Statistical Analysis and Results

# 1. Statistical Analysis

The present analysis was not pre-registered as we had no a priori hypotheses and thus was considered exploratory. Inferential statistics were treated as highly unstable local descriptions of the relations between model assumptions and data in order to acknowledge the inherent uncertainty in drawing generalised inferences from single and small samples (Amrhein, Trafimow, et al., 2019). For all analyses we opted to avoid dichotomising the existence of effects and therefore did not employ traditional null hypothesis significance testing on parameter estimates (Amrhein, Greenland, et al., 2019; McShane et al., 2019). Instead, we opted to take an estimation-based approach instead (Cumming, 2014), based within a Bayesian framework (Kruschke & Liddell, 2018). For all analyses model parameter estimates and their precision, along with conclusions based upon them, were interpreted continuously and probabilistically, considering data quality, plausibility of effect, and previous literature, all within the context of each model. We focused primarily on qualitative examination of our results based on visualization of the data and models for fixed effects, and exploration of variances using random effects.

All analysis was performed in R (version 4.2.3, The R Foundation for Statistical Computing, 2023) and all data and code is presented in the supplementary materials (https://osf.io/32758/). The dependent variable in our model was the perception of assistance provided which we transformed to the interval. As such we used beta regression for our model fit using the brms package (Bürkner, 2017). data was handled in long format with each row corresponding to an observation of a participants rating of the perception of assistance being provided. As such, participants had multiple rows of observations i.e., up to two observations (one for each forced repetition) for them as trainee, and up to two observations for them as spotter. Fixed effects for the actual assistance provided by the spotter (as a percentage of the load lifted by the trainee), the role of the participant for this observation (either the trainee or the spotter), the forced repetition number (either the first or the second repetition), and each of their two way interactions (i.e., actual assistance:role, actual assistance:force repetition, role:forced repetition) were included. We also used a maximal random effect structure for the parameters including random intercepts for participant and all of the aforementioned effects including interactions as random slopes. Given the novel study design and subject matter we did not have a clear intuition or informed opinion about what prior to set and so opted to use the default weakly regularising priors and “let the data speak”. Four Monte Carlo Markov Chains with 4000 warmup and 4000 sampling iterations were used in each model. All parameters in the model had values , trace plots demonstrated chain convergence, and the posterior predictive checks appeared appropriate (see https://osf.io/9s6jq). Model results were visualised by taking draws from the expected posterior distribution (n=4000) and taking the mean of these draws along with the 95% quantile (credible) interval for the fixed effects parameters, thus providing the overall grand mean effects for the population. As we were partly interested in exploring the identity of the relationships between actual assistance provided and perception of assistance provided, yet a nonlinear beta regression model was used in which the model parameters are on the latent logit scale and are not directly interpretable, we used the marginaleffects package (Arel-Bundock et al., 2022) and examined posterior draws for the average marginal effects (i.e., the slope) for the overall grand mean effects of fixed effects parameters at different values of actual capacity (0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%) across forced repetition number and role. We examined the conditional effects by taking draws from the expected posterior distribution incorporating both fixed and random effects and their uncertainty taking the mean of these draws along with all levels of quantile (credible) interval in order to examine the heterogeneity in effects. Lastly, we compared the fixed and random effects estimates for each model term and their ratio to determine if effect heterogeneity was meaningful where a rule of thumb of its standard deviation (random effect) being 0.25 or greater than its average (fixed) effect suggests noteworthy heterogeneity (Bolger et al., 2019). All data visualisations were made using ggplot2 (Wickham et al., 2022), the tidybayes package (Kay & Mastny, 2022), and the patchwork package (Pedersen, 2022). For interpretation of model estimates and visualisation we rescaled the expected posterior distribution of the interval back to the percentage scale.

# 2. Results

## 2.1 Main Effects

The overall grand means from the model for the fixed effects (i.e., without including the random effects) can be seen in [Figure 1](#fig-model-data-plot) across both the first and second forced repetitions, and for both trainee and spotter roles. The marginal effects for each independent variable are shown in [Figure 2](#fig-marg-effs-data-plot). The slopes for actual assistance provided were closer to the identity (i.e., a slope of one meaning that a one unit increase in actual assistance provided was related to a one unit increase perception of assistance provided) over the range in which there were most observations available (~20-40% actual assistance) in forced repetition one and for both trainee and spotter roles. During repetition two there was a tendency in both trainees and spotters for increases in the actual assistance provided to be perceived as a lesser change in assistance (i.e., slopes less than one; panel (A) of [Figure 2](#fig-marg-effs-data-plot)). Both trainees and spotters tended to perceive the assistance being provided as greater during the first repetition, and also over lower levels of actual assistance (panel (B) of [Figure 2](#fig-marg-effs-data-plot)). A slight interaction effect of actual assistance provided upon the slope of repetition number can also be seen in panel (B) of [Figure 2](#fig-marg-effs-data-plot) (i.e., a reversal of the slope sign suggesting at higher actual assistance repetition two was perceived as receiving less assistance). There was little interaction effects of role (panel (C) of [Figure 2](#fig-marg-effs-data-plot)), but trainees did tend to perceive the assistance being provided by spotters as being greater than it actually was to a greater degree than spotters themselves, at least over lower levels of actual assistance.

