

How Foreign Language Impacts Moral Decision Making

Lindsey King

University of Chicago

Introduction

Geipel et al. (2015) tested the effects of foreign language on moral decision making, finding that participants were more consequentialist, or willing to take an action that may be unmoral to ultimately save a life, when they are asked in their foreign language. It is also seen that people are in general less consequentialist when presented with the footbridge dilemma compared to the original trolley dilemma regardless of language. This is likely because of the increased involvement in the footbridge dilemma (having to physically push someone onto the train track) compared to the trolley dilemma (Greene et al., 2009). This involvement in the dilemma is tied to higher levels of emotion as well which has been shown to lead to less consequentialist choices in moral dilemmas (Huebner et al., 2009).

Given this information, it is plausible that foreign language would lead to higher consequentialist responses. This is because we know that people are more rational and systematic and less emotional in their decision making in a foreign language (Cipolletti et al., 2016; Costa et al., 2014). This lessened emotion and heightened rationality is what leads individuals speaking a foreign language to be more consequentialist in their decision making. They are more removed from the situation than they would be in their native language and can thus judge the pros and cons from a purely moral standpoint. This is why we see more individuals that would be willing to push a man off a bridge to save five lives in the foreign language condition. This choice would weigh on your moral conscience, but it ultimately saves more lives and is thus the more rational decision between the two.

One possible other reason for increased consequentialism in a foreign language is that the individual reading the scenario in their foreign language may assume that the individuals in the dilemma are in a foreign country where that language is spoken (Geipel et al., 2015). This is a different example of how the foreign language condition is potentially distancing people from the scenario allowing them to take a more consequentialist approach that they may not take in their native language.

Methods

The current study seeks to replicate the figures and analyses used in Geipel et al. (2015). This study assessed the foreign language effect on moral judgement by presenting participants with the footbridge and trolley dilemma. A non-moral dilemma was also designed and presented to test for understanding in the foreign language. If it was understood, it should have high endorsements in all conditions (Geipel et al., 2015). Two foreign language conditions were included (both English and German) to test if the effect is generalizable across different languages. We will not run every analysis included in Study 1 of Geipel et al. (2015), but rather will run an analysis to see the general trends of the effects while presenting figures and tables that suggest the more nuanced effects described in the original study. This is simply a replication aimed at recreating the figures and findings using the same data set as the original study.

Participants

One hundred and five students were recruited from The University of Trento in Trento, Italy. These students are native Italian speakers enrolled in either German or English courses. Participants were assigned to one of three conditions: 39 to the native language Italian condition, 37 to the foreign language English condition, and 29 to the foreign language German condition (Geipel et al., 2015). Those assigned to a foreign language condition were individuals that ranked six or higher on a ten point scale of foreign language proficiency. This was done to ensure that participants in the foreign language condition would be proficient enough in their foreign language to properly understand and respond to the materials presented to them. So while conditions were randomly assigned, they were randomly assigned with these parameters in place to ensure accuracy.

Material

This is a replication of Study 1 in the paper by Geipel et al. (2015). The trolley problem originates from Thomson (1985) and is adapted to test moral decision making across language conditions. Exact questions provided were modified from Greene et al.

(2001). Participants are presented with three moral dilemmas with different levels of personal involvement with the expectation that the less involved the participant is in the scenario, the easier it will be for them to make the rational, moral decision (Hare, 1981). Translated to the trolley problem, individuals are more likely to push a button to redirect a train killing one person and saving five, but are less likely to push a person off a bridge to stop the train and save five (Cushman et al., 2006).

Procedure

Participants were presented with each dilemma (footbridge, trolley, and non-moral) in a randomized and counterbalanced order. Each dilemma was presented to the participant containing an action like pushing the man or the button. They were then asked to select if they would (consequentialist response) or would not (deontological response) take that action by selecting Yes or No. Participants only took the assessment once in the language condition they were randomly assigned to, and materials were presented entirely in the language condition assigned.

