

The Stroop Effect: A Reproducible Analysis

Daniël Lakens

Industrial Engineering and Innovation Sciences, Eindhoven University of Technology

Author Note

Daniël Lakens  <https://orcid.org/0000-0002-8393-5316>

All data, code, and materials are available at https://github.com/Lakens/stroop_effect_paper.
The authors have no conflict of interest to declare. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Correspondence concerning this article should be addressed to Daniël Lakens, Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, Den Dolech 2, 5600 MB, Eindhoven, Email: D.Lakens@tue.nl

Abstract

Students in an introductory psychology course completed an online Stroop task, naming the font colors of words in congruent trials (e.g., the word 'red' printed in red) and incongruent trials (e.g., the word 'red' printed in green). We tested whether reaction times differed between the two conditions, as predicted by the Stroop effect.

Keywords: Stroop effect, reaction time, reproducibility

The Stroop Effect: A Reproducible Analysis

The Stroop effect – the finding that it takes longer to name the font color of a word when that word names a different color (e.g., the word “red” printed in green ink) than when the word and its color match – is one of the most robust and widely replicated phenomena in cognitive psychology (MacLeod, 1991). Here, we analyze a simple dataset of a Stroop experiment. Students in an introduction to psychology course completed an online Stroop task (<http://faculty.washington.edu/chudler/java/ready.html>) and named the colors in congruent trials (e.g., the word ‘red’ written in a red font) and in incongruent trials (e.g., the word ‘red’ written in a green font). The time they took to name all words was self-reported in seconds (e.g., 21.3 seconds) for both the congruent and incongruent blocks. In this analysis, we are interested in examining whether there is a Stroop effect.

Ethical approval

All participants completed the task voluntarily as part of a classroom demonstration of research methods, and no identifying information was collected. Because no personally identifiable information was recorded and the task carried no more than minimal risk, this classroom demonstration was exempt from formal review by the Eindhoven University of Technology Ethics Review Board.

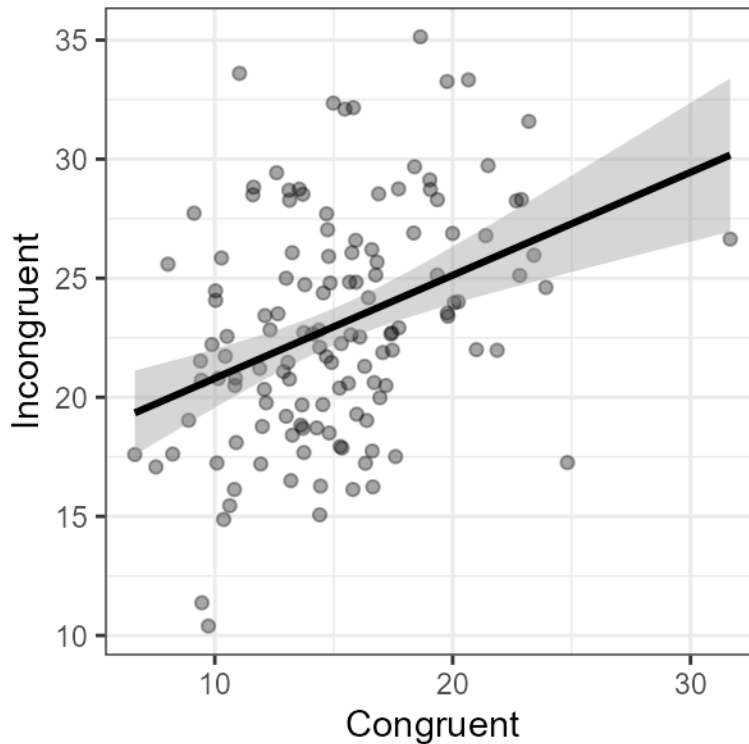
Power analysis

The Stroop effect is one of the largest and most consistently replicated effects in cognitive psychology (MacLeod, 1991), so we planned our sample size assuming a conservatively large effect.

