

Study Information

1. Title

Processing the word red and crystallized intelligence test performance

2. Authorship

Timo Gnambs

3. Research Questions

Perceiving color stimuli can influence psychological functioning including cognitive performance. In a series of experiments, Elliot and colleagues (Elliot, Maier, Moller, Friedman, & Meinhardt, 2007) showed that in an achievement context presenting a small stimulus colored red (e.g., a red square on a cover page of a test booklet) before an achievement test can significantly reduce test performance as compared to viewing other colors (e.g., green or gray). This effect has been demonstrated for verbal and numeric reasoning (Elliot et al., 2007; Maier, Elliot & Lichtenfeld, 2008), psychophysiological outcomes (Elliot, Payen, Brisswalter, Cury, & Thayer, 2011), and also various behavioral measures (e.g., walking speed; Meier, D’Agostino, Elliot, Maier, & Wilkowski, 2012) for various experimental manipulations (see Elliott & Maier, 2014, for a review). Theoretical and empirical evidence suggests that in an achievement context perceiving red color, due to its implicit association with caution and danger, influences achievement motivation; seeing red implicitly activates thoughts about failure and instigates avoidance motivation which, in turn, leads to poorer test performance (Elliot, Maier, Binser, Friedman, & Pekrun, 2009; Maier et al., 2008). Following the initial research on red color in achievement contexts, several independent follow-up studies corroborated these results (e.g., Bertrams, Baumeister, Englert, & Furley, 2015; Brooker & Franklin, 2016; Gnambs, Appel, & Batinic, 2010; Shi, Zhang, & Jiang, 2015; Zhang & Hang, 2014). Although some research failed to replicate the link between perceiving red and cognitive performance (e.g., Larsson & von Stumm, 2015; Smajic, Merritt, Banister & Blinebury, 2014), a meta-analysis (Pedley, 2016) suggested a robust red-performance link (pooled effect: Cohen’s $d = -0.28$).

Extending this line of research, Lichtenfeld and colleagues (Lichtenfeld, Maier, Elliot & Pekrun, 2009) showed that simply processing the *word* red is enough to yield effects that are comparable to actually seeing a red stimulus. In four experiments, the authors presented the word red before a reasoning test and observed significantly lower test scores when participants read the word red as compared to the word gray or green. The observed effects (Cohen’s d between 0.57 and 0.99) were rather impressive given the subtle color manipulations. For

example, in two experiments they manipulated a small copyright notice including seven words (font size: 10 points) in the bottom of the cover pages of the test booklet. In another experiment, an example item containing the word red or black was placed before a reasoning test. In all experiments, reading the word red consistently lead to poorer test performance as compared to reading another color word. If these findings can be substantiated, they might have important practical implications. Psychological and educational assessment typically do not consider color words when constructing and administering achievement tests. To prevent memory effects or cheating, it is not uncommon to administer parallel versions of a test to examinees (e.g., including different items or different item orders). However, if reading the word red influences subsequent test performance, different test versions might involuntarily bias, for example, certification programs or selection procedures; particularly, if color words are not matched across different test versions.

Considering the substantial effects previously triggered by the word red, the present study seeks to conceptually replicate and extend the study by Lichtenfeld et al. (2009). The aim of this preregistered experiment is to examine the effect of reading the word red as compared to two control colors (gray, green) on a subsequent achievement test. The study will extend Lichtenfeld et al. (2009) in three ways: First, the field of color psychology is dominated by research on student samples. Therefore, the present study seeks to examine the respective effect in an adult sample from the German population. Second, the study will adopt an appropriate sample size that is based on *a priori* power considerations to detect the expected effects. Many previous studies used rather small samples that are unlikely to identify small effects; for example, the sample sizes in Lichtenfeld et al. (2009) fell between 20 and 49 across the four experiments. Third, instead of fluid intelligence the present study will focus on a measure of crystallized intelligence. Thus, the experiment will evaluate whether the suggested color effect generalizes to test settings that are typical encountered in educational contexts. Finally, in line with Lichtenfeld et al. (2009) the study will try to replicate the implied mediation effect of avoidance motivation; thus, it is expected the worries about test performance will mediate the red color effect on test performance.

4. Hypotheses

Hypothesis 1: Knowledge test scores will be lower for respondents reading the word red before the achievement test as compared to respondents reading the word gray or green.

Hypothesis 2: Worry scores will be larger for respondents reading the word red before the achievement test as compared to respondents reading the word gray or green.

