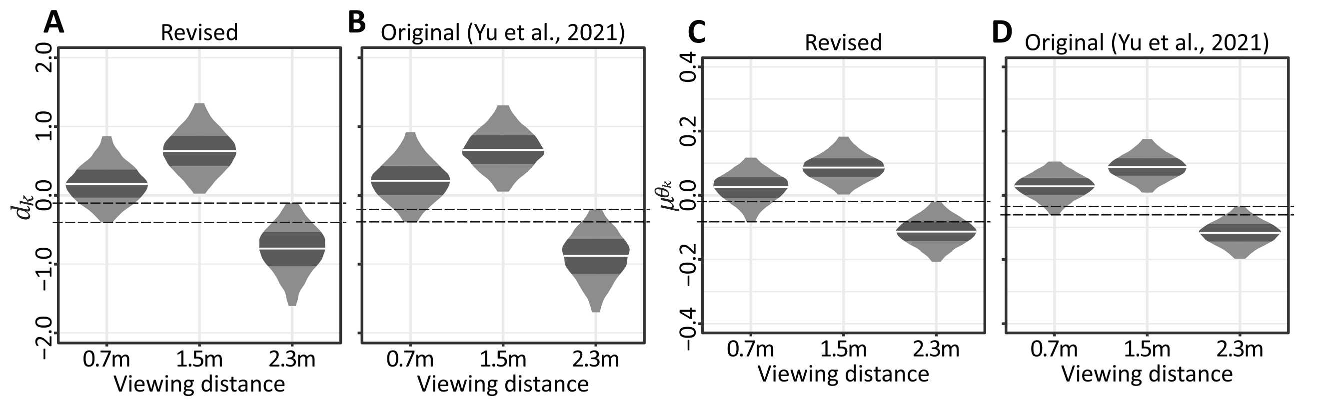
The results of the analyses showed that the revised model used in my study and the original model of Yu et al. (2021) fitted the empirical results almost equally well but they both showed similar discrepancies from the empirical results at the 0.7 m viewing distance (Figures 2, 3). These differences between the models and the empirical results can be attributed to the violation of i.i.d. in the empirical results at the 0.7 m viewing distance.

Consider the individual participants' empirical results at the 0.7 m viewing distance (Figure 4A). It is easy to see that the sample means of the log scaling factor were larger as the sample variances became smaller at the 0.7 m viewing distance (Spearman's rank correlation: $\rho(10) = -0.89$, $p = 9.2 \times 10^{-5}$). A similar trend cannot be seen clearly at the 1.5 m ($\rho(10) = -0.48$, $p = 0.12$) or at the 2.3 m ($\rho(10) = -0.44$, $p = 0.15$) viewing distances (Figures 4B, 4C). The trend at the 0.7 m viewing distance suggests that i.i.d. was violated. Note that if the assumption had been well satisfied, the dots representing the individual participants in Figure 4A would be distributed mirror-symmetrically around a horizontal line that represents the population mean and more widely as the sample variance became larger.

A violation of i.i.d. at the 0.7 m viewing distance could cause a discrepancy between the results of the analysis from the psychophysical results (see Figures 2, 3). Specifically, the mean of the log scaling factor was overestimated (Figures 2A, 2B) and the variance was underestimated (Figures 2C, 2D) at the 0.7 m viewing distance. It is possible to interpret the results of the analysis as being biased toward the results of the participants whose sample variances were small.

The bias at the 0.7 m viewing distance can be attributed to the overestimation of $d_1$ and $\mu^{\theta_1}$ but these parameters cannot be compared directly with the psychophysical results. Instead, the differences of $\mu^{\theta_1}$ (0.7 m) from $\mu^{\theta_2}$ (1.5 m) and from $\mu^{\theta_3}$ (2.3 m) can be compared with the differences of the sample means of the individual participants between 0.7 m and 1.5 m and between 0.7 m and 2.3 m (Figure 5). It can be seen in Figure 5 that both $\mu^{\theta_1} - \mu^{\theta_2}$ and $\mu^{\theta_1} - \mu^{\theta_3}$ are overestimated relative to the mean of the differences of the sample means. The HDIs of $\mu^{\theta_1} - \mu^{\theta_2}$ include zero for both models. The HDI of $\mu^{\theta_1} - \mu^{\theta_3}$ stably includes zero in the tail of the distribution for the revised model (see Kruschke, 2018). The HDI of $\mu^{\theta_1} - \mu^{\theta_3}$ barely includes zero for the original model in 65 out of the 100 sessions of the simulation. Note that the differences of the sample means between 0.7 m and 1.5 m (Spearman's rank

**Figure 3** ■ The Bayesian posterior distributions (95% HDIs) of (A, B) $d_k$ and (C, D) $\mu^{\theta_k}$ for the revised and original models. The horizontal dashed lines represent the bottom of the HDI at 0.7 m viewing distance and the top of the HDI at 2.3 m viewing distance.




correlation: $\rho(10) = -0.88$, $p = 0.00019$) and between 0.7 m and 2.3 m (Spearman's rank correlation: $\rho(10) = -0.75$, $p = 0.0075$) were also correlated with the sample variances at the 0.7 m viewing distance.

