

ML4 Results Section in rMarkdown

Richard A. Klein¹

¹ Université Grenoble Alpes

Author Note

This script generates the participants + results sections for the main ML4 manuscript. To knit this document you must install the papaja package from GitHub.

Correspondence concerning this article should be addressed to Richard A. Klein, Postal address. E-mail: raklein22@gmail.com

Abstract

Keywords: Terror Management Theory, mortality salience, replication, many labs

Word count: X

ML4 Results Section in rMarkdown

Participants**NOTE: CONFIRM FOOTNOTES ARE SHOWING UP**

21 labs participated and provided a total sample of 2,281 participants. In accordance with the pre-registration, we immediately excluded from all analyses participants who either failed to complete all 6 ratings of the essay authors, or who failed to complete both writing prompts within the mortality salience or control conditions (e.g., the between-subjects manipulation).¹ Thus, the usable $N = 2,220$ (see Table XX for a summary of sites). 1,157 participants (52.12%) reported being female and 708 participants (31.89%) reported being male; the remaining participants did not respond to the item, were asked about gender in a non-standard way, or chose a different response. Mean age was 19.87 years ($SD = 2.79$). Participant reported race was 910 (40.99%) White, 221 (9.95%) Asian, 120 (5.41%) Black or African American, 36 (1.62%) American Indian or Alaska Native, 20 (0.90%) Native Hawaiian or Pacific Islander, 114 (5.14%) Other. The remaining participants did not report their race, were not asked the item, or were asked the item in a non-standard way.

Data analysis

We used R (Version 3.6.0; R Core Team, 2019) and the R-packages *dplyr* (Version 0.8.1; Wickham et al., 2019), *effsize* (Version 0.7.4; Torchiano, 2018), *extrafont* (Version 0.17; Winston Chang, 2014), *forcats* (Version 0.4.0; Wickham, 2019a), *gginnards* (Version 0.0.2; Aphalo, 2016), *ggplot2* (Version 3.1.1; Wickham, 2016), *GPArotation* (Version 2014.11.1; Bernaards & I.Jennrich, 2005), *haven* (Version 2.1.0; Wickham & Miller, 2019), *Matrix* (Version 1.2.17; Bates & Maechler, 2019), *metafor* (Version 2.1.0; Viechtbauer, 2010), *metaSEM* (Version 1.2.2; Cheung, 2015), *metaviz* (Version 0.3.0; Kossmeier, Tran, &

Voracek, 2019), *OpenMx* (Version 2.13.2; Neale et al., 2016; Hunter, 2018; Pritikin, Hunter, & Boker, 2015), *papaja* (Version 0.1.0.9842; Aust & Barth, 2018), *psych* (Version 1.8.12; Revelle, 2018), *purrr* (Version 0.3.2; Henry & Wickham, 2019), *readr* (Version 1.3.1; Wickham, Hester, & Francois, 2018), *stringr* (Version 1.4.0; Wickham, 2019b), *tibble* (Version 2.1.3; Müller & Wickham, 2019), *tidyr* (Version 0.8.3.9000; Wickham & Henry, 2019), and *tidyverse* (Version 1.2.1.9000; Wickham, 2017) for all our analyses.

Results

Researcher Expectations and Characteristics

A total of 28 researchers from 21 participating sites completed the experimenter survey. Psychology research experience ranged from 0 to 28 years ($M = 9.32$, $SD = 8.80$). 1 (3.57%) researcher indicated they were an expert in TMT, 5 (17.86%) indicated they had “a lot” of TMT knowledge, 10 (35.71%) indicated “some” knowledge, 5 (17.86%) indicated little knowledge, 6 (21.43%) indicated zero knowledge, and 1 (3.57%) did not respond to the question.

When asked what outcome they wanted to happen, 13 (46.43%) indicated that they hoped for the project to successfully replicate the TMT effect, 10 (35.71%) indicated no preference, and 3 (10.71%) hoped the project would result in a failure to replicate, with 2 (7.14%) researchers leaving the question blank. On average, the teams estimated a 54.37% chance of successful replication with a wide range of estimates from 20% to 95% ($SD = 22.14$).

