

ORIGINAL RESEARCH REPORT

No Effect of Red on Personality Trait Self-Ratings: Testing for Effects of Font Color

Daniel Wolf and Astrid Schütz

Unlike most other mammals, humans are trichromats and have the ability to perceive the color red. An explanation for the evolution of humans' trichromatic color vision is that it offers humans the advantage to detect ripe fruit. Apart from this explanation, psychological theories have proposed that color, especially the color red, conveys information that affects psychological functioning, but results have been mixed. Whereas studies have extensively tested effects of red on performance measures, it is unclear whether this effect generalizes to self-ratings, one of the most frequently used methods in psychological research. In line with theory and empirical evidence, we argue that exposure to red can lead to distorted responses in self-ratings on the basis of the same underlying mechanism that affects results on performance measures. We varied the font color (hue values) of self-ratings in two online studies. In a first exploratory study, we found an effect of font color on personality trait self-ratings ($N = 145$). We attempted to rigorously replicate this finding in a larger sample ($N = 1,007$) but did not detect any effect. The findings underline the importance of rigorous research on effects of color on psychological functioning and call into question the proposition that red has ubiquitous effects.

Keywords: color; red; self-rating; psychological functioning

Human experience is colorful. Unlike the majority of mammals (i.e., dichromats), humans, apes, and Old World monkeys are trichromats, which means they can perceive more than 2.3 million colors with a unique ability to perceive the colors from the red spectrum (Jacobs, 2008; Pointer & Attridge, 1998). Whereas red color vision might be grounded in the advantage to detect ripe fruit (Osorio, Smith, Vorobyev, & Buchanan-Smith, 2004), it might also affect psychological functioning as proposed by the color-in-context theory (Elliot, 2015; Elliot & Maier, 2014). Research has suggested that perceiving red in evaluative contexts such as IQ testing has a negative impact on results (Elliot, Maier, Moller, Friedman, & Meinhardt, 2007). In this article, we argue that red may affect self-ratings on the basis of the same underlying mechanism that affects performance measure results. Thus, we conducted an empirical test of this hypothesis.

According to the color-in-context theory, (a) color can convey information that is relevant for psychological functioning, and (b) exposure to color can affect psychological functioning. (c) This occurs by activating color associations that in turn evoke unconscious affective, behavioral, and cognitive reactions. (d) These reactions are learned or based on biological predispositions. (e) This

effect can be reciprocal such that color perception can affect psychological functioning, and psychological functioning can affect color perception. (f) This effect is also moderated by the psychological context. For example, the effect of red in an achievement context may differ from that of red in a romantic context (Elliot & Maier, 2012; Maier, Hill, Elliot, & Barton, 2015).

Based on the color-in-context theory, various studies on the effects of red in achievement contexts have been published (see Elliot, 2015; Elliot & Maier, 2014; Maier et al., 2015). In this line of research, achievement contexts are defined as situations in which competence evaluations take place, and positive or negative outcomes are possible (Elliot, 1999; Elliot et al., 2007; Maier et al., 2015). The initial study suggested that perceiving red prior to and while working on tests of cognitive abilities, such as anagrams or IQ tests, reduces performance (Elliot et al., 2007). Effects have been observed not only in adults but also in children (Brooker & Franklin, 2016) and in different cultures (Shi, Zhang, & Jiang, 2015). Likewise, performance on knowledge tests was reduced when people were exposed to red while being tested, but the effect occurred only in males (Gnamb, Appel, & Batinic, 2010). Furthermore, seeing red prior to taking a pinchgrip or a handgrip task also impaired performance (Payen et al., 2011), whereas being confronted with red while performing isometric maximal voluntary contractions of the thigh enhanced performance on the task (Elliot &

Aarts, 2011). Interestingly, research has suggested that the effect of red on IQ test performance is also present when a person is simply exposed to the word red instead of perceiving the color red (Lichtenfeld, Maier, Elliot, & Pekrun, 2009).

The negative impact of perceiving red in achievement contexts has been explained by color associations: Red is typically associated with failure and negativity. It has been suggested that this association is learned (teachers mark mistakes in red, red traffic lights and traffic signs require the inhibition of action) but may partly be biologically based (blood is red, toxic animals are partly or completely red) and that consequently red evokes fear of failure in achievement contexts (Elliot et al., 2007). This last idea has been supported by research that showed that failure words were categorized more quickly when they were presented in red (Moller, Elliot, & Maier, 2009), that the use of red pens increased the rate at which words related to error and failure were completed, the rate at which errors were corrected, as well as the rate at which poor grades were given (Rutchick, Slepian, & Ferris, 2010). Further, the rates at which negative words were recalled were increased when the words were presented in red font (Kuhbandner & Pekrun, 2013). Supporting the assumption of a learned association, the negative effect of red on performance was reversed in Chinese stockbrokers – a finding that can be explained by the fact that increases in stock prices are presented in red in the Chinese stock market (Zhang & Han, 2014).

Meta-analytic evidence has suggested that performance on IQ tests is heavily influenced by test motivation (Duckworth, Quinn, Lynam, Loeber, & Stouthamer-Loeber, 2011), and test motivation in turn is affected by personality traits (Borghans, Meijers, & Ter Weel, 2008; Freund & Holling, 2011). Thus, unincentivized performance is basically driven by an individual's motive to present him or herself in a positive manner – of course as long as the relevant ability is present.

