Supplemental Material for

Processing the Word Red and Intellectual Performance: Four Replications Attempts

Timo Gnambs [1,2,*], Carrie Kovacs [3], & Barbara Stiglbauer [1]

[1] Johannes Kepler University Linz

[2] Leibniz Institute for Educational Trajectories

[3] University of Applied Sciences Upper Austria

[*] Corresponding author

Detailed Methods for Experiment 1

Power Analysis

The sample size was determined based upon *a priori* power analyses in *pwr* version 1.2-2 (Champely, 2018) to identify an effect of Cohen's $d = 0.70$ (as in Experiment 2 in Lichtenfeld et al., 2009) for a one-tailed *t*-test at a significance level of 5% and a power of 80%. This resulted in a minimum sample size of 52.

Participants

Sixty-nine students (41 female and 28 male) from two upper secondary schools ("*Gymnasium*") in Austria with a median age of 17 years (*Min* = 16, *Max* = 19) were randomly assigned to a red color ($n = 30$) and a gray color ($n = 39$) condition. The study was conducted in small groups at the students' respective schools. All participants gave written informed consent and had good or very good German proficiency. None of the students guessed the purpose of the experiment after finishing the test. In exchange for their participation, all students received a candy bar. All data was collected in the year 2017.

Materials

The participants were informed that they were about to work on a short intelligence test. Then the verbal analogy subtest of the *Intelligence Structure Test 2000 R* (Liepmann, Beauducel, Brocke, & Amthauer, 2007) also used in Lichtenfeld and colleagues (2009) was administered in both groups. The test included 20 multiple-choice items forming a word pair and the first word of a second pair (e.g., "fast : slow = young : ?"). For each item, five response options were presented, one of which was correct (e.g., "quick, long, tall, tardy, old"). The number of correct answers was the dependent variable. Missing responses were scored as incorrect. Before the actual test, two example items were presented to explain the logic of the test. After the test,

socio-demographic information (sex, age) was assessed and students' proficiency in German ("How well do you understand German?") was measured on a four-point response scale (1 = *very badly*, 4 = *very well*). Finally, students indicated their assumptions about the purpose of the study in an open response field.

Experimental Manipulation

Following Experiment 2 in Lichtenfeld and colleagues (2009), the second example item before the analogy test included the experimental manipulation: "animal : hound = plant : ?" with five response options "branch, red/gray-alder, root, tree, organism". The manipulation was instigated by the correct solution being presented either as "red-alder" (red color condition) or "gray-alder" (gray color condition). Moreover, a description below the item (one sentence) explained that "red/gray-alder" was the correct solution.

Statistical Software

The statistical analyses were conducted in *R* version 3.6.1 (R Core Team, 2019) using the packages *car* version 3.0-4 (Fox & Weisberg, 2019), *sjstats* version 0.17.6 (Lüdecke, 2019), and *TOSTER* version 0.3.4 (Lakens, 2017).

Detailed Methods for Experiment 2

Power Analysis

Following the social science replication project (Camerer et al., 2018) we aimed at identifying a smaller effect that was only 75% of the original effects size. Thus, the sample size was determined based upon *a priori* power analyses to identify an effect of Cohen's $d = 0.50$ for a one-tailed *t*-test at a significance level of 5%, and a power of 80%. This resulted in a minimum sample size of 102.

Participants

From an original sample including 106 students from an Austrian university, two students were excluded because of a suspected color vision deficiency. The remaining $N = 104$ students (35 female, 68 male, 1 without information on sex) with a median age of 22 years (*Min* = 18, *Max* = 57) were randomly assigned to a red color ($n = 53$) and a gray color ($n = 51$) condition. All participants gave written informed consent and had good or very good German proficiency. None of the students guessed the purpose of the experiment after finishing the test. In exchange for their participation, all participants were eligible to enter a lottery for one of three Amazon vouchers worth 50 Euro. All data was collected in 2019.

Materials

The procedure was identical to that of the first experiment. Participants were informed that they were about to work on a short intelligence test and were then administered the analogy subtest of the *Intelligence Structure Test 2000 R* (Liepmann et al., 2007). After the test, socio-demographic information (sex, age) was assessed and students' proficiency in German ("How well do you understand German?") was measured on a four-point response scale (1 = *very badly*,

4 = *very well*). Moreover, participants indicated whether they had a color vision deficiency. Finally, students were asked about the assumed purpose of the study.