TMT Replication Results

The primary finding of interest from Greenberg et al., 1994, was that participants who underwent the mortality salience treatment showed greater preference for the pro-US essay author compared to the anti-US essay author. To assess whether the replication results support the original, we followed a similar analysis plan as in the original article. Scores from the three items evaluating the authors of the anti-American essays were averaged ($\alpha = 0.90$) and then subtracted from the average of the three items evaluating authors of the pro-American essays ($\alpha = 0.89$).² An independent samples t-test was then conducted comparing those in the “subtle own death salient” (MS) condition with scores from the “TV salient” (control) condition. Original authors were not entirely in agreement about what exclusions should be implemented. So, we repeated our analyses under different exclusion criteria suggested by original authors:

Exclusion set 1: Include all participants who completed the materials (e.g., wrote something for both writing prompts, and completed all six items evaluating the essay authors). Reduces the usable N from 2,281 to 2,220 participants.

Exclusion set 2: All prior exclusions, and further exclude participants who did not identify as White or who indicated they were born outside the United States. Reduces N to 1,874.

Exclusion set 3: All prior exclusions, and further exclude participant who responded lower than 7 on the American Identity item. Further reduces the usable N to 1,843 participants.

Exclusion sets 2 and 3 were specifically recommended by original authors and these criteria were used to analyze the data from Author Advised labs. However, the data required to make these exclusions were often not collected at In House replication sites because they

made independent decisions about design and demographic measures for potential exclusion, and these measures were not in the original article. Thus, for all analyses only Exclusion Set 1 was used for In House participants. The full data cleaning and analysis scripts are available on the OSF (<https://osf.io/8ccnw/>). All analysis plans and procedures were pre-registered on the OSF prior to data collection.

Meta-analytic results across all labs (random effects meta-analysis).

Some labs administered both Author Advised and In House protocols. To account for this nesting of effect sizes within labs, a three-level random-effects meta-analysis was conducted using the MetaSEM package (Cheung, 2014) in R.³ This analysis produces the grand mean effect size across all sites and versions.

The most basic question is whether we found the predicted effect of mortality salience. Regardless of which exclusion criteria were used, we observed an effect that was not consistent with the predicted mortality salience effect: Exclusion set 1: *Hedges'* $g = 0.03$, 95% CI = [-0.05, 0.11], $SE = 0.04$, $Z = 0.71$, $p = 0.48$. Exclusion set 2: *Hedges'* $g = 0.07$, 95% CI = [-0.04, 0.19], $SE = 0.06$, $Z = 1.26$, $p = 0.21$. Exclusion set 3: *Hedges'* $g = 0.04$, 95% CI = [-0.09, 0.18], $SE = 0.07$, $Z = 0.61$, $p = 0.54$. Forest plots showing the effects for individual sites and the aggregate are available in Figure XX for exclusion set 1, and on the OSF page for the other two exclusion rules (which look very similar).

We also examined how much variation was observed among effect sizes (e.g., heterogeneity). For example, there may have been a mortality salience effect at some sites and not others. For exclusion sets 1 and 3, this sort of variation did not exceed variation expected by chance (e.g., sampling variance), exclusion set 1: $Q(20) = 25.82$, $p = 0.17$; exclusion set 3: $Q(20) = 29.61$, $p = 0.08$. The amount of variation between did exceed chance for exclusion set 2, $Q(20) = 36.32$, $p = 0.01$. However, though statistically significant,

it was small in magnitude, $\text{Tau}^2_{\text{within labs}} = 0.00$, $\text{Tau}^2_{\text{between labs}} = 0.02$.

In sum, we observed no evidence for an overall effect of mortality salience in these replications. And, overall there was only weak evidence of heterogeneity in effect sizes. This lack of variation suggests that it is unlikely we will observe an effect of Author Advised versus In House protocols or other moderators such as differences in samples or TMT knowledge. Even so, the plausible moderation by Author Advised/In House protocol is examined in the following section.