The motive to present oneself in a positive manner also affects the results of self-ratings and is often referred to as socially desirable responding (Paulhus, 2002). Meta-analytic evidence has shown that there is a general positivity bias in self-ratings (Viswesvaran & Ones, 1999). Additionally, research has indicated that self-ratings can be regarded as having positive or negative outcomes (Bäckström, Björklund, & Larsson, 2009) and in this respect resemble achievement contexts: Almost everybody wants to be more emotionally stable, more extraverted, more open to experience, more agreeable, and more conscientious (Hudson & Roberts, 2014). In a similar vein, research regarding a general factor of personality, which can be seen as a sort of meta-trait representing high emotional stability, high extraversion, high openness to experience, high agreeableness, and high conscientiousness, has suggested that there is a general evaluative factor in self-ratings of personality (Bäckström, 2007; Bäckström, Björklund, & Larsson, 2009; Schermer & MacDougall, 2013; Schermer & Vernon, 2010). This general factor of personality is also related to affect and depression (Rushton & Erdle, 2010). Thus, these relations illustrate how vulnerable self-ratings are to negative affect

and parallel the effects of fear of failure on performance measures.

To sum up, past research has suggested that red has negative effects on performance. This effect has been attributed to the fact that red evokes fear of failure in achievement contexts. We argue that, just like performance measures, self-ratings imply an evaluative context because in self-ratings, there are likewise potential positive or negative outcomes. In some studies on the effects of red on performance, self-ratings of states such as anxiety and worry were used as manipulation checks, and sometimes there was no effect on self-ratings (Elliot et al., 2007), but sometimes such an effect occurred (e.g., there was an increase in worry) (Lichtenfeld et al., 2009; Zhang & Han, 2014). The results have been interpreted as indicators of an unconscious effect of red on performance (Maier et al., 2015). Such concerns could likewise affect personality trait self-ratings.

Self-ratings are among the most often applied measures in psychological research and can be seen as indispensable in online studies. However, conducting research online increases the chances that participants will be exposed to color much more than in classical paper-and-pencil studies. Color may be used intentionally or unintentionally in web designs and may lead to bias on self-ratings if there are unconscious effects of color. Still, a search on the effects of red on self-ratings did not provide anything but the aforementioned results on increases in worry (Lichtenfeld et al., 2009; Zhang & Han, 2014).

On the basis of the reasoning presented above, we aimed to test the effect of red on self-ratings. We hypothesized that the same underlying mechanisms that alter performance results would lead to distortions in self-ratings as well. We expected that fear of failure would lead to exaggerations of one's positive features and thus higher scores on traits with a positive valence (e.g., conscientiousness) and lower scores on traits with a negative valence (e.g., neuroticism).

Reviews of effects of color on psychological functioning have called for rigor in empirical work to clarify the effects of color on psychological functioning, which have been masked by mixed results in pre-21st-century research when researchers have failed to consider the multidimensionality of color (Elliot, 2015; Elliot & Maier, 2014). In line with this call, we aimed to address the weaknesses that were criticized in earlier research on color, and we controlled for the multidimensionality of color at the device level, participants' color deficiencies, and the duration of color exposure.

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the studies (Simmons, Nelson, & Simonsohn, 2012). Supplemental information for this article is available online at the Open Science Framework (<https://osf.io/cmqfd/>). All stimuli, presentation materials, participant data, and analysis scripts are also available online at the Open Science Framework (<https://osf.io/x493s/>).

Study 1

Study 1 was conducted in an exploratory manner. We administered several state and trait self-ratings in three different font colors: the chromatic colors red and blue

as well as the achromatic color black. We used both state and trait self-ratings because both types of measures can potentially be influenced by red, and an influence on trait measures could partly be explained by changes in state (Steyer, Mayer, Geiser, & Cole, 2015; Steyer, Schmitt, & Eid, 1999). We decided to use blue as a conservative chromatic contrast because it has a universally different appearance from red (Abramov & Gordon, 1994; Bornstein, 1973) but can be matched with red on chroma and lightness. Moreover, blue is a relatively common color as we often write in blue ink, and blue seems to be the most preferred color (Crozier, 1999). Black was used as an additional achromatic contrast to red. We manipulated the font color because we assumed that font color would be perceived on the one hand but would not be obtrusive on the other hand (e.g., like background color would be). We tested for a possible effect of color on self-ratings of state and trait concepts separately.

Method

Participants

Data were collected online. Between November 2015 and March 2016, we registered 216 first-page views, of which 188 individuals completed the study. On a dichotomously scored variable “I have completed all questionnaires with care” versus “I have completed at least one questionnaire without care (e.g., because I wanted to finish up fast or just clicked through),” 147 participants indicated that they had completed the study with care. Of these, two participants self-reported a red-green color vision impairment (one in the blue condition, one in the black condition) and were excluded. The remaining $N = 145$ participants (77% women) had a mean age of 28.6 years ($SD = 10.8$). Not a single participant correctly anticipated the purpose of the study.

Participants were recruited mostly among students from a University in southern Germany and additionally from various social networks (e.g., Facebook groups). Students received course credit for participation upon request (course credit was requested 50 times). To incentivize nonstudent participants, we implemented the additional opportunity to obtain individual feedback on one's personality profile (feedback was requested 96 times). All participants provided informed consent. The study was approved by the ethics committee of the University of Bamberg (dossier number 2019-07/21).

Measures

We decided to use short versions of questionnaires whenever available and accepted a tradeoff in reliability to minimize the load on participants. Affective state was assessed in direct and indirect ways. For direct assessment, we applied the German versions of (a) the International Positive and Negative Affect Schedule Short Form (I-PANAS-SF) measuring positive (PA) and negative affect (NA) with five affective markers each rated on a 5-point scale ranging from 1 (*not at all*) to 5 (*strongly*) (Thompson, 2007), whereas translations for the markers were taken from the Positive and Negative Affect Schedule (Krohne, Egloff, Kohlmann, & Tausch, 1996); and (b) the Activation-Deactivation Adjective Check List (AD ACL) measuring

energy and tension with 10 affective markers each rated on a 4-point scale ranging from 1 (*not at all*) to 4 (*strongly*) (Imhof, 1998). For indirect assessment, we applied the Implicit Positive and Negative Affect Test (IPANAT) measuring implicit positive (iPA) and implicit negative affect (iNA) (Quirin & Bode, 2014). The IPANAT measures implicit affect by requesting participants to rate the extent to which made-up words express different kinds of affective markers. Each of six made-up words has to be rated on six identical affective markers on a 4-point scale ranging from 1 (*does not fit at all*) to 4 (*fits very well*).