*Summary*

My analysis in the present study, showed that the result in Yu et al. (2021) could be biased because their data did not satisfy two of the assumptions used in the original Bayesian statistical model of Yu et al. (2021). The first violated assumption was that the variances of the participants' group were equal across the viewing distance conditions. The second violated assumption was that the means and variances of the individual participants individually satisfied i.i.d. and that there was no interdependence between the means and the variances. I scrutinized the data obtained in Experiment 1 of Yu et al. (2021) and showed that these data did not satisfy these assumptions (Figures 1, 4).

Their original model was revised in the present study so that the revised model could allow the variance to vary depending on the condition. These two models tended to show similar trends. There was a difference, however, between these two models. It was observed in the comparison of $d_k$ and $\mu^{\theta_k}$ between 0.7 m and 2.3 m viewing distance, but note that this difference was rather small (Figure 5).

Both of the models fitted the empirical results almost equally well but they both showed similar discrepancies from the psychophysical results at the 0.7 m viewing distance (Figures 2, 3). These discrepancies could be attributed to the violation of i.i.d. The violation was in the correlation between the sample means and the sample variances of the log scaling factor across the participants at the 0.7 m viewing distance (Figure 4A). This violation overestimated the mean of the log scaling factor (Figure 2) at the 0.7 m viewing distance.

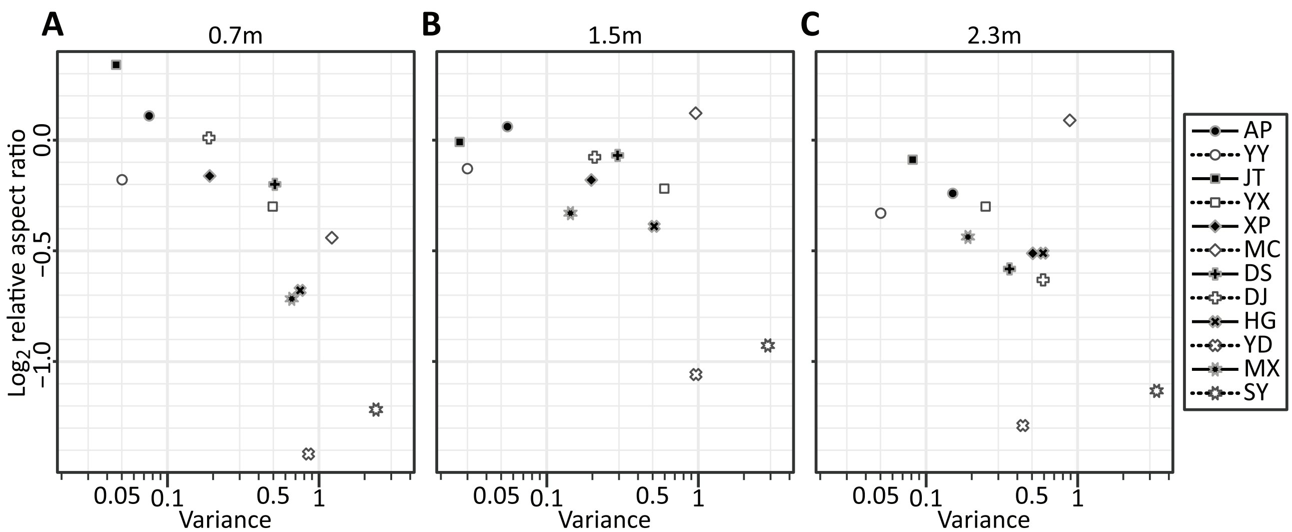
Note that the trend in the psychophysical results between 0.7 m and 2.3 m was consistent with the claim of Yu et al. (2021) that the depth-interval became more compressed as the viewing distance and the trend between 0.7 m and 1.5 m was not consistent with the claim, but note that these trends were rather weak (Figure 1). The effect of violating i.i.d. led to an overestimation of the consistent trend between 0.7 m and 2.3 m and to an underestimation of the inconsistent trend between 0.7 m and 1.5 m in the results of the Bayesian analysis. The consistent trend between 0.7 m and 2.3 m was also slightly magnified by the violation of equal variance.

The violation of i.i.d. that is caused by the correlation between the sample means and the sample variances of the individual participants is less critical with the common conventional statistical methods used in Psychophysical studies, such as t-tests and ANOVAs. These conventional methods do not make use of the sample variances of the individual participants when they are used to analyze the results of a population of participants.