Hypothesis 3: Worry will mediate the color effect on knowledge test performance. Respondents reading the word red will exhibit higher scores on the worry scale as compared to respondents reading the word gray or green. In turn, higher scores on the worry scale will be associated with lower scores on the knowledge test.

Sampling Plan

5. Existing data

Not data has been collected prior to the generation of this document.

6. Explanation of existing data

Not applicable.

7. Data collection procedures.

Data will be collected using an unproctored, web-based survey. The field time will be up to two weeks. The length of the survey is expected to be about 10 minutes.

8. Sample size

The target sample size is 1,400. A combined quota sample according to sex and age will be drawn from an online access panel (see Table 1).

Table 1. Sample quotas

		Sex		Total
		Women	Men	
Age	16 to 25 years	17%	17%	34%
	26 to 45 years	17%	16%	33%
	46 years and older	17%	16%	33%
Total		51%	49%	100%

9. Sample size rationale

Lichtenfeld and colleagues (2009) conducted four experiments that examined the effect of processing the word red on verbal and numeric reasoning abilities. These studies resulted in effect sizes (Cohen’s d) of 0.57, 0.73, 0.64, and 0.99. However, these effects are likely to be an overestimation of the true effect. A recent meta-analysis (Pedley, 2016) on exposure to red color and cognitive performance identified pronounced experimenter effect. Research groups involving the original authors of the color red effect (Elliot et al., 2007) typically reported substantially larger effects ($d = 1.08$) as compared to independent replication attempts ($d = 0.11$). Moreover, other research on behavioral priming

effects (e.g., on the impact of presenting words connected to an action or a goal) also derived considerably smaller effects (meta-analytic estimate: $d = 0.35$; Weingarten et al., 2016). Therefore, the present study adopts a rather conservative target effect size of Cohen’s $d = 0.30$ that is less than half the effect reported in Lichtenfeld et al. (2009).

Power analyses (see attached simulation syntax) were conducted to identify the sample size requirements to identify a $d = 0.30$ with $\alpha = 0.05$. To guard against a type II error, the power was set to 0.95. Hypotheses 1 and 2 focused on main effects of the color condition (see section 18). To identify the ANOVA main effect a sample size of 780 is required (see Figure 1), whereas for Tukey’s post-hoc test a sample of 1,080 is needed (see Figure 2). Because it is unclear whether the three experimental conditions will result in equal sample sizes (e.g., on account of different dropout rates), it was also evaluated whether an unequal allocation of respondents to the three color conditions (33% / 33% / 33% or 25% / 37% / 37%) might affect sample size requirements. For unequal sample sizes slightly larger samples are needed, $N = 920$ for the ANOVA main effect and $N = 1,180$ for Tukey’s HSD (see Figures 1 and 2).