For trait assessment, the German versions of different questionnaires were administered: The Big Five Inventory (BFI) measuring neuroticism (N) with seven items, extraversion (E) with eight items, openness (O) with 10 items, agreeableness (A) with eight items, and conscientiousness (C) with nine items (Lang, Lüdtke, & Asendorpf, 2001); the 20-item version of the Balanced Inventory of Desirable Responding (BIDR-20) measuring self-deceptive enhancement (SDE) and impression management (IM) with 10 items each (Musch, Brockhaus, & Bröder, 2002); the Satisfaction with Life Scale (SWLS) measuring satisfaction with life (SWL) with five items (Glaesmer, Grande, Braehler, & Roth, 2011); the revised Rosenberg Self-Esteem Scale (RSES) measuring self-esteem (SE) with 10 items (von Collani & Herzberg, 2003); and a short version of the Narcissistic Personality Inventory (NPI-15) measuring narcissism (NAR) with 15 items (Spangenberg et al., 2013). To avoid confusing participants, we used the same rating scale for each questionnaire. All items were rated on a 5-point scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*), except the dichotomous forced-choice items from the NPI-15, which required participants to choose one response out of two alternatives.

Procedure

We created three equivalent online studies with Questback's EFS Survey software for academic use (UNIPARK; <http://www.unipark.com/>). They were identical except for the font color of all text elements (e.g., general instructions, measures). We decided to color all text elements from the beginning to prevent participants from becoming suspicious about the purpose of the study. All text elements were shown on a plain white background (font family “Arial, Helvetica, Sans Serif,” font size 15px, and line height 1.4 em). The black colored logo of the University was centered at the top of each page.

To gain insight into effects of color on psychological functioning, it is important to take into account the fact that color is complex. Color can be decomposed into basic dimensions (e.g., lightness, chroma, and hue model [LCH]; hue, saturation, and value model [HSV]; or red, green, and blue model [RGB]) (Fairchild, 2015). The LCH model is a commonly used color model in psychological research because the dimensions of the LCH model correspond to color qualities that are perceptible by humans. In this model, lightness (or brightness) refers to the amount of light emitted by an object, ranging from bright to dark. Chroma (or saturation/intensity) is the amount of hue emitted by an object, ranging from saturated to unsaturated. And hue is the pigment of

color, probably the most salient quality of color. It refers to the aspect most humans think of when talking about color. Hue evolves from the perceptible spectrum of light (wavelength) reflected by an object, ranging from blue to red (i.e., the shortest to longest perceptible wavelengths for trichromats).

Humans' specific ability to perceive the colors from the red spectrum suggests that red may play an important role in psychological functioning. Furthermore, published research on effects of color on psychological functioning in achievement contexts and other contexts (e.g., a romantic context) is primarily comprised of effects of red. Both lines of research support the assumption that effects of color are most likely to occur with regard to the color red. Whereas recent research has consistently aimed to hold chroma and lightness values constant for comparisons of different colors, hue values of colors from the red spectrum vary within the literature. We manipulated the hue values for chromatic font colors (red and blue) and kept the lightness and chroma values constant. We used black as an achromatic control. The lightness and chroma values of the font colors were matched to be as close as possible in a trial-and-error process with the free color tool ColorHexa (<http://www.colorhexa.com/>). Finally, we used hexadecimal color #A10000 (entering values LCH 33, 74, 40 into ColorHexa resulted in exact values LCH = 33.090, 73.838, 39.732) for the red font color, and #0045BC (values LCH 33, 74, 293; exact values LCH = 33.745, 73.697, 294.650) for the blue font color. The black font color was hexadecimal color #000000 (LCH = 0.000, 0.000, 360.000). The font colors we used are illustrated in Figure S1 in the supplemental information (available online at <https://osf.io/cm9fd/>).

After the general instructions and a request for consent were presented the questionnaires were each administered on a single page in a fixed order for all participants. The order of the measures was: I-PANAS-SF, BFI, BIDR-20, SWLS, RSES, NPI-15, AD ACL, IPANAT, and I-PANAS-SF. The I-PANAS-SF was presented twice, once at the beginning and once at the end. After completing the measures section, participants were asked to provide demographic information (gender, age, first language, highest school degree, final grade in that degree (i.e. GPA), employment status, as well as information about their current study status). The next to last page asked participants to self-evaluate the quality of the data they had provided with a dichotomous variable "I have completed all questionnaires with care" versus "I have completed at least one questionnaire without care (e.g., because I wanted to finish up fast or just clicked through)" and probed for their potential awareness of the purpose of the study (with open questions "Did you recognize anything suspicious while participating in the study?" and "What do you think we are trying to find out?"). Participants' self-reported color vision impairment was assessed on the last page (participants were required to choose one of the four alternatives "red-green color vision impairment," "blue-yellow color vision impairment," "color vision impairment of another type," or "not aware of any color vision impairment"). Individual feedback was

available after the study was completed. In order to be eligible to continue, participants were required to answer all items, except the free-text items about their potential awareness of the purpose of the study.