***Other issues in Yu et al. (2021)***

There are a few additional issues about the analyses in Yu et al. (2021). First, Yu et al. (2021) could have mis-perceived the results of their analysis. They observed that "the HDIs do not overlap at all for the near and far distances" (p. 7 in Yu et al., 2021). The "near" and "far" distances referred to the 0.7 m and 2.3 m viewing distances (Y. Yu & A. A. Petrov, personal communication, August 2, 2023). This observation was an important part of their "clear evidence" (p. 12

**Figure 4** ■ The sample means of the log scaling factor of the individual participants (ordinates) plotted as a function of their sample variances (abscissas) at the (A) 0.7 m, (B) 1.5 m, and (C) 2.3 m viewing distances.




in Yu et al., 2021) used to support their claim, but this observation is not exactly consistent with their graph in Figure 6 of Yu et al. (2021). The HDIs at 0.7 m and 2.3 m viewing distance slightly overlapped one another in their Figure 6 (Kruschke, 2018). This inconsistency was small but their description was quite binary. Currently, authors working in Psychology are encouraged to describe the results of their analyses quantitatively.

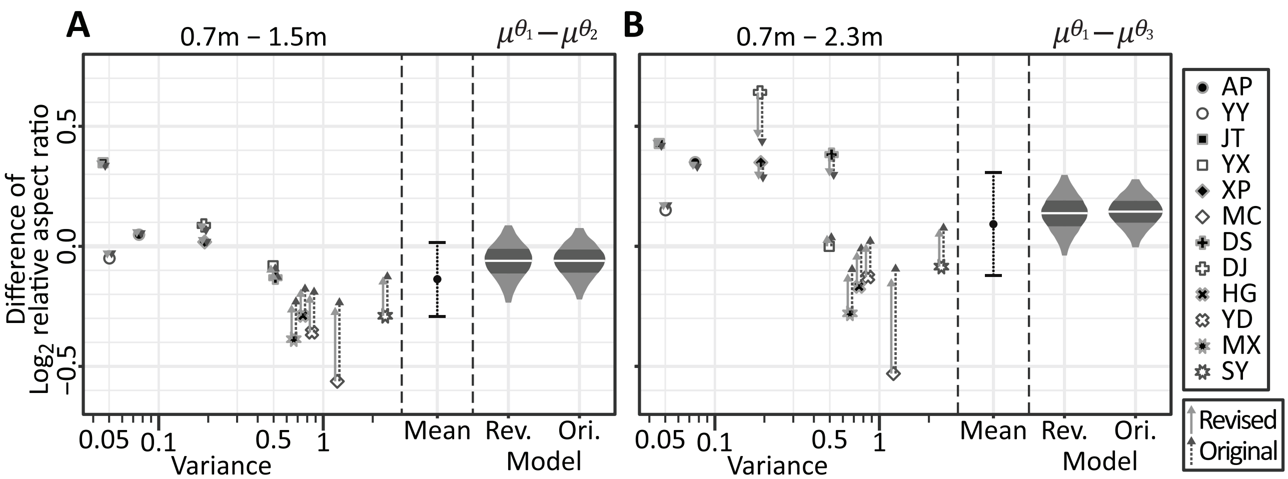
Next, the authors of Yu et al. (2021) might have been misled by how they plotted their results (Figure 5 in Yu et al., 2021). Yu et al. (2021) plotted the results of individual participants in separate panels that had different ranges for their ordinates (Figure 5 in Yu et al. 2021). A small range of the ordinate would enhance visually any trend in the plot while a large range would suppress any trend (see Huff, 1954; Yang et al., 2021). The participants whose results were consistent with the claim of Yu et al. (2021) that the depth-interval became more compressed as the viewing distance became larger were plotted in Yu et al. (2021, Figure 5) with the short (1.2) or middle (1.7) ranges of the ordinates, so these consistent results tended to be visually enhanced in Yu et al.'s (2021) Figure 5. Their figure could have misled the authors of Yu et al. (2021) and it can mislead their readers about the overall trend of their results.

## General Discussion

The present study discussed a case (Yu et al., 2021) in which a Bayesian statistical analysis supported some trends in psychophysical data while these trends were not supported by a conventional analysis (t-tests). This difference in their results could be explained by a violation of the assumptions, which were used in the Bayesian statistical model, used in the analysis. The assumptions violated were: (i) equal variance across the conditions and (ii) independent and identically distributed random variables (i.i.d.). These are two of the most common assumptions that are used in many statistical analyses. I scrutinized the data and showed that these data did not satisfy these assumptions. I also conducted Bayesian statistical analyses of their data using two different models. One of the models was formulated by Yu et al. (2021) with the assumption of equal variance. The other model in this study, was formulated without the assumption of equal variance. There was a difference between these two models but the difference was rather small. The models showed similar discrepancies from the psychophysical data and these discrepancies could be explained by the violation of i.i.d. The violation of these assumptions biased the results of the analysis in such a way that the results appeared more supportive of Yu et al.'s (2021) conclusion.