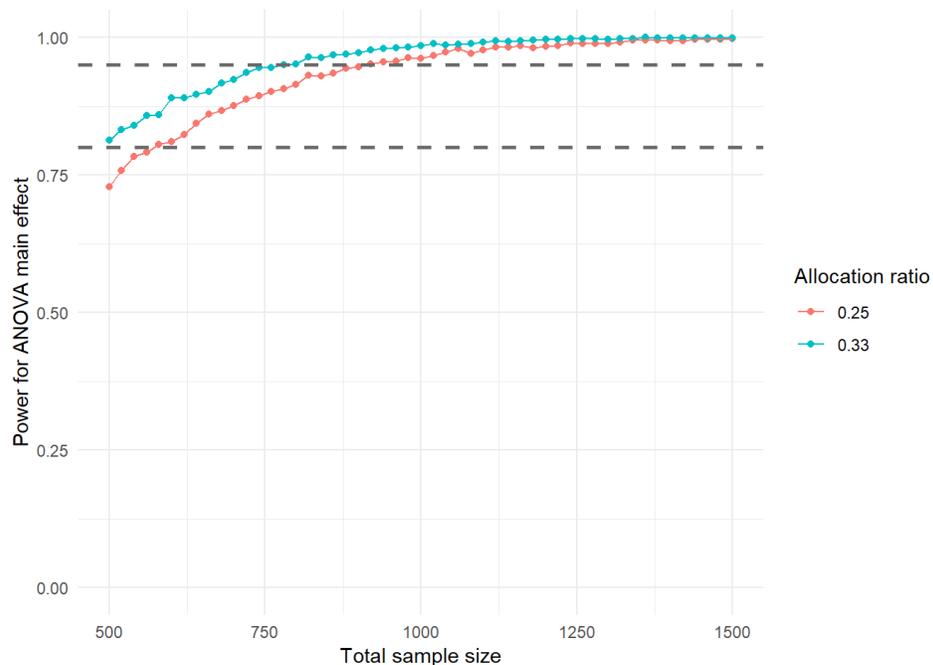


Figure 1. Estimated power of ANOVA main effect for $d = 0.30$ (group 1 versus group 2 or 3), $\alpha = 0.05$, and different allocation ratios (25% or 33% of respondents in group 1). Dashed lines represent a power of 80% and 95%.

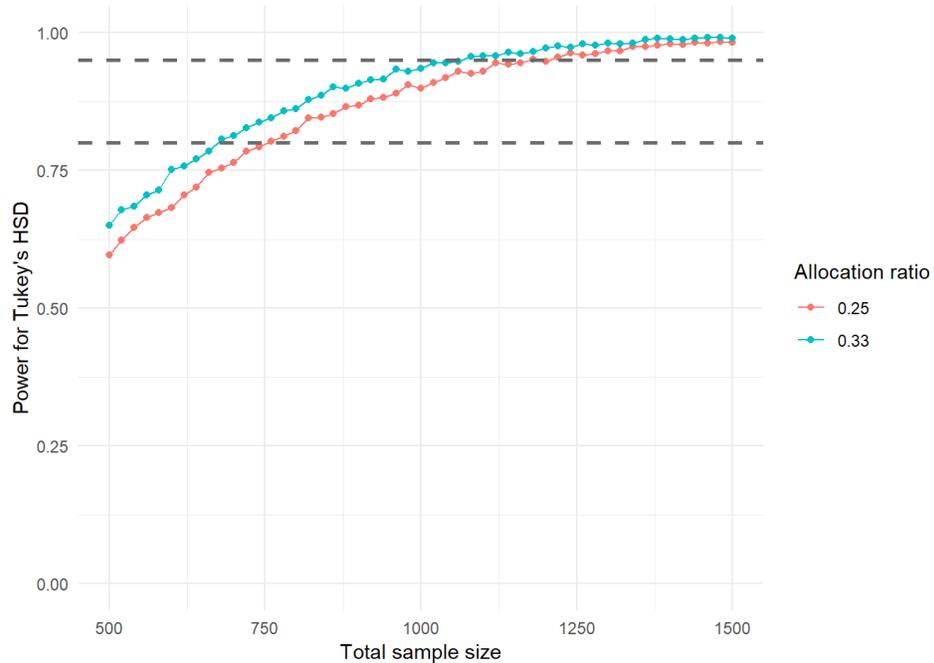


Figure 2. Estimated power of Tukey’s HSD for $d = 0.30$, $\alpha = 0.05$, and different allocation ratios (25% or 33% of respondents in group 1). Dashed lines represent a power of 80% and 95%.