Table 1

Demographics With Consequentialist Means

Condition	N	Trolley Mean %	Footbridge Mean %	Non-Moral Mean %
Italian	39	54%	13%	87%
English	28	73%	43%	92%
German	36	61%	36%	93%

Data analysis

We used R (Version 4.3.2; R Core Team, 2023) and the R-packages *broom* (Version 1.0.5; Robinson et al., 2023), *car* (Fox et al., 2022; Version 3.1.2; Fox & Weisberg, 2019), *carData* (Version 3.0.5; Fox et al., 2022), *dplyr* (Version 1.1.3; Wickham, François, et al., 2023), *forcats* (Version 1.0.0; Wickham, 2023a), *ggplot2* (Version 3.4.4; Wickham, 2016), *ggsci* (Version 3.0.0; Xiao, 2023), *kableExtra* (Version 1.4.0; Zhu, 2024), *knitr* (Version 1.45;

Xie, 2015), *koRpus* (Michalke, 2020a; Version 0.13.8; Michalke, 2021), *koRpus.lang.en* (Version 0.1.4; Michalke, 2020a), *lme4* (Version 1.1.35.1; Bates et al., 2015), *lubridate* (Version 1.9.3; Grolemund & Wickham, 2011), *Matrix* (Version 1.6.5; Bates et al., 2024), *papaja* (Version 0.1.2; Aust & Barth, 2023), *psych* (Version 2.3.12; William Revelle, 2023), *purrr* (Version 1.0.2; Wickham & Henry, 2023), *readr* (Version 2.1.4; Wickham, Hester, et al., 2023), *Require* (Version 0.3.1; McIntire, 2023), *reshape2* (Version 1.4.4; Wickham, 2007), *scales* (Version 1.2.1; Wickham & Seidel, 2022), *stringr* (Version 1.5.1; Wickham, 2023b), *sylly* (Version 0.1.6; Michalke, 2020b), *tibble* (Version 3.2.1; Müller & Wickham, 2023), *tidyr* (Version 1.3.0; Wickham, Vaughan, et al., 2023), *tidyverse* (Version 2.0.0; Wickham et al., 2019), and *tinylabels* (Version 0.2.4; Barth, 2023) for all our analyses.

All statistical outputs are interpreted through the framework provided by Cumming (2013). Descriptive statistics were calculated to give the percent of consequentialist responses across language and dilemma type as seen in Table 1. An analysis of covariance (ANCOVA) was conducted to test language differences across the three conditions. The output of this ANCOVA can be seen in Table 2. Language proficiency was also considered as a potential mediator of the foreign language effect, but Figure 3 shows no result worth investigating further.

Table 2

Language and Dilemma Types ANCOVA

	Sum Sq	df	F Statistic	p-value
(Intercept)	11.168912	1	64.46800	0.0000000
language.condition	1.839164	1	10.61583	0.0012478
dilemma.type	18.718447	2	54.02231	0.0000000
Residuals	52.840448	305	NA	NA

Results

A significant effect of dilemma type on consequentialism across language conditions ($p < .001$) is seen showing that any differences in consequentialism are not due to misunderstanding in the foreign language. This is because we expect consequentialism to be highest in the non-moral dilemma ($\mu = 0.90$) and lowest in the footbridge dilemma ($\mu = 0.30$). This is because there is no moral dilemma in the non-moral dilemma and therefore no reason to not choose the consequentialist response where as the inverse is true for the footbridge dilemma. The trolley dilemma falls in the middle ($\mu = 0.63$). To understand these percentages better, they can be visualized in Figure 1 and Figure 2.