Experimental Manipulation

As in Experiment 1, the experimental manipulation was instigated by the second response option (i.e., "red/gray-alder") of the second example item explaining the test procedure.

Statistical Software

The statistical analyses were conducted in *R* version 3.6.1 (R Core Team, 2019) using the packages *car* version 3.0-4 (Fox & Weisberg, 2019), *sjstats* version 0.17.6 (Lüdecke, 2019), and *TOSTER* version 0.3.4 (Lakens, 2017).

Detailed Methods for Experiment 3

Power Analysis

The sample size was determined based upon *a priori* power analyses to identify an effect of Cohen's $d = 0.50$ for a one-tailed *t*-test at a significance level of 5% and a power of 80%. This resulted in a minimum sample size of 102.

Participants

From an original sample including 111 students attending a German university, four students were excluded because they failed to give the correct response to the item implementing the color manipulation (see below). The remaining $N = 107$ students (81 female and 26 male) with a median age of 21 years (*Min* = 19, *Max* = 44) gave informed consent and were randomly assigned to a red color ($n = 56$) and a green color ($n = 51$) condition. None of the students guessed the purpose of the experiment after finishing the test. The study was conducted in small groups. Students received course credits in exchange for their participation. All data was collected in 2012.

Materials

Participants were told that they were about to work on a short general knowledge test. First, socio-demographic information (sex, age) was collected. Then, a short version of the *General Knowledge Test – German* (GKT-D; Lynn, Wilberg, & Margraf-Stiksrud, 2004) was administered in small groups. The test included 37 items from different domains that had to be answered in open response fields. The number of correct responses after the experimental manipulation (i.e., based on 18 items of the GKT-D) was the dependent variable. Missing responses were scored as incorrect.

Experimental Manipulation

The experimental manipulation was implemented by changing the wording of item 19 of the GKT-D. In the red color condition, the item asked about the color of a ripe tomato (correct answer: “red”), whereas the control color condition inquired about the color of a ripe cucumber (correct answer: “green”). Four respondents failed to give the correct response to this item (i.e., either “yellow” or no response at all) and, thus, were excluded from the analyses.

Statistical Software

The statistical analyses were conducted in *R* version 3.6.1 (R Core Team, 2019) using the packages *car* version 3.0-4 (Fox & Weisberg, 2019), *sjstats* version 0.17.6 (Lüdecke, 2019), and *TOSTER* version 0.3.4 (Lakens, 2017).

Detailed Methods for Experiment 4

Power Analysis

Although Lichtenfeld and colleagues (2009) reported effect sizes between Cohen's *d* = 0.57 and 0.99 for their color manipulations, other research on behavioral priming has typically identified substantially smaller effects. For example, meta-analytic estimates for action and goal priming using incidentally presented words have been about *d* = 0.35 (Weingarten et al., 2016). In order to increase statistical power to detect even such small a small effect, the present study used a more conservative effect size estimate of *d* = 0.30 (i.e., less than half the effect reported in Lichtenfeld et al., 2009). Moreover, to guard against type II error, the power was set to 95%. An *a priori* power analysis estimated a required sample size of *N* = 1,180 to identify a Cohen's *d* of 0.30 using a significance level of 5% (two-tailed) and a power of 95% for an experimental setup with three color conditions (red, gray, and green) analyzed with a one-factorial analysis of variance and Tukey's (1949) honest significant difference post-hoc test. The study was conducted as an unproctored, web-based test. Because 10 to 20 percent of the respondents were expected to be screened out according to the exclusion criteria given below (cf. Chandler & Paolacci, 2017), the target sample size was set to *N* = 1,400.