In a first step, the study was introduced as a study of personality and emotion during a single lecture that was primarily attended by first-semester psychology students. Students could sign up for a mailing list to receive a link to participate. Interested students were randomly assigned to one condition, and an e-mail with a link for participation was sent to them. Additional participants were recruited in different seminars following the same course of action. At a later stage, we offered the opportunity to receive individual feedback as an incentive for students who did not need course credit and nonstudents. To promote the study in social networks, we created a referral page that forwarded participants to one color condition in an alternating fashion. Visits to the referral page were monitored with cookies to ensure that participants were forwarded to the same color condition when they clicked the link a second time.

Results

We used G*Power (Version 3.1.9.2) to calculate the power (Faul, Erdfelder, Lang, & Buchner, 2007) and IBM SPSS Statistics (Version 25) for data analysis. In addition, we used the SPSS script provided by Wuensch (2012) to calculate confidence intervals for Cohen's *ds*. We had expected positivity effects of red on state and trait variables. To rule out influences of potential covariates due to unsuccessful random assignment of participants to different color conditions, which could provide an alternative explanation for group differences because age and gender differences in traits are well established (e.g., for the Big Five) (Feingold, 1994; Soto, John, Gosling, & Potter, 2011), we initially checked for group differences regarding age and gender by applying an ANOVA and a chi-square test for homogeneity. Mean scores of all items for all assessed states and traits rated on a 5-point scale were calculated for the analyses. The sum score of the dichotomous NPI-15 items was divided by three to achieve better comparability to all other 5-point scales. To detect effects of color condition for multiple dependent variables, we applied two MANOVAs with eight and 10 dependent variables separately for all state and all trait variables. We examined the Pillai-Bartlett trace statistic due to its robustness against violations of MANOVA assumptions and its high statistical power for dependent variables with a multidimensional structure (Olson, 1976). Uncorrected post hoc *t* tests were applied to gain more insight into potential effects on single variables.

The ANOVA conducted to detect an effect of color condition (red vs. blue vs. black) on age revealed no statistically significant effect of color condition, $F(2, 142) = 0.07, p = 0.930, \eta^2 = 0.001$. The chi-square test of homogeneity revealed no statistically significant differences in the gender ratios in the groups, $\chi^2(2) = 0.004, p = 0.998$. The results pointed to success in the random assignment of participants to the color conditions regarding the potential covariates of age and gender.

We had not determined the sample size a priori but tried to recruit as many participants as possible. Sensitivity analysis with a 95% chance of detecting an existing effect revealed a detectable effect size of $\eta^2 = .09$ for eight dependent variables (regarding the sample size we had, power to detect a small effect of $\eta^2 = .01$ was <50%, a medium effect of $\eta^2 = .06$ was 72%, and a large effect of $\eta^2 = .14$ was >99%) and of $\eta^2 = .10$ for 10 dependent variables (regarding the sample size we had, power to detect a small effect of $\eta^2 = .01$ was <50%, a medium effect of $\eta^2 = .06$ was 66%, and a large effect of $\eta^2 = .14$ was >99%) for a global effect in a MANOVA.

Table 1 presents an overview of descriptive statistics (means, standard deviations, effect sizes) and the reliabilities (Cronbach's alpha) for all explicitly and implicitly assessed states as well as all explicitly assessed traits. We had assumed that color would influence explicit and implicit states. To detect an effect of color on all assessed self-rated states (PA, NA, energy, tension, iPA, iNA, PA, NA), we submitted the data to a MANOVA with color condition as a between-subjects factor (red vs. blue vs. black). Indicated by the Pillai-Bartlett trace statistic, the MANOVA demonstrated no statistically significant effect across the states for the between-subjects factor color condition, $V = .12$, $F(16, 272) = 1.06$, $p = .393$, $\eta^2 = .06$. Uncorrected post hoc t tests were all statistically nonsignificant with $ps \geq .254$ and Cohen's $ds \leq |0.23|$. We did not find the assumed effect of color on affective states.

We had further hypothesized that color would influence personality trait self-ratings. To detect an effect of color on all self-rated personality traits (N, E, O, A, C, SDE, IM, SWL, SE, NAR), we submitted the data to a MANOVA with color condition as a between-subjects factor (red vs. blue vs. black). Indicated by the Pillai-Bartlett trace statistic, the MANOVA demonstrated a statistically significant effect across the personality traits for the between-subjects factor color condition, $V = .22$, $F(20, 268) = 1.66$, $p = .041$, $\eta^2 = .11$. Table S1 in the supplemental information presents the intercorrelations of all trait mean scores (available online at <https://osf.io/cmqfd/>). To further explore the effect of color on traits we conducted uncorrected post hoc t tests revealing statistically significant effects on O for red versus blue, $t(95) = 2.69$, $p = .008$, Cohen's $d = 0.55$, and SWL, $t(95) = -2.19$, $p = .031$, Cohen's $d = -0.44$, and on A for red versus black, $t(95) = -2.25$, $p = .027$, Cohen's $d = -0.46$, and for blue versus black, $t(94) = -2.82$, $p = .006$, Cohen's $d = -0.58$. All other effects remained nonsignificant with $ps \geq .08$ and Cohen's $ds \leq |0.35|$. Results of all conducted uncorrected post hoc t tests as well as 95% confidence intervals are included in **Table 1**, complementary the distribution of p values of all tests are illustrated in **Figure 1**. The chance of detecting at least two statistically significant results when running 10 significance tests as done for the conservative red versus blue comparisons of traits was 8.61% and thereby lower than the chance of detecting at least one statistically significant result as obtained for the weaker red versus black and blue versus black comparisons, which was 40.13%. At first view, we found supporting evidence for the hypothesized effect of red on personality traits.

However, for the conservative red versus blue comparisons only the effect for openness pointed towards the expected direction of positivity effects of red.