Bayesian statistical modeling is a descriptive method that can be used to formulate a model to analyze data with a high degree of flexibility. Flexibility is an advantage of Bayesian statistical modeling but the flexibility can provide a "researcher degrees of freedom" (Wicherts et al., 2016; see also Goodboy & Kline, 2017; Seaman & Weber, 2015). It demands a high degree of literacy on the part of the researcher who must carefully observe the data and formu-

**Figure 5** ■ The differences of the sample means of the individual participants (A) between 0.7 m and 1.5 m and (B) between 0.7 m and 2.3 m (ordinates) as functions of the sample variances of the individual participants at 0.7 m viewing distance (abscissas). The arrows represent (A) $\mu_{i1} - \mu_{i2}$ and (B) $\mu_{i1} - \mu_{i3}$ of the individual participants. The styles of arrows represent the revised and original models. On the right sides of the individual graphs, the mean of the difference of the sample means across the participants with their 95% confidence intervals based on a t-distribution and the 95% HDIs of Bayesian posterior distributions of (A) $\mu^{\theta_1} - \mu^{\theta_2}$ and (B) $\mu^{\theta_1} - \mu^{\theta_3}$ for the revised and original models are plotted for comparison.




late the model while examining and implementing properties of the data and it also makes similar demands on the part of the readers. The researcher should be able to recognize which assumptions were used in the model and he/she should consider the validity of these assumptions about the data. Note that the number of assumptions used in the model increases as the model uses more parameters and as the model becomes more complex. These features of the model and the properties of the data should be clearly reported. This is especially critical if any inference is going to be made on the basis of the results of the analysis that are not clearly visible in the data. Careful scrutinization of the data is important in all statistical analyses, including Bayesian analysis.

All statistical analyses, including Bayesian statistical analysis, are formulated on the basis of assumptions about the data, and the results of the analysis become unreliable if these assumptions are violated. The present study describes an example of such a violation of assumptions of equal variance and i. i. d. and its effect on the results of a Bayesian statistical analysis. Bayesian models can be formulated in individual studies depending on their data but these models can also have individual problems. Therefore, regardless of the statistical method used, it is important to carefully scrutinize the data and understand the statistical method.

## Author's note

The author thanks the authors of Yu et al. (2021), Drs. Ying Yu, James T. Todd, and Alexander A. Petrov, for sharing their code.

The data and the code used in this study are uploaded to osf.io/4tu5h/ except for the code of the original model of Yu et al. (2021). Requests for the code of the original model should be directed to the authors of Yu et al. (2021).

## References

Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2018). Effect of variance ratio on anova robustness: Might 1.5 be the limit? *Behavior Research Methods, 50*(3), 937–962. doi: 10.3758/s13428-017-0918-2.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)* Lawrence Erlbaum Associates.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to loftus and masson's method. *Tutorials in quantitative methods for psychology, 1*(1), 42–45. doi: 10.20982/tqmp.01.1.p042.

Cousineau, D., Goulet, M.-A., & Harding, B. (2021). Summary plots with adjusted error bars: The superb framework with an implementation in r. *Advances in Methods and Practices in Psychological Science, 4*(3), 1–18. doi: 10.1177/25152459211035109.

Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use welch's t-test instead of student's t-test. *International Review of Social Psychology*, *30*(1), 92–101. doi: 10.5334/irsp.82.

Fagerland, M. W. (2012). T-tests, non-parametric tests, and large studies—a paradox of statistical practice? *BMC Medical Research Methodology*, *12*(1), 78–99. doi: 10.1186/1471-2288-12-78.

Faulkenberry, T. J. (2022). *A note on the normality assumption for bayesian models of constraint in behavioral individual differences.* arXiv: 2112.05503). doi: 10.48550/arXiv.2112.05503.

Goodboy, A. K., & Kline, R. B. (2017). Statistical and practical concerns with published communication research featuring structural equation modeling. *Communication Research Reports*, *34*(1), 68–77. doi: 10.1080/08824096.2016.1214121.

Gottardo, R., & Raftery, A. (2009). Bayesian robust transformation and variable selection: A unified approach. *Canadian Journal of Statistics*, *37*(3), 361–380. doi: 10.1002/cjs.10021.

Grissom, R. J. (2000). Heterogeneity of variance in clinical data. *Journal of Consulting and Clinical Psychology*, *68*(1), 155–165. doi: 10.1037/0022-006X.68.1.155.

Huff, D. (1954). *How to lie with statistics.* W. W. Norton.

Jordan, M. I. (2004). Graphical models. *Statistical Science*, *19*(1), 140–155. doi: 10.1214/088342304000000026.

Kasuya, E. (2001). Mann-whitney u test when variances are unequal. *Animal Behaviour*, *61*(6), 1247–1249. doi: 10.1006/anbe.2001.1691.

Khan, A., & Rayner, G. D. (2003). Robustness to nonnormality of common tests for the many-sample location problem. *Journal of Applied Mathematics and Decision Sciences*, *7*(4), 187–206. doi: 10.1207/S15327612JAMD0704_1.

Kruschke, J. K. (2018). Rejecting or accepting parameter values in Bayesian estimation. *Advances in Methods and Practices in Psychological Science*, *1*(2), 270–280. doi: 10.1177/2515245918771304.

Lee, M. D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psychonomic Bulletin & Review*, *15*(1), 1–15. doi: 10.3758/PBR.15.1.1.

Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian cognitive modeling: A practical course.* Cambridge University Press.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105*, 156–166. doi: 10.1037/0033-2909.105.1.156.