|  |
| --- |
| Figure 1: Global grand means and credible interval estimates from the expectation of the posterior predictive distribution for the relationship of actual assistance provided with perception of assistance provided over both forced repetitions and participant roles. The dashed line across the identity is provided to aid with visual interpretation where strong identity would be indicated where model estimates straddle this line, where they fall below the line it would indicate that perception underestimates actual assistance provided, and vice versa where they fall above the line. |

|  |
| --- |
| Figure 2: Average marginal effect estimates (i.e., slopes) from expectations of the posterior predictive distributions for (A) the relationship actual assistance provided with perception of assistance provided over both forced repetitions and participant roles, (B) the effect of forced repetition number over both actual assistance provided and participant role, and (C) the effect of participant role over actual assistance provided and forced repetition number. For panel (A) the dashed line indicates a slope value of one for actual assistance provided and perception of assistance provided, where a value of one indicates strong identity i.e., a one to one change in the magnitude of perception for a change in the magnitude of actual assistance provided. In both other panels the dashed line indicates a value of zero. |

# 3. Heterogeneity of effects

Despite these qualitative trends in our modeled results, there was evidently considerable variance in effects between participants. [Figure 3](#fig-individual-data-plot) shows the individual data for each participant, and [Figure 4](#fig-cond-effs-data-plot) shows the conditional effects from the expected posterior distribution (i.e., including the fixed and random effects and their uncertainty). With the exception of spotter at low levels of actual assistance provided, the variance in predicted effects covered almost the entire range of the dependent variable (i.e., perception of assistance provided; 0-100%) suggesting all effects to be quite heterogeneous. [Table 1](#tbl-random-effs) shows that all model terms, with the exception of the intercept and actual assistance provided, showed noteworthy heterogeneity (i.e., ratio of standard deviation to mean estimate )

|  |
| --- |
| Figure 3: Individual data plot. Each panel shows the raw data for an individual participant. |

|  |
| --- |
| Figure 4: Conditional estimates plot. The dashed line across the identity is provided to aid with visual interpretation where strong identity would be indicated where model estimates straddle this line, where they fall below the line it would indicate that perception underestimates actual assistance provided, and vice versa where they fall above the line. |

Table 1: Parameter estimates for both fixed and random effects (i.e., standard deviations between participants) on the logit scale.

|  | Means | | | Standard Deviations | | | Heterogeneity |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model Term | Estimate (Mean) | Lower 95% CI (Mean) | Upper 95% CI (Mean) | Estimate (SD) | Lower 95% CI (SD) | Upper 95% CI (SD) | Ratio of SD to Mean Estimate |
| Intercept | -3.77 | -4.68 | -2.88 | 0.41 | 0.02 | 0.96 | -0.11 |
| Actual Assistance Provided | 0.07 | 0.04 | 0.11 | 0.02 | 0.00 | 0.04 | 0.23 |
| Role: (Trainee) | 1.38 | 0.40 | 2.38 | 2.75 | 2.10 | 3.53 | 1.98 |
| Forced Repetition Number | 1.30 | 0.63 | 1.96 | 0.60 | 0.26 | 0.98 | 0.46 |
| Actual Assistance Provided \* Role: (Trainee) | -0.01 | -0.04 | 0.02 | 0.02 | 0.00 | 0.04 | -1.80 |
| Actual Assistance Provided \* Forced Repetition Number | -0.02 | -0.05 | 0.00 | 0.01 | 0.00 | 0.02 | -0.28 |
| Role: (Trainee) \* Forced Repetition Number | -0.34 | -1.11 | 0.43 | 1.80 | 1.35 | 2.35 | -5.25 |
| Note: |  |  |  |  |  |  |  |
| CI = credible interval |  |  |  |  |  |  |  |
| SD = standard deviation |  |  |  |  |  |  |  |

# 4. References

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature*, *567*(7748), 305–307. <https://doi.org/10.1038/d41586-019-00857-9>

Amrhein, V., Trafimow, D., & Greenland, S. (2019). Inferential Statistics as Descriptive Statistics: There Is No Replication Crisis if We Don’t Expect Replication. *The American Statistician*, *73*(sup1), 262–270. <https://doi.org/10.1080/00031305.2018.1543137>

Arel-Bundock, V., Diniz, M. A., & Greifer, N. (2022). *Marginaleffects: Marginal Effects, Marginal Means, Predictions, and Contrasts*. <https://CRAN.R-project.org/package=marginaleffects>

Bolger, N., Zee, K. S., Rossignac-Milon, M., & Hassin, R. R. (2019). Causal processes in psychology are heterogeneous. *Journal of Experimental Psychology: General*, *148*(4), 601–618. <https://doi.org/10.1037/xge0000558>

Bürkner, P.-C. (2017). Brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*, 1–28. <https://doi.org/10.18637/jss.v080.i01>

Cumming, G. (2014). The New Statistics: Why and How. *Psychological Science*, *25*(1), 7–29. <https://doi.org/10.1177/0956797613504966>

Kay, M., & Mastny, T. (2022). *Tidybayes: Tidy Data and ’Geoms’ for Bayesian Models*. <https://CRAN.R-project.org/package=tidybayes>

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*(1), 178–206. <https://doi.org/10.3758/s13423-016-1221-4>

McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon Statistical Significance. *The American Statistician*, *73*(sup1), 235–245. <https://doi.org/10.1080/00031305.2018.1527253>

Pedersen, T. L. (2022). *Patchwork: The Composer of Plots*. <https://CRAN.R-project.org/package=patchwork>

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., & RStudio. (2022). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. <https://CRAN.R-project.org/package=ggplot2>