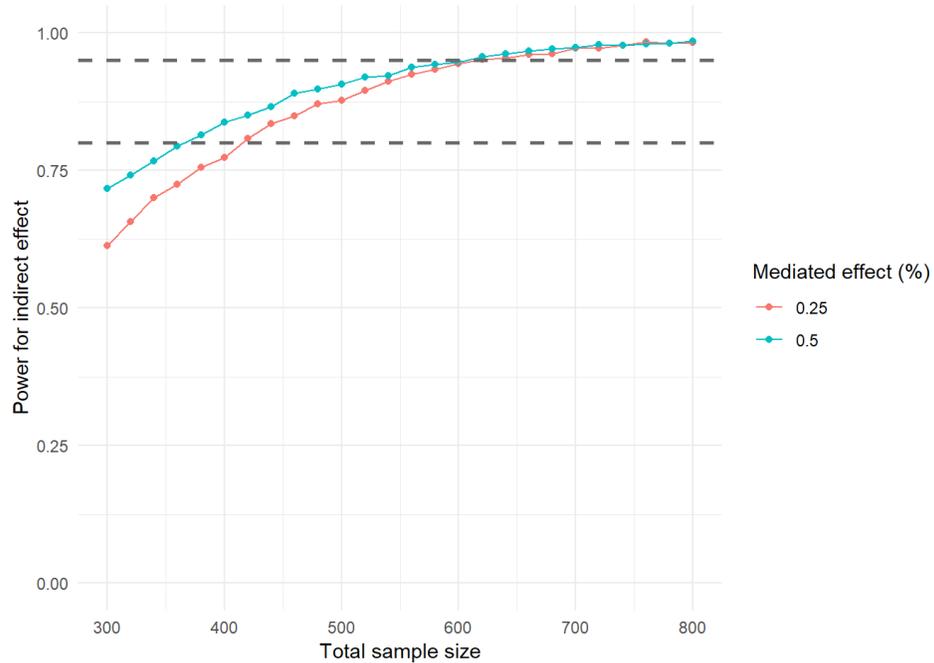


Figure 3. Estimated power of an indirect effect for at total effect of $d = 0.30$, $\alpha = 0.05$, and different indirect effects (25% or 50% of the total effect). Dashed lines represent a power of 80% and 95%.

Hypothesis 3 addresses a mediation effect. Two scenarios were considered, that is, the indirect effect accounts for (a) 25% or (b) 50% of the total effect. The required sample size to identify these indirect effects is 620 in both cases (see Figure 3).

Table 2 summarizes the required sample sizes for the different analyses under various conditions. These results suggest a minimum sample size of $N = 1,180$.

Table 2. Required sample sizes for different scenarios

Condition	Required sample size
ANOVA main effect for equal groups	780
ANOVA main effect for unequal groups	920
Tukey’s HSD for equal groups	1,080
Tukey’s HSD for unequal groups	1,180
Indirect effect of 25% of the total effect	620
Indirect effect of 50% of the total effect	620

Note. Estimated sample sizes for $d = 0.30$ and $\alpha = \beta = 0.05$.

About 10-20% of the respondents are expected to be screened due to problematic response behavior or other exclusion criteria (see section 22). Therefore, the required sample size increases to $N = 1,400$ for the present study.

10. Stopping rule

Sampling will stop as soon as the designated sample size is reached.

Variables

11. Manipulated variables

Participants will be randomly assigned to one of three experimental conditions and will be presented either with the word red, the word gray, or the word green. The experimental manipulation will be implemented by presenting an example item before the knowledge test. The multiple-choice example item will be identical in the three conditions except for the color word “rot” (“red”), “grau” (“gray”), or “grün” (“green”). The following item will implement the experimental manipulation:

Welcher dieser Bäume ist ein Laubbaum?	(Which of these trees is a leaf tree?)
○ Nordmann-Tanne	(Nordmann-Fir)
x Rot/Grau/Grün-Erle	(Red/Gray/Green-Alder)
○ Sargent-Fichte	(Sargent-Spruce)
○ Berg-Kiefer	(Mountain-Pine)

The second response option will include either the word “red”, the word “gray”, or the word “green” depending on the experimental condition. The experimental manipulation is similar to the one used in experiment 2 of Lichtenfeld et al. (2009). The authors used an example item for a verbal reasoning test and also included “red/gray alder” as one of five response options to implement the color manipulation.

The example item will be accompanied by the following description: “Auf den nächsten Seiten werden Sie 12 Fragen finden. Zu jeder Frage gibt es vier Antwortmöglichkeiten. Es ist immer nur eine Antwortmöglichkeit richtig. Bitte wählen Sie bei jeder Frage die Ihrer Meinung nach richtige Antwort aus. In diesem Beispiel ist die zweite Antwortmöglichkeit richtig, da die Erle ein Laubbaum ist und alle anderen Bäume nicht. Bitte markieren Sie nun die zweite Antwortmöglichkeit.” [On the following pages you will be presented with 12 questions. For each question there are four response options. Only one response option is correct. Please select the response option for each question that, in your opinion, indicates the correct response. In this example, the second response option is correct because the alder is a leaf tree, whereas the other trees are not. Please select the second response option.]

For respondents selecting a distractor instead of the correct response option, a warning message will be presented to indicate the mistake. Respondents can proceed to the actual knowledge test only after selecting the correct response option including the experimental manipulation.

12. Measured variables

General knowledge will be assessed using the BEFKI GC-K (Schipolowski et al., 2014), a short scale for the measurement of crystallized intelligence. The test includes 12 multiple-choice items with four response options (with one option being correct). The BEFKI GC-K has been developed as a short screening instrument to be administered in large-scale studies with limited testing time. In a representative sample of German adults, the test demonstrated a unidimensional factor structure, satisfactory reliability ($\alpha = .81$), and validity comparable to a longer version of the instrument (Willhelm, Schroeders, & Schipolowski, 2014). In the present study, each item will be presented individually on the screen. Respondents will not have the option to return to previous items or change their responses.