Oberfeld, D., & Franke, T. (2013). Evaluating the robustness of repeated measures analyses: The case of small sample sizes and nonnormal data. *Behavior Research Methods*, *45*(3), 792–812. doi: 10.3758/s13428-012-0281-2.

Petrov, A. A., Yu, Y., & Todd, J. T. (2022). Are mirror-symmetric objects of special importance for 3d shape perception? a reply to sawada and pizlo (2022). *Journal of Vision*, *22*(4), 1–5. doi: 10.1167/jov.22.4.16.

Plummer, M. (2003). Jags: A program for analysis of Bayesian graphical models using gibbs sampling. In E. Master (Ed.), *Proceedings of the 3rd international workshop on distributed statistical computing* (pp. 1–9). Distributed Statistical Computing.

Rossell, D., & Rubio, F. J. (2018). Tractable Bayesian variable selection: Beyond normality. *Journal of the American Statistical Association*, *113*(524), 1742–1758. doi: 10.1080/01621459.2017.1371025.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. doi: 10.3758/PBR.16.2.225.

Ruscio, J., & Roche, B. (2012). Variance heterogeneity in published psychological research: A review and a new index. *Methodology*, *8*(1), 1–11. doi: 10.1027/1614-2241/a000034.

Sawada, T. (2021). Conditions of the central-limit theorem are rarely satisfied in empirical psychological studies. *Frontiers in Psychology*, *12*(76241), 1–4. doi: 10.3389/fpsyg.2021.762418.

Sawada, T., & Pizlo, Z. (2022). Testing a formal theory of perception is not easy: Comments on yu, todd & petrov (2021) and yu, petrov & todd (2021). *Journal of Vision*, *22*(15), 1–4. doi: 10.1167/jov.22.4.15.

Seaman, C. S., & Weber, R. (2015). Undisclosed flexibility in computing and reporting structural equation models in communication science. *Communication Methods and Measures*, *9*(4), 208–232. doi: 10.1080/19312458.2015.1096329.

Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E.-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical bayesian methods. *Cognitive Science*, *32*(8), 1248–1284. doi: 10.1080/03640210802414826.

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., Bahník, Š., Bai, F., Bannard, C., Bonnier, E., Carlsson, R., Cheung, F., Christensen, G., Clay, R., Craig, M. A., Dalla Rosa, A., Dam, L., Evans, M. H., Flores Cervantes, I., & Nosek, B. A. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*, *1*(3), 337–356. doi: 10.1177/2515245917747646.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. doi: 10.1177/0956797611417632.

van den Bergh, D., Wagenmakers, E.-J., & Aust, F. (2023). Bayesian repeated-measures analysis of variance: An updated methodology implemented in jasp. *Advances in Methods and Practices in Psychological Science*, *6*. doi: 10.1177/25152459231168024.

Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., van Aert, R. C. M., & van Assen, M. A. L. M. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology*, *7*, 1–9. https://www.frontiersin.org/articles/10.3389/fpsyg.2016.01832

Yang, B. W., Restrepo, V., C., S., L., M., & Marsh, E. J. (2021). Truncating bar graphs persistently misleads viewers. *Journal of Applied Research in Memory and Cognition*, *10*(2), 298–311. doi: 10.1016/j.jarmac.2020.10.002.

Yu, Y., Todd, J. T., & Petrov, A. A. (2021). Failures of stereoscopic shape constancy over changes of viewing distance and size for bilaterally symmetric polyhedra. *Journal of Vision*, *21*(5), 1–19. doi: 10.1167/jov.21.6.5.

**Appendix**

The Bayesian statistical model proposed in my study was formulated by revising the model of Yu et al. (2021). The original model of Yu et al. (2021) was a Bayesian analog of a one-way repeated-measures ANOVA. The original model was formulated for an experiment that had a single within-subject factor with multiple conditions. Multiple measurements were taken from each participant in each condition. The design of the experiment was fully balanced, namely, each participant was tested in all the conditions and ran the same number of trials in each condition. Note that the original model assumed that the variance of the participants' population was constant across the conditions. The original model was revised by introducing a difference in the variance of the participants' population that depended on the viewing distance conditions.

This appendix describes how I formulated the revised model. Note that the description of the revised model provided just below is mostly true for the original model of Yu et al. (2021) as well. The difference between the revised model and the original model is summarized in this appendix in the sub-section called "Difference from Yu et al.'s (2021) model".