Research Question 2: Moderation by Author Advised/In House protocol

A covariate of protocol type was added to the random effects model to create a three-level mixed-effects meta-analysis. The addition of the argument “x = version” to the prior metaSEM R code can be seen here:

```
meta3(y=es, v=var, cluster=Location, x=version, data=dataset)
```

NOTES: [Rick] This is the primary preregistered analysis NOTES: [Brian]
See the prior section for some idea of shortening. Just provide the summary of the first analysis meaning and then show just the analytic outcomes for the 2nd and 3rd exclusion criteria. You also can remove description of those exclusions because you’ve said all that twice already. It can be one short paragraph of results.

This analysis again produces an overall grand mean effect size, and those were again null across all three exclusion sets: Exclusion set 1: *Hedges’ g* = 0.02, 95% CI = [-0.10, 0.14], *SE* = 0.06, *Z* = 0.32, *p* = 0.75. Exclusion set 2: *Hedges’ g* = 0.04, 95% CI = [-0.10, 0.17], *SE* = 0.07, *Z* = 0.51, *p* = 0.61. Exclusion set 3: *Hedges’ g* = 0.02, 95% CI = [-0.11, 0.16], *SE* = 0.07, *Z* = 0.34, *p* = 0.73.

This analysis additionally tests protocol version (In House vs Author Advised) as a moderator. Across the three exclusion sets, protocol version did not significantly predict replication effect size. Exclusion set 1: $b = 0.02$, $Z = 0.29$, $p = 0.77$; exclusion set 2: $b = 0.11$, $Z = 0.92$, $p = 0.36$; exclusion set 3: $b = 0.10$, $Z = 0.57$, $p = 0.57$.

Variation among effect sizes also followed the previously observed pattern. Modest (but statistically significant) heterogeneity for exclusion set 2, $Q(20) = 36.32$, $p = 0.01$, $\text{Tau}^2_{\text{within labs}} = 0.00$, $\text{Tau}^2_{\text{between labs}} = 0.02$; while variation did not exceed chance for exclusion sets 1 $Q(20) = 25.82$, $p = 0.17$; or exclusion set 3: $Q(20) = 29.61$, $p = 0.08$.

Research Question 3: Effect of Standardization

Finally, we tested whether In House protocols displayed greater variability in effect size than Author Advised protocols. To test this hypothesis, we ran the mixed-effects models but constrained the variances at both level 2 and level 3 to 0, effectively creating fixed-effects models. These models were then compared with a chi-squared differences test to assess whether the fit significantly changed. In this case, none of the three models significantly decreased in fit: Exclusion rule #1: $\chi^2(2) = 0.16$, $p = 0.92$; Exclusion rules #2: $\chi^2(2) = 0.85$, $p = 0.65$; Exclusion rules #3: $\chi^2(2) = 0.27$, $p = 0.87$. Overall, there was no evidence that In House protocols elicited greater variability than Author Advised protocols.

Follow-Up Exploratory Analyses

Results for TMT-knowledgeable sites. One principal investigator reported being an expert in TMT, while 5 others indicated having “a lot” of knowledge about TMT. One might expect that these locations would have greater success at replicating the mortality salience effect. Aggregating across these sites, and using only the first exclusion rule, these sites did not detect a difference between the mortality salience group ($M = 1.05$, $SD = 2.22$)

and the control group ($M = 0.97$, $SD = 2.19$), $t(522.61) = 0.45$, $p = 0.66$, *Hedges'* $g = 0.04$, 95% CI = [-0.13, 0.21].