Discussion

In line with the color-in-context theory and earlier findings, which suggested that exposure to red had detrimental effects in achievement contexts, we tested whether exposure to red would lead people to provide more favorable ratings of themselves. Self-ratings were presented in red, blue, and black font colors, and the results were somewhat inconclusive. We observed no differences in states and inconsistent differences in personality traits: When exposed to red in comparison with blue, self-ratings increased for openness but decreased for satisfaction with life, whereas self-ratings of agreeableness decreased for red and blue in comparison with black. As differences between chromatic (red or blue) and achromatic colors (black) can be attributed not only to hue but also to chroma and lightness, the effects for openness and satisfaction with life can be attributed to differences in hue, whereas the effects that were observed for agreeableness can also be attributed to differences in chroma or lightness. Only the effect of openness was in the expected direction, a tendency to present oneself in a favorable manner when exposed to red. Of course the possibility that the findings were due to chance has to be taken into account. The evidence concerning systematic effects of red in our study was weak, but due to sample size limitations, our study did not allow us to conclude that there was no effect of red on personality trait self-ratings because we could not rule out the possibility that there were small effects that we simply did not detect. To clarify this issue, we decided to conduct a high-powered replication. Such a replication would need to provide an overall test of the effect of red on personality trait self-ratings. Because the Big Five were substantially correlated with the other personality traits and states that we measured in Study 1, and because they can provide a broad picture of personality, we focused on these five traits in the next study.

Study 2

Given that we found some expected and some unexpected effects in our first exploratory study, a second high-powered study was needed to answer the question of whether there is an effect of red on personality trait self-ratings. Because we were interested in an effect of the color red (i.e., hue, which is the most salient quality of color and the aspect most humans think of when talking about color) and not effects of chroma or lightness, we decided to maximize our chances of obtaining the effect by focusing on a comparison of the chromatic colors red and blue, which we had already used in Study 1. To increase power and thus to be able to detect even small effects as observed in Study 1, we aimed to collect a sample that was large enough to detect even small effects. To achieve a generalization of possible effects regarding the applied measures, we complemented Study 1 by using another personality trait measure to assess the Big Five. We also decided to reduce the study to include only the

Table 1: Means, Standard Deviations, Effect Sizes and 95% Confidence Intervals for Scale Mean Scores in Different Font-Color Groups.

Concept	Measure	Scale	Cronbach's alpha (N = 145)	Font color			Cohen's d			95% CI		
				Red (n = 49)	Blue (n = 48)	Black (n = 48)	Red vs. blue	Red vs. black	Blue vs. black	Red vs. blue	Red vs. black	Blue vs. black
State	I-PANAS-SF	PA t1	.76	2.95 (0.64)	2.93 (0.69)	3.07 (0.60)	0.03	-0.19	-0.22	[-0.37, 0.43]	[-0.59, 0.21]	[-0.62, 0.18]
		NA t1	.62	1.33 (0.42)	1.43 (0.49)	1.40 (0.40)	-0.20	-0.16	0.06	[-0.60, 0.20]	[-0.56, 0.24]	[-0.34, 0.46]
		PA t2	.84	2.91 (0.70)	2.96 (0.77)	3.01 (0.67)	-0.07	-0.14	-0.07	[-0.46, 0.33]	[-0.54, 0.26]	[-0.47, 0.33]
		NA t2	.65	1.28 (0.41)	1.28 (0.38)	1.33 (0.43)	-0.01	-0.11	-0.10	[-0.41, 0.38]	[-0.51, 0.29]	[-0.50, 0.30]
AD ACL	Energy	.89	2.57 (0.61)	2.58 (0.70)	2.54 (0.60)	-0.01	0.06	0.06	[-0.23, 0.22]	[-0.33, 0.46]	[-0.34, 0.46]	
	Tension	.89	1.90 (0.59)	1.86 (0.60)	1.73 (0.52)	0.06	0.30	0.23	[-0.34, 0.46]	[-0.10, 0.70]	[-0.17, 0.64]	
IPANAT	iPA	.65	2.33 (0.42)	2.37 (0.38)	2.34 (0.29)	-0.10	-0.04	0.07	[-0.50, 0.30]	[-0.44, 0.36]	[-0.33, 0.47]	
	iNA	.67	1.81 (0.39)	1.85 (0.36)	1.80 (0.32)	-0.12	0.03	0.16	[-0.52, 0.28]	[-0.37, 0.43]	[-0.24, 0.56]	
BFI	N	.83	3.12 (0.74)	3.08 (0.75)	2.89 (0.80)	0.04	0.30	0.25	[-0.35, 0.44]	[-0.11, 0.70]	[-0.15, 0.65]	
	E	.86	3.48 (0.62)	3.33 (0.71)	3.33 (0.81)	0.23	0.22	0.01	[-0.17, 0.63]	[-0.18, 0.62]	[-0.39, 0.41]	
O	A	.80	3.97 (0.53)	3.66 (0.58)	3.81 (0.64)	0.55**	0.27	-0.24	[0.14, 0.95]	[-0.13, 0.67]	[-0.64, 0.16]	
	C	.69	3.68 (0.46)	3.60 (0.49)	3.92 (0.60)	0.15	-0.46**	-0.58**	[-0.25, 0.55]	[-0.86, -0.05]	[-0.98, -0.17]	
SDE	IM	.84	3.46 (0.67)	3.53 (0.62)	3.67 (0.63)	-0.10	-0.33	-0.23	[-0.50, 0.29]	[-0.73, 0.07]	[-0.63, 0.17]	
	SWL	.66	3.03 (0.49)	3.05 (0.48)	3.10 (0.58)	-0.04	-0.12	-0.08	[-0.44, 0.35]	[-0.52, 0.28]	[-0.48, 0.32]	
SWLS	SE	.63	3.01 (0.61)	3.03 (0.63)	2.93 (0.52)	-0.03	0.14	0.16	[-0.43, 0.37]	[-0.26, 0.53]	[-0.24, 0.56]	
	RSES	.85	3.29 (0.80)	3.65 (0.84)	3.57 (0.79)	-0.44*	-0.35	0.10	[-0.85, -0.04]	[-0.75, 0.05]	[-0.30, 0.50]	
NPI-15	NAR	.91	3.61 (0.79)	3.72 (0.80)	3.77 (0.93)	-0.14	-0.19	-0.06	[-0.53, 0.26]	[-0.58, 0.21]	[-0.46, 0.34]	
	NAR	.75	1.61 (1.05)	1.44 (0.98)	1.33 (0.94)	0.16	0.29	0.12	[-0.23, 0.56]	[-0.12, 0.69]	[-0.28, 0.52]	