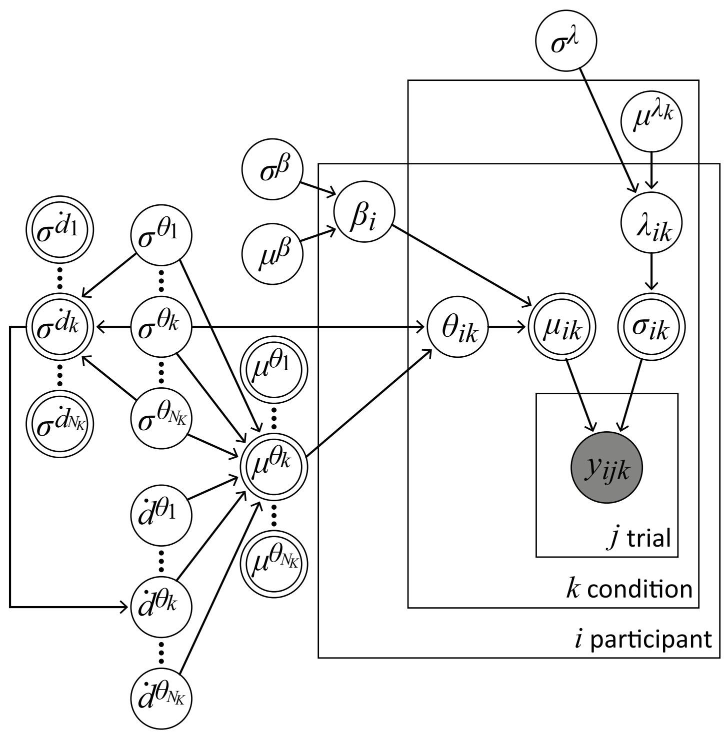
The input to the model is data taken from a psychophysical experiment. The model is also given the number of participants $N_I$, the number of conditions $N_K$, and the number of trials in each condition $N_J$. Note that the data used in the analyses of this study were taken from Yu et al. (2021). There were 12 participants and 3 conditions of viewing distance and each participant ran 30 trials in each condition: $N_I = 12$, $N_K = 3$, $N_J = 30$. The data put into the model was the set of 1080 responses of the participants.

The revised Bayesian statistical model used in my study is represented graphically in Figure A1. Note that this representation is the same as the representation of the original model provided in Yu et al. (2021, Figure A1) except for a few revised parts (see the sub-section "Difference from Yu et al.'s (2021) model"). The input to the model is represented by $y_{ijk}$ where the subscripts $i$, $j$, and $k$ represent participant $i$, condition $k$, and trial $j$. The participants, conditions, and trials are represented by the nested plates in the figure. The response $y_{ijk}$ is regarded as a random variable that is independently sampled from a normal distribution with the mean $\mu_{ik}$ and the variance ${\sigma_{ik}}^2$. Both $\mu_{ik}$ and ${\sigma_{ik}}^2$ can vary depending on the participants and the conditions and they are constant across the trials. The distribution of these two parameters $\mu_{ik}$ and ${\sigma_{ik}}^2$ characterize the performance of each participant in each condition.

The natural logarithm $\lambda_{ik}$ of the standard deviation $\sigma_{ik}$ is regarded as an independent random sample from a normal distribution whose mean and variance are $\mu^{\lambda_k}$ and $(\sigma^{\lambda})^2$. The mean $\mu^{\lambda_k}$ can vary depending on the conditions and is constant across the participants. The variance $(\sigma^{\lambda})^2$ is constant across the conditions and the participants. Priors are individually assigned to the parameters $\mu^{\lambda_k}$ and $\sigma^{\lambda}$ (see the sub-section "Priors").

The mean $\mu_{ik}$ of the distribution for participant $i$ in condition $k$ is represented as a sum of two parameters $\beta_i$ and $\theta_{ik}$: $\mu_{ik} = \beta_i + \theta_{ik}$. The parameter $\beta_i$ represents the mean performance of participant $i$ across the conditions. The parameter $\theta_{ik}$ represents the performance of participant $i$ in condition $k$ relative to the mean $\beta_i$ of the participant. The parameter $\beta_i$ is regarded as a random sample from a normal distribution whose mean and variance are $\mu^{\beta}$ and $(\sigma^{\beta})^2$. These two parameters are constant across the participants and the conditions. The parameter $\theta_{ik}$ is a random sample in a normal distribution whose mean and variance are $\mu^{\theta_k}$ and $(\sigma^{\theta_k})^2$. These two parameters can vary depending on the conditions. Both $\mu^{\theta_k}$ and $(\sigma^{\theta_k})^2$ are constant across the participants. The parameter $\mu^{\beta}$ represents the mean performance of the population of the participants across the conditions. The parameter $\mu^{\theta_k}$ represents the performance of the participants' population in condition k relative to the mean $\mu^{\beta}$ of the population. Note that the mean $\mu_{ik}$ of the distribution for participant $i$ in condition k can also be regarded as a random sample from a normal distribution whose mean and variance are $\mu^{\beta} + \mu^{\theta_k}$ and $(\sigma^{\beta})^2 + (\sigma^{\theta_k})^2$. This distribution characterizes the population of the participants in condition $k$. Priors are individually assigned to the parameters $\mu^{\beta}$, $\sigma^{\beta}$, and $\sigma^{\theta_k}$ (see the sub-section "Priors").

**Figure A1** ■ A graphical model representing structure of the Bayesian statistical model proposed in my study (Jordan, 2004; Lee, 2008; Shiffrin et al., 2008). The parameter $y_{ijk}$ with a shaded circle represents the input to the model. The parameters with single-bordered circles represent latent variables that were sampled from their individual priors. The parameters with double-bordered circles represent latent variables that were computed deterministically from the other parameters.