Discussion

References

- Aphalo, P. J. (2016). *Learn r ...as you learnt your mother tongue*. Leanpub. Retrieved from <https://leanpub.com/learnr>
- Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown*. Retrieved from <https://github.com/crsh/papaja>
- Bates, D., & Maechler, M. (2019). *Matrix: Sparse and dense matrix classes and methods*. Retrieved from <https://CRAN.R-project.org/package=Matrix>
- Bernaards, C. A., & Jennrich, R. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement*, 65, 676–696.
- Cheung, M. W.-L. (2015). metaSEM: An r package for meta-analysis using structural equation modeling. *Frontiers in Psychology*, 5(1521). doi:10.3389/fpsyg.2014.01521
- Henry, L., & Wickham, H. (2019). *Purrr: Functional programming tools*. Retrieved from <https://CRAN.R-project.org/package=purrr>
- Hunter, M. D. (2018). State space modeling in an open source, modular, structural equation modeling environment. *Structural Equation Modeling*, 25(2), 307–324. doi:10.1080/10705511.2017.1369354
- Kossmeier, M., Tran, U. S., & Voracek, M. (2019). *Metaviz: Forest plots, funnel plots, and visual funnel plot inference for meta-analysis*. Retrieved from <https://CRAN.R-project.org/package=metaviz>
- Müller, K., & Wickham, H. (2019). *Tibble: Simple data frames*. Retrieved from <https://CRAN.R-project.org/package=tibble>

- Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., ... Boker, S. M. (2016). OpenMx 2.0: Extended structural equation and statistical modeling. *Psychometrika*, 81(2), 535–549. doi:10.1007/s11336-014-9435-8
- Pritikin, J. N., Hunter, M. D., & Boker, S. M. (2015). Modular open-source software for Item Factor Analysis. *Educational and Psychological Measurement*, 75(3), 458–474.
- R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Revelle, W. (2018). *Psych: Procedures for psychological, psychometric, and personality research*. Evanston, Illinois: Northwestern University. Retrieved from <https://CRAN.R-project.org/package=psych>
- Torchiano, M. (2018). *Effsize: Efficient effect size computation*. doi:10.5281/zenodo.1480624
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. Retrieved from <http://www.jstatsoft.org/v36/i03/>
- Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from <https://ggplot2.tidyverse.org>
- Wickham, H. (2017). *Tidyverse: Easily install and load the 'tidyverse'*. Retrieved from <https://CRAN.R-project.org/package=tidyverse>
- Wickham, H. (2019a). *Forcats: Tools for working with categorical variables (factors)*. Retrieved from <https://CRAN.R-project.org/package=forcats>
- Wickham, H. (2019b). *Stringr: Simple, consistent wrappers for common string operations*. Retrieved from <https://CRAN.R-project.org/package=stringr>

Wickham, H., François, R., Henry, L., & Müller, K. (2019). *Dplyr: A grammar of data manipulation*. Retrieved from <https://CRAN.R-project.org/package=dplyr>

Wickham, H., & Henry, L. (2019). *Tidyr: Easily tidy data with 'spread()' and 'gather()' functions*. Retrieved from <https://CRAN.R-project.org/package=tidyr>

Wickham, H., Hester, J., & François, R. (2018). *Readr: Read rectangular text data*. Retrieved from <https://CRAN.R-project.org/package=readr>

Wickham, H., & Miller, E. (2019). *Haven: Import and export 'spss', 'stata' and 'sas' files*. Retrieved from <https://CRAN.R-project.org/package=haven>

Winston Chang. (2014). *Extrafont: Tools for using fonts*. Retrieved from <https://CRAN.R-project.org/package=extrafont>

Footnotes

¹The latter exclusion criteria applied only to participants from Author Advised sites, because the necessary data was not always available for In House sites.

²Supplemental analyses treating these as two separate dependent variables are available in the online supplement (<https://osf.io/xtg4u/>) but generally follow the same pattern of results reported below.

³Sample code to run this analysis is: `meta3(y=es, v=var, cluster=Location, data=dataset)`. In this sample code, “y=es” directs the program to the column of effect sizes, “v=var” indicates the variable to be used as the sampling variance for each effect size, and the “cluster=Location” command groups the effect sizes by a location variable in the dataset (in this case, a unique identifier assigned to each replication site).