Participants

Participants were members of an online access panel in Germany that received bonus points (that could be exchanged for monetary incentives) for completing the survey. A sample of *N* = 1,492 respondents finished the web-based test. Participants were excluded from the analyses based on six *a priori* specified criteria: (a) respondents with poor German proficiency (*n* = 35), (b) respondents who failed a seriousness check using a self-reported diligence item (*n* = 106), (c) respondents with a suspected color vision deficiency (*n* = 185), (d) participants taking an

unusually short amount of time[1] to complete the test (n = 73), (e) respondents guessing the hypotheses (i.e., mentioning the effect of any color with regard to cognitive abilities) at the end of the study (n = 0), and (f) respondents with missing values on all items of the knowledge test (n = 0). After applying these exclusion criteria, N = 1,149 participants (596 female, 552 male, and 1 without specified gender) with a median age of 38 years (*Min* = 16, *Max* = 85) remained, roughly half having been randomly assigned to either a red color (n = 563) or a gray color (n = 586) condition. All data was collected in 2019.

Materials

After giving informed consent, participants were told that they were about to work on a general knowledge test and receive feedback on their individual performance in reference to a representative norm sample. Then, the *BEFKI GC-K* (Schipolowski, Wilhelm, & Schroeders, 2013), a short instrument for the measurement of crystallized intelligence, was administered. The test includes 12 multiple-choice items with four response options each (with one option being correct). Each item was presented individually on the screen, without the possibility of returning to previous items and changing a response. The number of correct solutions was the dependent variable. Missing responses were scored as incorrect. Lichtenfeld and colleagues (2009) assumed that worries about the test performance would mediate the color effect on test performance. Therefore, after the knowledge test, worries with regard to the test performance were measured with three items (e.g., "I am not satisfied about my performance in the test.") based on Morris, Davis and Hutchings (1981) on seven-point response scales from 1 (*does not apply at all*) to 7 (*strongly applies*). Then, socio-demographic information and respondents' proficiency in

[1] All respondents falling in the lowest five percentile of the testing time, that is, those taking 6.5 minutes or less for the entire test, were excluded.

German (on a four-point response scale) were measured. After participants indicated the assumed purpose of the study as an open response, they completed one item of Ishigara's (1985) test for color blindness by identifying a colored number presented within a colored circle. Finally, a diligence item (see Aust, Diedenhofen, Ullrich, & Musch, 2013) asked respondents whether they had worked on the test in a serious manner (1= *not true at all*, 5 = *completely true*).

Experimental Manipulation

Although we planned to implement three color conditions (red, gray, and green), a programming error resulted in the green color condition not being administered. Therefore, the experiment included only two color conditions (red, gray), identical to the previous studies. The experimental manipulation was implemented in a similar way as in Experiment 2 of Lichtenfeld et al. (2009). Before the knowledge test, the following example item explaining the logic of the test was presented: "Which of these trees is a leaf tree?" with four response options "Nordmann-fir, red/gray-alder, Sargent-spruce, mountain-pine". The manipulation was again instigated by the correct solution being presented either as "red-alder" (red color condition) or "gray-alder" (gray color condition). In addition, a description below the item (one sentence) explained that "red/gray-alder" was the correct solution. To enforce the processing of the color word, respondents had to give the correct response to the manipulated example item before being able to proceed to the knowledge test.

Statistical Software

The statistical analyses were conducted in *R* version 3.6.1 (R Core Team, 2019) using the packages *car* version 3.0-4 (Fox & Weisberg, 2019), *sjstats* version 0.17.6 (Lüdecke, 2019), and *TOSTER* version 0.3.4 (Lakens, 2017).

Additional References

Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods, 45*, 527-535. https://doi.org/10.3758/s13428-012-0265-2

Champely, S. (2018). *pwr: Basic Functions for Power Analysis*. R package version 1.2-2. https://CRAN.R-project.org/package=pwr

Chandler, J. J., & Paolacci, G. (2017). Lie for a dime: When most prescreening responses are honest but most study participants are impostors. *Social Psychological and Personality Science, 8*, 500-508. https://doi.org/10.1177%2F1948550617698203

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression* (3rd edition). Thousand Oaks, CA: Sage.

Ishihara, S. (1985). *Ishihara's test for colour deficiency*. Göttingen, Germany: Hogrefe.

Lüdecke, D. (2019). *sjstats: Statistical Functions for Regression Models* (Version 0.17.6). https://doi.org/10.5281/zenodo.1284472

R Core Team (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org

Tukey, J. (1949). Comparing individual means in the analysis of variance. *Biometrics, 5*, 99-114. https://doi.org/10.2307/3001913%20