Group-level parameters:

$\sigma^{\theta_k} \sim Uniform(0,10)$

$(\sigma^{\dot{d}_k})^2 \leftarrow (\sigma^d)^2 \dfrac{({N_K}^2 - N_K)(\sigma^{\theta_k})^2 - \sum_{q=1}^{N_K}(\sigma^{\theta_q})^2}{(N_K-1)(N_K-2)(\sigma^{\theta_k})^2}$

$\dot{d}^{\theta_k} \sim Gaussian(0, \sigma^{\dot{d}_k})$

$\mu^{\theta_k} \leftarrow d^{\theta_k}\sigma^{\theta_k} = \dot{d}^{\theta_k}\sigma^{\theta_k} - \dfrac{\sum_{q=1}^{N_K}(\dot{d}^{\theta_q}\sigma^{\theta_q})}{N_K}$

$\mu^{\beta} \sim Gaussian(0, 10)$

$\sigma^{\beta} \sim Uniform(0,10)$

$\mu^{\lambda_k} \sim Gaussian(0,2)$

$\sigma^{\lambda} \sim Uniform(0,10)$

Individual-level parameters:

$\theta_{ik} \sim Gaussian(\mu^{\theta_k}, \sigma^{\theta_k})$

$\beta_i \sim Gaussian(\mu^{\beta}, \sigma^{\beta})$

$\lambda_{ik} \sim Gaussian(\mu^{\lambda_k}, \sigma^{\lambda})$

$\sigma_{ik} \leftarrow \exp(\lambda_{ik})$

$\mu_{ik} \leftarrow \beta_i + \theta_{ik}$

Likelihood:

$y_{ijk} \sim Gaussian(\mu_{ik}, \sigma_{ik})$

Consider a case in which $N_K \geq 3$. The parameter $\mu^{\theta_k}$ is constrained as:

$$0 = \sum_{k=1}^{N_K} \mu^{\theta_k} \tag{A1}$$

so that $\mu^{\theta_k}$ represents the population of the participants in condition $k$ relative to the mean $\mu^{\beta}$ of the population across the conditions. The parameter $\mu^{\theta_k}$ is characterized by the effect size $d^{\theta_k}$:

$$\mu^{\theta_k} = d^{\theta_k}\sigma^{\theta_k} \tag{A2}$$

where $d^{\theta_k}$ is referred to as the effect size in condition $k$ (Lee & Wagenmakers, 2014). The effect size $d^{\theta_k}$ is regarded as a sample from a normal distribution whose mean and variance are 0 and $(\sigma^d)^2$ that are constant across the conditions (see the sub-section "Priors"). To satisfy all these conditions, $d^{\theta_k}$ and $\mu^{\theta_k}$ are formulated as:

$$\mu^{\theta_k} = d^{\theta_k}\sigma^{\theta_k} = \dot{d}^{\theta_k}\sigma^{\theta_k} - \frac{\sum_{q=1}^{N_K}(\dot{d}^{\theta_q}\sigma^{\theta_q})}{N_K} \tag{A3}$$

where $\dot{d}^{\theta_k}$ is a parameter to which a normal distribution is assigned as its prior. The mean of the prior of $\dot{d}^{\theta_k}$ is 0 and its variance $(\sigma^{\dot{d}_k})^2$ is:

$$(\sigma^{\dot{d}_k})^2 = \left(\sigma^d\right)^2 \frac{\left({N_K}^2 - N_K\right)\left(\sigma^{\theta_k}\right)^2 - \sum_{q=1}^{N_K}\left(\sigma^{\theta_q}\right)^2}{(N_K-1)(N_K-2)\left(\sigma^{\theta_k}\right)^2} \tag{A4}$$

When $N_K = 2$, $0 = \mu^{\theta_1} + \mu^{\theta_2}$ and the effect size $d^{\theta}$ between the two conditions is defined as (Cohen, 1988):

$$d^{\theta} = \frac{\mu^{\theta_1} - \mu^{\theta_2}}{\sqrt{\frac{(\sigma^{\theta_1})^2 + (\sigma^{\theta_2})^2}{2}}} \tag{A5}$$

The prior of $d^{\theta}$ is a normal distribution whose mean and variance are 0 and $(2\sigma^d)^2$ (see the sub-section "Priors").

***Priors***

The parameter $\mu^{\lambda_k}$ and $(\sigma^{\lambda})^2$ are the mean and the variance of a distribution from which the natural logarithm $\lambda_{ik}$ of the standard deviation $\sigma_{ik}$ of each participant in condition $k$ is sampled. The prior $\mu^{\lambda_k}$ is a normal distribution whose mean and variance are 0 and 22. The prior $\sigma^{\lambda}$ is a uniform distribution between 0 and 10. These prior distributions are sufficiently wide, considering Yu et al.'s (2021) data. The natural logarithm of the sample standard deviations of the individual participants is between -1.8 and 0.60 in the data.