An a priori power analysis for a paired-samples t -test, conducted using the `pwr.t.test()` function from the `pwr` package (Champely, 2020), indicated that to detect a true effect of 0.8 standard deviations of the difference scores, with an alpha level of 0.01, and a desired power of 90%, at least 27 participants would be needed. We report data from all students who completed the task in a single course offering, exceeding this minimum.

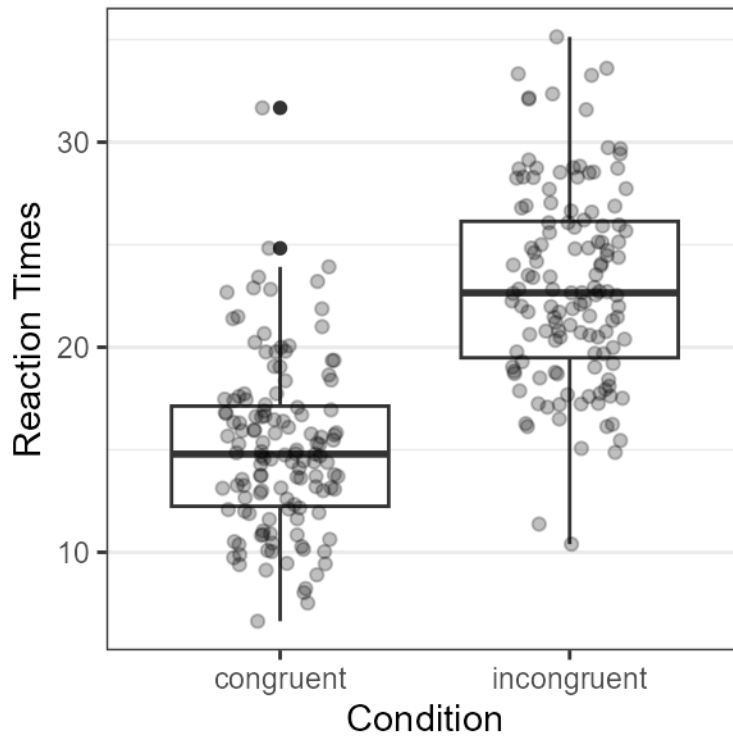
Plotting the data

When we plot the reaction times, we can visually see that there are no extreme outliers. It is also clear that response times are generally faster for the congruent trials, compared to the incongruent trials.



Results

The mean reaction time (in seconds) of participants in the Congruent condition ($M = 15.1$, $SD = 4.1$) was lower than the mean of participants in the Incongruent condition ($M = 23$, $SD = 4.78$). Reaction times in the two conditions were positively correlated, $r = 0.37$, $t(129) = 4.53$, $p < 0.001$. A dependent t -test indicated that based on our alpha level of 0.01 we could reject the null-hypothesis, $t(130) = 18.04$, $p < 0.001$, Cohen's $d_z = 1.58$, 95% CI [1.32, 1.83] (Lakens, 2013). As we can expect from the Stroop effect, the standardized effect size is very large. The congruency effect is very clear when we plot the data from the two groups.



Differences between cohorts

The students were taught in two different years, 2013 and 2014. To examine whether the Stroop effect (the reaction time in the incongruent condition minus the reaction time in the congruent condition) differed between the two cohorts, we performed a Welch's t -test. The Stroop effect in the 2013 cohort ($M = 7.77$) did not differ from the Stroop effect in the 2014 cohort ($M = 8.13$), $t(120.37) = -0.42$, $p = 0.673$, Cohen's $d = -0.07$, 95% CI $[-0.42, 0.27]$.

Open Practices Statement

The raw data, the processed data, the analysis code, and the code used to generate this manuscript are all openly available in a GitHub repository at https://github.com/Lakens/stroop_effect_paper. No preregistration was filed for this classroom demonstration.

References

- Champely, S. (2020). *Pwr: Basic functions for power analysis*. <https://doi.org/10.32614/CRAN.package.pwr>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00863>
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109(2), 163–203. <https://doi.org/10.1037/0033-2909.109.2.163>