Note: M (SD) in column Font color. AD ACL = Activation-Deactivation Adjective Check List; A = agreeableness; BIDR-20 = 20 item version of the Balanced Inventory of Desirable Responding; BFI = Big Five Inventory; C = conscientiousness; E = extraversion; iNA = implicit negative affect; iPANAT = implicit positive affect; iPANAT = Implicit Positive and Negative Affect Test; IM = impression management; I-PANAS-SF = International Positive and Negative Affect Schedule Short Form; t1 = measured at the beginning; t2 = measured in the end; NAR = narcissism; NPI-15 = 15 item version of the Narcissistic Personality Inventory; NA = negative affect; N = neuroticism; O = openness; PA = positive affect; RSES = revised Rosenberg Self-Esteem Scale; SWL = satisfaction with life; SWLS = Satisfaction with Life Scale; SDE = self-deceptive enhancement; SE = self-esteem. * $p < .05$. ** $p < .01$.

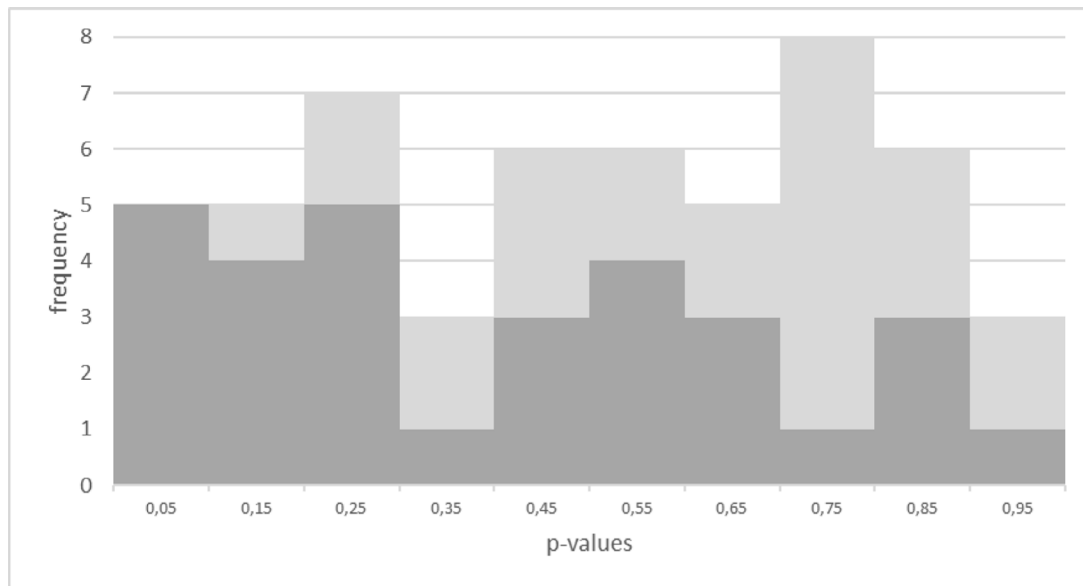


Figure 1: Frequency Distribution of italicized Values of Observed Effects.

Note. *P* values for all color comparisons for measured traits in dark grey as well as states and traits together in light grey; *p* values on the x-axis represent class midpoints.

Big Five because a shorter study would be more attractive to potential participants, and thus, it would be possible to recruit a large sample.

Method

Participants

Data were again collected online. The study was advertised as a study of personality. Participants were recruited from a research mailing list to which they had subscribed, Internet platforms (<http://www.psychologie-heute.de>, <https://www.surveycircle.com>), social media groups (Facebook, Xing), and students from a University in southern Germany. Students received partial course credit for participation upon request (course credit was requested 109 times). All participants could request feedback on their personality profile (feedback was requested 969 times). Participants requesting feedback were statistically significantly different from those not requesting feedback only on trait O for both applied measures, thus supporting the general validity of the measurements. All participants provided informed consent. The study was approved by the ethics committee of the University of Bamberg (dossier number 2019-07/21).

During October 2016 and December 2016, we registered 1,412 first-page views, of which 1,194 individuals completed the study. On a dichotomously scored variable “I have completed all questionnaires with care” versus “I have completed at least one questionnaire without care (e.g., because I wanted to finish up fast or just clicked through),” 1,049 participants indicated that they had completed the study with care. Three participants were excluded because they classified themselves to the gender category “other” (group estimates would be biased by the low *n*). Thirty-eight others were excluded because they had self-reported a color vision impairment (20 reported a red-green color vision impairment, 3 reported a blue-yellow color vision impairment, and 15 reported a color

vision impairment of another type). Another participant was excluded on the basis of a suspicious data pattern (e.g., age of 5 years, overall response time of 178 s). None of the remaining *N* = 1,007 participants (76% women) correctly anticipated the purpose of the study. The mean age of participants was 35.0 years (*SD* = 5.6).