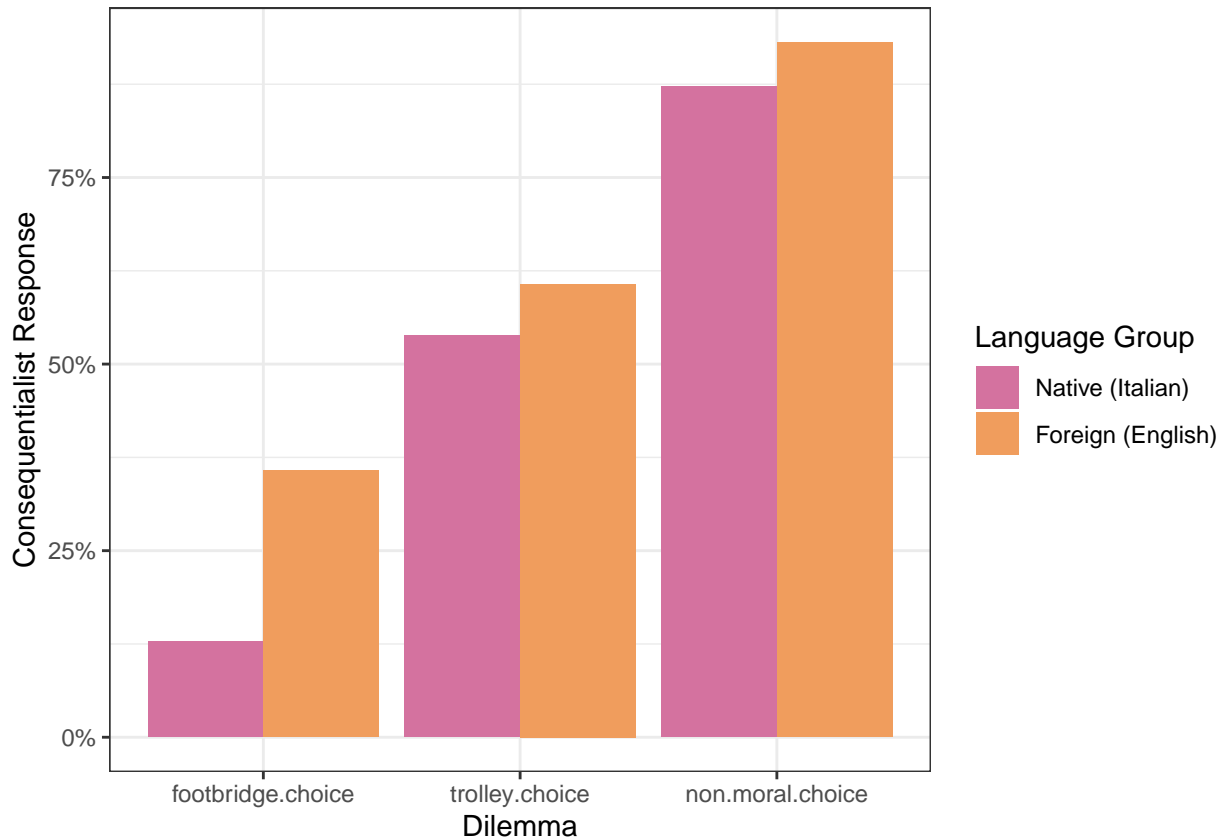
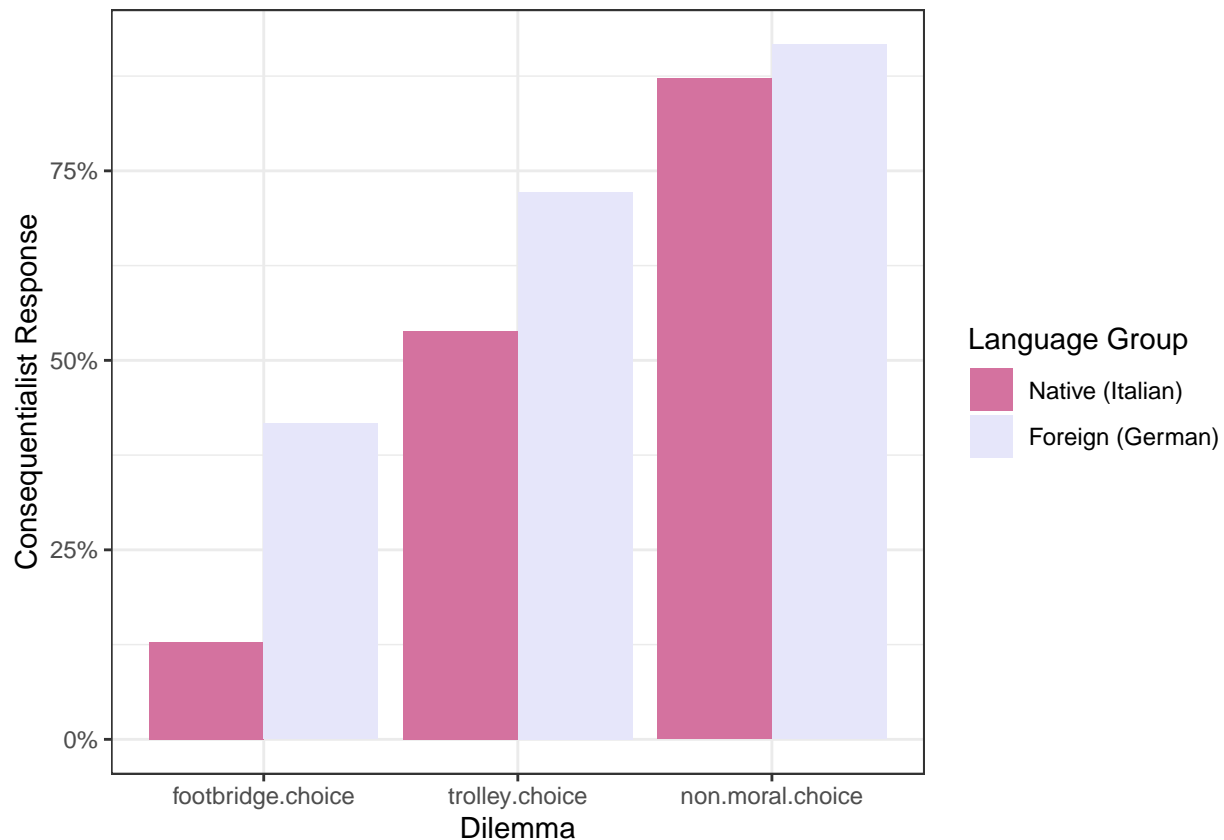