Worry will be measured with three items based upon Morris, Davis, and Hutchings (1981) on seven-point response scales from 1 = “trifft überhaupt nicht zu” (“does not apply at all”) to 7 = “trifft stark zu” (“strongly applies”):

1. In diesem Test habe ich wahrscheinlich nicht so gut abgeschnitten wie ich eigentlich könnte (In this test I may not have done as well as I could have).
2. Andere Personen wären wahrscheinlich von meiner Leistung in diesem Test enttäuscht (Others would be disappointed in my test performance.)
3. Ich bin nicht zufrieden was meine Leistung in diesem Test betrifft (I am not satisfied about my performance on this test)

The following socio-demographic information will be collected:

- Sex will be measured using a single-choice item (“Ihr Geschlecht”) with three categories: male (“männlich”), female (“weiblich”), other (“anderes”)
- Age will be measured with an open response (“Wie alt sind Sie?”).
- Education will be measured with a single-choice item (“Welchen höchsten allgemeinen Schulabschluss haben Sie?”) using three categories: secondary general school (“Hauptschule”), intermediate secondary school (“Realschule / mittlere Reife”), general higher education (“Fachhochschulreife / Abitur”)
- Proficiency in German will be measured with a single item (“Wie gut verstehen Sie Deutsch”) using a four-point rating scale using the response options “sehr schlecht” (“very poorly”), “schlecht” (“poorly”), “gut” (“well”) to “sehr gut” (“very well”).

Next, the research question hypothesized by the participants will be assessed with an open-ended question as “Was untersucht diese Studie Ihrer Meinung nach? Bitte fassen Sie Ihre Vermutung kurz mit wenigen Stichworten zusammen. Wenn Sie keine konkrete Vermutung haben, lassen Sie dieses Feld frei” [What do you think, which question does this study examine? Please shortly summarize your assumptions with few keywords. If you do not have any specific assumption, leave the field empty].

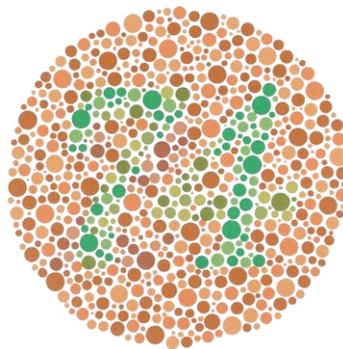


Figure 4. Item of Ishihara’s (1985) test for color deficiency. Source: Wikipedia.

On the following page one item from Ishihara’s (1985) test of color blindness will be presented. Respondents are required to indicate the number (74) they see as an open-ended response.

Finally, a diligence item will be presented. The wording is “Ich habe ernsthaft und nach bestem Wissen und Gewissen an dieser Studie teilgenommen. Meine Antworten können sinnvoll für die Forschung verwendet werden“ [I was serious and answered the survey questions to the best of my knowledge. It makes sense to use my answers for your research]. Answers are required on a five-point scale with verbal options “trifft gar nicht zu” (not true at all), “trifft eher nicht zu” (trending not true), “teils-teils” (undecided), “trifft eher zu” (trending true) to “trifft voll und ganz zu” (completely true).

13. Indices

For the general knowledge test correct responses will be scored as 1 and incorrect or missing responses will be scored as 0. Scores for each respondent will be derived by calculating the sum of the 12 test items.

For each respondent the mean of the three items of the worry scale will be calculated.

Design Plan

14. Study type

Experiment

15. Blinding

The experimenter will not be present because this will be an online experiment. The presentation of the stimuli, the administration of the items, and the allocation of the participants to the experimental conditions are completely computerized without any human involvement during the data collection.

Participants will not know about the purpose of the study (examining color effects). As knowledge of the purpose might elicit demand effects, participants will be asked about the supposed purpose at the end of the survey (see exclusion criteria).

16. Study design

This is a cross-sectional web-based experiment following a one-factorial design with three color conditions (red versus gray versus green).

17. Randomization

Participants will be randomly assigned to the color conditions via the online survey tool.