Measures

The BFI described in Study 1 was complemented by a brief German version of the NEO Five-Factor Inventory (NEO-FFI-30) (Körner et al., 2015, 2008) also measuring N, E, O, A, and C with six items each. All items were rated on a 5-point scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*).

Procedure

Study 2 was designed to be similar to Study 1 except that the black font color was not included due to the reasons given above. Red and blue font colors were matched on chroma and lightness and were identical to those used in Study 1 (LCH = 33.090, 73.838, 39.732 for the red font color and LCH = 33.745, 73.697, 29.650 for the blue font color). The measures used were the BFI and the NEO-FFI 30. Different from Study 1, the measures were presented in a randomized order, but again, each measure was presented on a single browser page. Further deviating from Study 1 with respect to the measures, we added an additional item to the last page (on which participants’ self-reported color vision impairment was assessed) to assess participants’ favorite color. Participants were required to choose their favorite color from a list of color names containing red, green, blue, yellow, white, grey, and black. In line with previous findings (Crozier, 1999), blue was the color most often preferred by participants. The distribution of favorite color frequencies is illustrated in Figure S2 in the supplemental information (available online at <https://osf.io/cm9fd/>).

Results

The types of software and scripts used for power calculations and data analyses were identical to Study 1. We used JASP (Version 0.8.3.1) for additional calculations of Bayes factors in support of the null hypothesis. Again, we hypothesized an effect of red on self-rated personality traits independent of the type of questionnaire. Again, we initially checked for group differences regarding age and gender by applying an independent *t* test and a chi-square test of homogeneity. Mean scores including all items for measuring each trait were calculated for the analyses. To detect effects of hue for multiple dependent variables in a repeated-measures design, we computed a MANOVA with the 10 dependent variables. Again, we examined the Pillai-Bartlett trace statistic. Uncorrected post hoc *t* tests were computed to gain further insights into potential effects on single variables.

The independent *t* test computed to detect an effect of hue (blue vs. red) on age revealed no statistically significant effect of hue, $t(1005) = 0.46, p = .646, d = 0.03$. However, the chi-square test of homogeneity revealed a statistically significant difference in the gender ratio in the groups, $\chi^2(1) = 3.83, p = .050$. Therefore, following a conservative approach we included gender as a covariate in the main analysis to rule out systematic bias due to the confounding variable.¹

We did not determine sample size a priori but tried to recruit as many participants as possible to minimize the chance of false positive findings. A sensitivity analysis with a 95% chance of detecting an existing effect revealed a detectable effect size of $\eta^2 = .02$ for 10 dependent variables (regarding the sample size we had, power to detect a small effect of $\eta^2 = .01$ was 77%, a medium effect of $\eta^2 = .06$ was >99%, and a large effect of $\eta^2 = .14$ was >99%) for a global effect in a MANOVA.

An overview of the descriptive statistics (means, standard deviations, effect sizes) and achieved reliability (Cronbach's alpha) for all self-rated personality traits can be found in **Table 2**. We had hypothesized that red would influence the personality trait self-ratings. To detect an effect of hue on all self-rated personality traits, we submitted the data to a repeated-measures MANOVA with hue as a between-subjects factor (blue vs. red), measure as a within-subjects factor (BFI vs. NEO-FFI 30), and gender (female vs. male) as a covariate. As indicated by the Pillai-Bartlett trace statistic, the MANOVA demonstrated no statistically significant effect across the traits for the between-subjects factor hue, $V = .001, F(5, 1000) = 0.13, p = .987, \eta^2 = .001$, a statistically significant effect for the covariate gender, $V = .12, F(5, 1000) = 27.01, p < .001$, a statistically significant effect for the within-subjects factor measure, $V = .52, F(5, 1000) = 216.91, p < .001, \eta^2 = .52$, a statistically significant effect for the interaction of the factor measure and the covariate gender, $V = .03, F(5, 1000) = 5.95, p < .001, \eta^2 = .03$, and a statistically non-significant effect for the interaction of the within-subjects factor measure and the between-subjects factor hue, $V = .002, F(5, 1000) = 0.34, p = .884, \eta^2 = .002$. Uncorrected post hoc *t* tests were all non-significant with $ps \geq .365$ and Cohen's $ds \leq 0.06$. The power to detect an effect of $d = 0.10$ with the group sample sizes we obtained was 35.41%. With this level of power, the chance of detecting at least one statistically significant result was 98.74%. Additional calculations of Bayes factors (null/alternative) for a standard normally distributed prior suggested at least 10.56:1 times higher support for the null hypothesis by the data. To sum up, we could not find any supporting evidence for our hypothesized effect of hue on self-ratings of personality traits. Instead, we obtained clear evidence in support of the null hypothesis.

Table 2: Means, Standard Deviations, Effect Sizes and 95% Confidence Intervals for Scale Mean Scores in Different Font-Color Groups.

Measure	Scale	Cronbach's alpha (N = 1007)	Font color		Cohen's <i>d</i>	95% CI
			Red (n = 500)	Blue (n = 507)		
BFI	N	.85	2.99 (0.85)	2.99 (0.81)	0.01	[-0.12, 0.13]
	E	.89	3.30 (0.86)	3.32 (0.81)	-0.02	[-0.15, 0.10]
	O	.83	3.75 (0.64)	3.77 (0.62)	-0.03	[-0.16, 0.09]
	A	.73	3.55 (0.61)	3.58 (0.60)	-0.06	[-0.18, 0.07]
	C	.82	3.59 (0.66)	3.58 (0.64)	0.02	[-0.11, 0.14]
NEO-FFI 30	N	.88	2.68 (1.00)	2.70 (0.97)	-0.02	[-0.14, 0.11]
	E	.78	3.14 (0.78)	3.17 (0.67)	-0.05	[-0.17, 0.08]
	O	.78	3.69 (0.79)	3.69 (0.79)	0.00	[-0.12, 0.13]
	A	.75	3.85 (0.72)	3.87 (0.70)	-0.02	[-0.15, 0.10]
	C	.80	3.95 (0.69)	3.94 (0.67)	0.01	[-0.12, 0.13]

Note: *M* (*SD*) in column Font color. A = Agreeableness; BFI = Big Five Inventory; C = Conscientiousness; E = Extraversion; N = Neuroticism; NEO-FFI 30 = 30 item version of the NEO Five-Factor Inventory; O = Openness.