Figure 1

Italian versus English

Table 1 shows the breakdown of the language categories as well as the mean consequentialism participants exhibited for the different dilemmas. There is a significant

**Figure 2***Italian versus German*

effect of language condition on consequentialism ($p = .001$). This was run using native versus foreign language rather than testing the three separate language conditions. This is because we are not interested in the differences between English and German as they are the two foreign language groups. We are simply interested in native versus foreign language and can therefore group English and German into one category and compare that to the native Italian group. The significant effect tells us that participants do in fact respond differently to moral dilemmas depending on if they are using their native or foreign language, confirming the hypothesis in Geipel et al. (2015).

Discussion

Figure 3 was generated as an inquiry of language proficiency for those in the foreign language condition. I was curious to see if there was any difference in consequentialism as

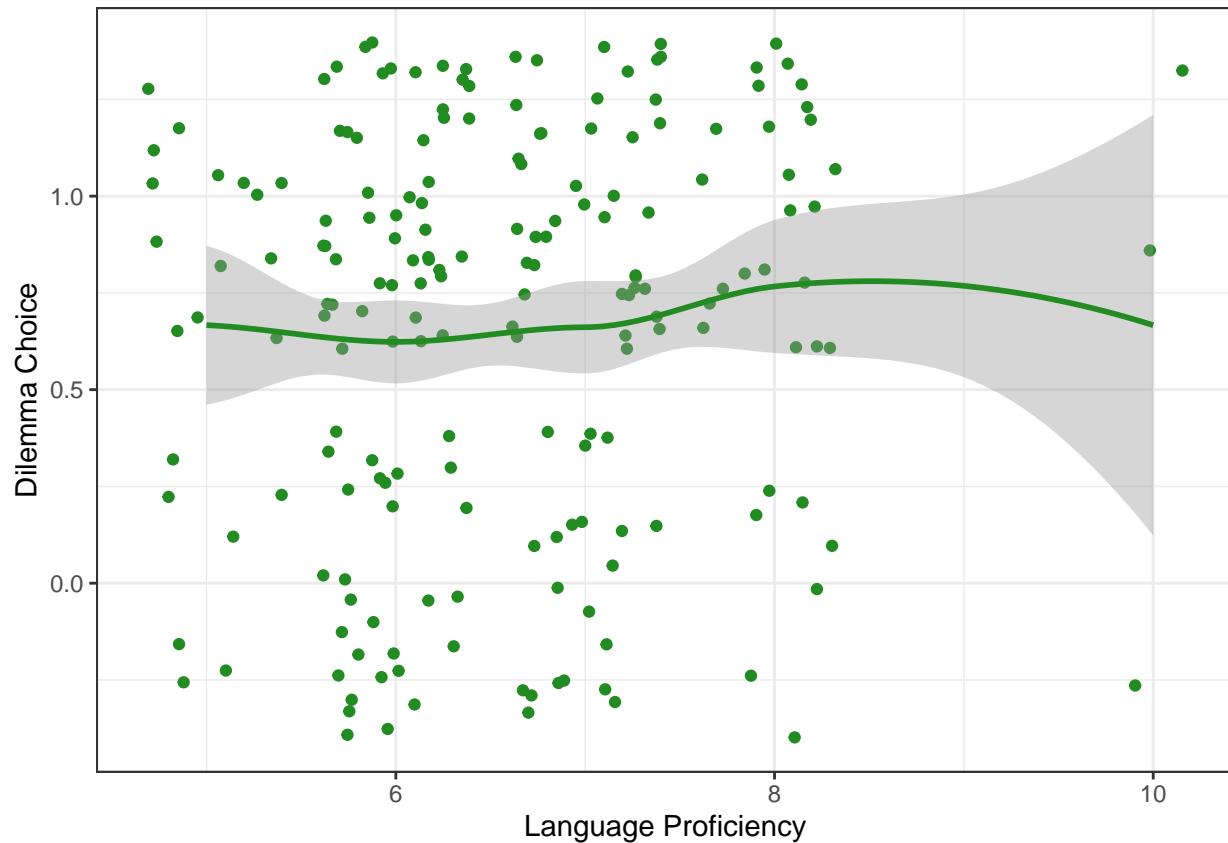


Figure 3

Proficiency of foreign language.

individuals become more proficient in their foreign language. I theorized that if a difference were to exist, individuals may become more consequential the more proficient they are in the foreign language. This is because they may be speaking the language more fluently thus relying less on translating materials to their native language in their mind and eliminating the foreign language effect. The figure shows that this does not seem to be the case. This makes sense because individuals assigned to a foreign language condition were not ranked below a six in proficiency which was ranked from one to ten. This means participants were all decently proficient in their foreign language. We can also tell this from the results of the non-moral dilemma which was designed to elicit a high consequentialist response rate when understood properly. This was the case in these participants as seen in the non-moral dilemma means in Table 1.

Results were successfully recreated from Geipel et al. (2015). Not all analyses from Study 1 were conducted, only the ANCOVA was conducted to see the total effect of language condition and dilemma type. The original study ran multiple chi-square tests to compare the responses for each dilemma and each language condition (e.g., comparing Italian versus German in the footbridge dilemma). This would result in six different chi-square tests and was beyond the scope of this paper. The analyses run in the current paper sufficiently show the general effect through its analysis and the specific effects through Figure 1 and Figure 2.