Analysis Plan

18. Statistical models

Hypothesis 1 will be tested using a one-factorial (color: red versus gray versus green) analysis of variance with knowledge test scores as dependent variable. The hypothesis to be tested is represented by the main effect of color. Given a significant main effect, pairwise comparisons between the three color conditions will be conducted using Tukey’s (1949) honest significant difference (HSD) test. Hypothesis 1 is supported if (a) the ANOVA main effect is significant, (b) Tukey’s (1949) HSD test indicates significant differences between the red and gray conditions as well as between the red and green conditions, and (c) mean test scores are smaller in the red condition as compared to the gray and green condition. The analyses will be conducted in *R* (R Core Team, 2018) using the *lm* and *TukeyHSD* functions from the *stats* package.

Hypothesis 2 will be tested in the same way as hypothesis 1 except for the use of the worry scores as dependent variable. Hypothesis 2 is supported if (a) the ANOVA main effect is significant, (b) Tukey’s (1949) HSD test indicates significant differences between the red and gray conditions as well as between the red and green conditions, and (c) mean test scores are larger in the red condition as compared to the gray and green condition.

Hypothesis 3 will be tested using a path model in *lavaan* (Rosseel, 2012). A mediation model will be specified with knowledge test scores as outcome, color condition (coded 0 for gray/green and 1 for red) as independent variable, and worry scores as mediator. The hypothesis to be tested is represented by the indirect effect.

19. Transformations

No transformations are planned.

20. Follow-up analyses

Because some research suggested sex differences in color effects (Gnamb et al., 2010; Ioan et al., 2007), all analyses will be repeated including the main effect of sex and the interaction between sex and the color condition (see Lichtenfeld et al., 2009, for similar analyses).

Although no age effects are expected, all analyses will be repeated controlling for the respondents age; that is, analyses of covariance will be used to test hypotheses 1 and 2, while the path model for hypothesis 3 will be extended by regressing all three variables on age. Moreover, potential interaction effects will be explored by regressing the knowledge test scores on the color condition (coded 0 for gray/green and 1 for red), the age of the respondents (centered), and the respective interaction effect. Age effects are supported if the interaction effect is significant.

21. Inference criteria

We will use the standard null hypothesis testing $p < .05$ criteria for determining if the effects are significantly different from those expected under the null hypothesis (two-tailed). The size of the effects will be calculated as standardized mean differences (Cohen's d) or eta squared.

22. Data exclusion

Participants will be excluded from the analyses according to the following criteria:

1. Respondents who check the self-reported diligence item with an answer other than “trifft eher zu” (trending true) or “trifft voll und ganz zu” (completely true) will be excluded.
2. Participants taking an unusually short amount of time to complete the survey will be excluded.
3. Respondents failing to give the correct response to the item testing for color deficiency will be excluded.
4. Respondents indicating in an open-ended question at the end of the questionnaire that they were able to guess the hypotheses (i.e., mentioning the effect of any color with regard to cognitive abilities) will be excluded.
5. Respondents with missing values on all items of the knowledge test or worry scale will be excluded.
6. Respondents who check the German proficiency item with an answer other than “gut” (“well”) or “sehr gut” (“very well”) will be excluded.

Potential outliers will be kept in the analyses.

23. Missing data

Respondents with missing values will be excluded according to the conditions given in section 22. Missing values on items of the knowledge test will be scored as incorrect (Schipolowski et al., 2014). Items with missing values on the worry scale will be excluded from the scoring.