The parameter $\mu^{\beta}$ and $(\sigma^{\beta})^2$ are the mean and the variance of a distribution from which the mean performance $\beta_i$ of each participant across the conditions is sampled. The prior $\mu^{\beta}$ is a normal distribution whose mean and variance are 0 and 102. The prior $\sigma^{\lambda}$ is a uniform distribution between 0 and 10. These prior distributions are sufficiently wide, considering the data. The maximum and the minimum of the sample means of the individual participants are -1.4 and 0.34 in the data.

The parameter $(\sigma^{\theta_k})^2$ is the variance of a distribution from which $\theta_{ik}$ is sampled for each participant in each condition. The mean of this distribution is $\mu^{\theta_k}$. The parameter $\theta_{ik}$ represents the performance of the participant in the condition relative to $\beta_i$ of the participant. The prior of $\sigma^{\theta_k}$ is a uniform distribution between 0 and 10. This prior distribution is sufficiently wide, considering the range of the sample means of the individual participants is between -1.4 and 0.34 in the data.

When $N_K \geq 3$, the parameter $\mu^{\theta_k}$ is characterized by the effect size $d^{\theta_k}$ and $d^{\theta_k}$ is formulated with $\dot{d}^{\theta_k}$. The effect size $d^{\theta_k}$ is regarded as a sample from a normal distribution whose mean and variance are 0 and $(\sigma^d)^2$. The variance $(\sigma^d)^2$ is set to 1 (Rouder et al., 2009; Yu et al., 2021). This setting of $(\sigma^d)^2$ determines the variance $(\sigma^{\dot{d}_k})^2$ of the prior of $\dot{d}^{\theta_k}$ (Equation A4) and this is a normal distribution whose mean is 0. The effect size $d^{\theta_k}$ is computed using Equation (A3). When $N_K = 2$, the difference between $\mu^{\theta_1}$ and $\mu^{\theta_2}$ is characterized by the effect size $d^{\theta}$. The prior of $d^{\theta}$ is a normal distribution whose mean is 0 and whose variance $(2\sigma^d)^2$ where $(\sigma^d)^2$ is set to 1.

***Difference from Yu et al.'s (2021) model***

The difference between the revised model and the original model of Yu et al. (2021) concerns the following parameters: $N_K$, $(\sigma^{\theta})^2$, and $d^{\theta_k}$.

The original model of Yu et al. (2021) only considered cases in which the number of conditions $N_K$ was 2 and 3. This was the case because their model was specifically formulated for analyzing the results of their psychophysical experiment.

The most important difference between the revised model and the original model of Yu et al. (2021) concerns $(\sigma^{\theta_k})^2$. The variance of the performance of the participnts' population in condition $k$ is characterized as $(\sigma^{\beta})^2 + (\sigma^{\theta_k})^2$. The parameter $(\sigma^{\beta})^2$ is constant across the conditions for both the revised model and the original model. The parameter $(\sigma^{\theta_k})^2$ varies depending on the conditions in the revised model. However, the original model assumed that $(\sigma^{\theta_k})^2$ was constant across the conditions. The parameter $(\sigma^{\theta_k})^2$ was referred to as $(\sigma^{\theta})^2$.

Yu et al. (2021) formulated $d^{\theta_k}$ in a simpler but less-precise way than the revised model does when $N_K = 3$. The effect sizes $d^{\theta_1}$ and $d^{\theta_3}$ in conditions 1 and 3 (0.7 m and 2.3 m viewing distance) were samples taken from their prior, which was a standard normal distribution, whose mean and variance were 0 and 1. The effect size $d^{\theta_2}$ in condition 2 (1.5 m viewing distance) was computed as $d^{\theta_2} = -d^{\theta_1} - d^{\theta_3}$. Note that $0 = d^{\theta_1}\sigma^{\theta} + d^{\theta_2}\sigma^{\theta} + d^{\theta_3}\sigma^{\theta}$ and Equation (A1) was satisfied. With this formulation, $d^{\theta_2}$ could be regarded as a sample from a normal distribution whose mean and variance were 0 and 2. The variance of the distribution of $d^{\theta_2}$ was larger than the variance of the prior of $d^{\theta_1}$ and $d^{\theta_3}$. I confirmed that this difference in the variances did not produce any observable effects on the results of the analysis of Yu et al.'s (2021) data. Note that, when $N_K = 2$, the prior $d^{\theta_1}$ is a standard normal distribution and $d^{\theta_2}$ is computed as $d^{\theta_2} = -d^{\theta_1}$.

**Open practices**

The *Open Data* badge was earned because the data of the experiment(s) are available on osf.io/4tu5h/
The *Open Material* badge was earned because supplementary material(s) are available on osf.io/4tu5h/

CrossMark
click for updates

**Citation**

Sawada, T. (2023). Effects of violating the assumptions of equal variance and independent and identically distributed random variables: A case using Bayesian statistical modeling. *The Quantitative Methods for Psychology*, *19*(3), 281–295. doi: 10.20982/tqmp.19.3.p281.



Received: 13/03/2023 ~ Accepted: 14/08/2023