The significant effects found in this study are quite profound as the p-values for both the effects of language condition (native or foreign) and dilemma type (non-moral, trolley, and footbridge) were all highly significant ($p = .001$, $p < .001$). This tells us that those in the foreign language condition were indeed more consequentialist in their responses, as hypothesized by Geipel et al. (2015). We also see that the dilemma types successfully capture differing levels of involvement. One limitation given by Geipel et al. (2015) include the limited number of moral dilemmas presented. Future research should consider more dilemmas to present to participants to get a better understanding of the effect. We see a difference between dilemma types, so it is important to expand this knowledge.

References

- Aust, F., & Barth, M. (2023). *papaja: Prepare reproducible APA journal articles with R Markdown*. <https://github.com/crsh/papaja>
- Barth, M. (2023). *tinylabels: Lightweight variable labels*. <https://cran.r-project.org/package=tinylabels>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bates, D., Mächler, M., & Jagan, M. (2024). *Matrix: Sparse and dense matrix classes and methods*. <https://CRAN.R-project.org/package=Matrix>
- Cipolletti, H., McFarlane, S., & Weissglass, C. (2016). The Moral Foreign-Language Effect. *Philosophical Psychology*, 29(1), 23–40. <https://doi.org/10.1080/09515089.2014.993063>
- Costa, A., Foucart, A., Hayakawa, S., Aparici, M., Apesteguia, J., Heafner, J., & Keysar, B. (2014). Your Morals Depend on Language. *PLOS ONE*, 9(4), e94842. <https://doi.org/10.1371/journal.pone.0094842>
- Cumming, G. (2013). *Understanding The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. Routledge.
- Cushman, F. A., Young, L., & Hauser, M. (2006). The Role of Conscious Reasoning and Intuition in Moral Judgment: Testing Three Principles of Harm. *Psychological Science*, 17(12), 1082–1089. <https://doi.org/10.1111/j.1467-9280.2006.01834.x>
- Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (Third). Sage. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>
- Fox, J., Weisberg, S., & Price, B. (2022). *carData: Companion to applied regression data sets*. <https://CRAN.R-project.org/package=carData>
- Geipel, J., Hadjichristidis, C., & Surian, L. (2015). The Foreign Language Effect on Moral Judgment: The Role of Emotions and Norms. *PLOS ONE*, 10(7), e0131529. <https://doi.org/10.1371/journal.pone.0131529>

- Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, 111(3), 364–371.
<https://doi.org/10.1016/j.cognition.2009.02.001>
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI Investigation of Emotional Engagement in Moral Judgment | Science. *Science*, 293(5537), 2105–2108. <https://doi.org/10.1126/science.1062872>
- Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, 40(3), 1–25. <https://www.jstatsoft.org/v40/i03/>
- Hare, R. M. (1981). *Moral Thinking: Its Levels, Method, and Point*. Oxford University Press.
- Huebner, B., Dwyer, S., & Hauser, M. (2009). The role of emotion in moral psychology. *Trends in Cognitive Sciences*, 13(1), 1–6. <https://doi.org/10.1016/j.tics.2008.09.006>
- McIntire, E. J. B. (2023). *Require: Installing and loading r packages for reproducible workflows*. <https://CRAN.R-project.org/package=Require>
- Michalke, M. (2020a). *koRpus.lang.en: Language support for 'koRpus' package: english*. <https://reaktanz.de/?c=hacking&s=koRpus>
- Michalke, M. (2020b). *Sylly: Hyphenation and syllable counting for text analysis*. <https://reaktanz.de/?c=hacking&s=sylly>
- Michalke, M. (2021). *koRpus: Text analysis with emphasis on POS tagging, readability, and lexical diversity*. <https://reaktanz.de/?c=hacking&s=koRpus>
- Müller, K., & Wickham, H. (2023). *Tibble: Simple data frames*. <https://CRAN.R-project.org/package=tibble>
- R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Robinson, D., Hayes, A., & Couch, S. (2023). *Broom: Convert statistical objects into tidy tibbles*. <https://CRAN.R-project.org/package=broom>
- Thomson, J. J. (1985). The Trolley Problem. *The Yale Law Journal*, 94, 1395–1415.

- Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, 21(12), 1–20. <http://www.jstatsoft.org/v21/i12/>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>
- Wickham, H. (2023a). *Forcats: Tools for working with categorical variables (factors)*. <https://CRAN.R-project.org/package=forcats>
- Wickham, H. (2023b). *Stringr: Simple, consistent wrappers for common string operations*. <https://CRAN.R-project.org/package=stringr>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A grammar of data manipulation*. <https://CRAN.R-project.org/package=dplyr>
- Wickham, H., & Henry, L. (2023). *Purrr: Functional programming tools*. <https://CRAN.R-project.org/package=purrr>
- Wickham, H., Hester, J., & Bryan, J. (2023). *Readr: Read rectangular text data*. <https://CRAN.R-project.org/package=readr>
- Wickham, H., & Seidel, D. (2022). *Scales: Scale functions for visualization*. <https://CRAN.R-project.org/package=scales>
- Wickham, H., Vaughan, D., & Girlich, M. (2023). *Tidyr: Tidy messy data*. <https://CRAN.R-project.org/package=tidyr>
- William Revelle. (2023). *Psych: Procedures for psychological, psychometric, and personality research*. Northwestern University. <https://CRAN.R-project.org/package=psych>
- Xiao, N. (2023). *Ggsci: Scientific journal and sci-fi themed color palettes for 'ggplot2'*. <https://CRAN.R-project.org/package=ggsci>

Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Chapman; Hall/CRC.

<https://yihui.org/knitr/>

Zhu, H. (2024). *kableExtra: Construct complex table with 'kable' and pipe syntax*.

<https://CRAN.R-project.org/package=kableExtra>