24. References

- Bertrams, A., Baumeister, R. F., Englert, C., & Furley, P. (2015). Ego depletion in color priming research: self-control strength moderates the detrimental effect of red on cognitive test performance. *Personality and Social Psychology Bulletin*, *41*, 311-322. doi:10.1177/0146167214564968
- Brooker, A., & Franklin, A. (2016). The effect of colour on children's cognitive performance. *British Journal of Educational Psychology*, *86*, 241-255. doi:10.1111/bjep.12101
- Elliot, A. J., & Maier, M. A. (2014). Color psychology: effects of perceiving color on psychological functioning in humans. *Annual Review of Psychology*, *65*, 95-120. doi:10.1146/annurev-psych-010213-115035
- Elliot, A. J., Maier, M. A., Binser, M. J., Friedman, R., & Pekrun, R. (2009). The effect of red on avoidance behavior. *Personality and Social Psychology Bulletin*, *35*, 365-375. doi:10.1177/0146167208328330
- Elliot, A. J., Maier, M. A., Moller, A. C., Friedman, R., & Meinhardt, J. (2007). Color and psychological functioning: The effect of red on performance attainment. *Journal of Experimental Psychology: General*, *136*, 154-168. doi:10.1037/0096-3445.136.1.154
- Elliot, A. J., Payen, V., Brisswalter, J., Cury, F., & Thayer, J. F. (2011). A subtle threat cue, heart rate variability, and cognitive performance. *Psychophysiology*, *48*, 1340-1345. doi:10.1111/j.1469-8986.2011.01216.x
- Gnams, T., Appel, M., & Batinic, B. (2010). Color red in web-based knowledge testing. *Computers in Human Behavior*, *26*, 1625-1631. doi:10.1016/j.chb.2010.06.010
- Ioan, S., Sandulache, M., Avramescu, S., Ilie, A., Neacsu, A., Zagrean, L., & Moldovan, M. (2007). Red is a distractor for men in competition. *Evolution and Human Behavior*, *28*, 285-293. doi:10.1016/j.evolhumbehav.2007.03.001
- Ishihara, S. (1985). *Ishihara's test for colour deficiency*. Göttingen, Germany: Hogrefe.
- Larsson, E. E. C., & von Stumm, S. (2015). Seeing red? The effect of colour on intelligence test performance. *Intelligence*, *48*, 133-136. doi:10.1016/j.intell.2014.11.007
- Lichtenfeld, S., Maier, M. A., Elliot, A. J., & Pekrun, R. (2009). The semantic red effect: Processing the word red undermines intellectual performance. *Journal of Experimental Social Psychology*, *45*, 1273-1276. doi:10.1016/j.jesp.2009.06.003
- Maier, M. A., Elliot, A. J., & Lichtenfeld, S. (2008). Mediation of the negative effect of red on intellectual performance. *Personality & Social Psychology Bulletin*, *34*, 1530-1540. doi:10.1177/0146167208323104
- Meier, B. P., D'Agostino, P. R., Elliot, A. J., Maier, M. A., & Wilkowski, B. M. (2012). Color in context: Psychological context moderates the influence of

- red on approach-and avoidance-motivated behavior. *PloS One*, 7(7), e40333. doi:10.1371/journal.pone.0040333
- Morris, L. W., Davis, M. A., & Hutchings, C. H. (1981). Cognitive and emotional components of anxiety: Literature review and a revised worry-emotionality scale. *Journal of Educational Psychology*, 73, 541-555. doi:10.1037//0022-0663.73.4.541
- Pedley, A. (2016). *The influence of red stimuli on cognitive performance in achievement context settings*. Unpublished doctoral thesis (University of Surrey).
- R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48, 1-36. doi:10.18637/jss.v048.i02
- Schipolowski, S., Wilhelm, O., Schroeders, U., Kovaleva, A., Kemper, C. J., & Rammstedt, B. (2014). *Eine kurze Skala zur Messung kristalliner Intelligenz* [A short scale for the measurement of crystallized intelligence]. GESIS Working Paper 24. Mannheim, Germany: GESIS. URL: <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-377864>
- Smajic, A., Merritt, S., Banister, C., & Blinebry, A. (2013). The red effect, anxiety, and exam performance: A multistudy examination. *Teaching of Psychology*, 41, 37-43. doi:10.1177/0098628313514176
- Shi, J., Zhang, C., & Jiang, F. (2015). Does red undermine individuals' intellectual performance? A test in China. *International Journal of Psychology*, 50, 81-84. doi:10.1002/ijop.12076
- Tukey, J. (1949). Comparing individual means in the analysis of variance. *Biometrics*, 5, 99-114. doi:10.2307/3001913
- Weingarten, E., Chen, Q., McAdams, M., Yi, J., Hepler, J., & Albarracín, D. (2016). On priming action: conclusions from a meta-analysis of the behavioral effects of incidentally-presented words. *Current Opinion in Psychology*, 12, 53-57. doi:10.1016/j.copsyc.2016.04.015
- Wikipedia (2018). *Ishihara 9*. URL: https://commons.wikimedia.org/wiki/File:Ishihara_9.png
- Wilhelm, O., Schroeders, U., & Schipolowski, S. (2014). *Berliner Test zur Erfassung fluider und kristalliner Intelligenz für die 8. bis 10. Jahrgangsstufe (BEFKI 8-10)*. Göttingen, Germany: Hogrefe.
- Zhang, T., & Han, B. (2014). Experience reverses the red effect among Chinese stockbrokers. *PLoS ONE*, 9(2). doi:10.1371/journal.pone.0089193