Discussion

Study 2 was a follow-up study that was designed to replicate and further investigate the inconclusive results from Study 1 regarding an effect of red on personality trait self-ratings. We aimed to recruit a large sample so that we would have enough power to detect even small effects. There were no effects of red on self-rated personality traits. A sensitivity analysis had revealed high power and thus the ability to detect even very small effects. The results impressively failed to support the hypothesis that red impacts personality trait self-ratings. Thus, the hypothesis that red has negligible effects, if any, on personality trait self-ratings is much more likely to hold.

General Discussion

In line with previously reported detrimental effects of perceiving red on results of performance measures, we had argued that perceiving red might also lead to distortions on self-ratings. Just like performance measures, self-ratings can be perceived as achievement contexts because evaluations of the results are inherent. In line with the color-in-context theory, it may thus be argued that fear of failure will trigger the motive to present oneself in a favorable manner if one is confronted with red during personality trait self-rating.

As self-ratings are among the most often applied assessment measures in psychological research, knowing about possible distortions caused by exposure to red is of great importance.

Furthermore, possible effects of red on self-ratings would expand the framework of the color-in-context theory by linking effects of red to self-ratings, thus expanding the claim of general effects of color on psychological functioning in achievement contexts. In Study 1, in line with the color-in-context theory, we expected that exposure to red would lead to more positive self-ratings, that is, higher scores on traits with a positive valence (e.g., extraversion, openness, agreeableness, conscientiousness) and lower scores on traits with a negative valence (e.g., neuroticism). However, we found only weak evidence for this argument. Still, due to power limitations, we were not able to rule out the possibility that there may be small effects. Inconclusive effects on differences between red or blue in comparison with black could have been attributed to chroma or lightness.

Thus, we conducted a high-powered second study that would be able to detect even small effects with two measures of the Big Five, and we restricted the study to the chromatic colors red and blue already used in Study 1, which were matched on their qualities of chroma and lightness. Despite the fact that we applied rigorous methods, we could not find any effect of exposure to red on the results of personality trait self-ratings.

The results imply that, contrary to our expectation that self-ratings would also be affected by efforts to present oneself favorably when exposed to red, there was no effect of red. The finding may be considered an indication that the color-in-context theory needs to be updated regarding the contextual effects of color on psychological functioning.

One potential weakness of our study could also be considered a strength. Because we conducted our research online, we could not control for color at the spectral level but only at the device level. However, control at the spectral level would be a threat to external validity. Control at the spectral level reduces color to a level of unnatural appearance because color in a natural environment is far from consistent at the spectral level. Consider a stop sign, for example: Nobody would claim that the sign is not red, whether it is sunny, cloudy, or even rainy outside, and no matter what the position of the sun is during the day. The spectral level changes, but the color is perceived as red and fulfills the function of catching drivers' attention. Thus, not controlling for color at the spectral level resembles reality. In fact, perceived nuances of hue might furthermore not be the same for every human individual (e.g., what one person calls blue may be another person's green). But at least red and blue as categories do not overlap and have universally different appearances (Abramov & Gordon, 1994; Bornstein, 1973). From a statistical perspective, not controlling for color at the spectral level increases noise in the data and has the potential to mask effects. Still, even if effects were masked by noise, unmasking such effects would not increase their size.

Participants were offered the opportunity to instantly obtain feedback on their personality characteristics. We did this to intensify the evaluative context in line with research on effects of red in achievement contexts (Elliot et al., 2007). One might claim that an anonymous online study is a weak manipulation that does not provide a close enough resemblance to an achievement context, or that our intention to intensify the evaluation by offering feedback could also have undermined the potential for responses to become distorted because participants could have tried to rate themselves as accurately as possible in order to obtain realistic insights into their personality. However, the motive to present oneself in a positive manner is considered to have conscious and unconscious components (He et al., 2015; Paulhus & Reid, 1991). If effects of color on psychological functioning are unconscious – as proposed by the color-in-context theory – efforts to rate oneself as accurately as possible would still be undermined with respect to the unconscious component. Furthermore, research has also suggested that especially failure in an anonymous context leads to favorable views of oneself in self-ratings (Brown & Gallagher, 1992).

We manipulated the font color because we thought certain other manipulations (e.g., of the background color) would be too obtrusive. In line with our decision to manipulate font color, not a single participant guessed the purpose of the study. Even if a more obtrusive manipulation might be considered more powerful, a red background should be considered the exception rather than the rule in online research. Thus, an unobtrusive manipulation has more external validity. Moreover, manipulations in research on performance measures tend to be rather unobtrusive (Elliot et al., 2007).

We did not pre-register the studies nor did we calculate a priori sample sizes because we aimed to

recruit as many participants as possible in order to achieve maximum power. Instead, we conducted a detailed sensitivity analysis to illustrate the implications of the sample size that we had procured in this manner. Further, we report all data exclusions, all manipulations, and all measures in the study (Simmons, Nelson, & Simonsohn, 2012).

To summarize, on the basis of the color-in-context theory and its claim that red might trigger fear of failure in achievement contexts, we expected that there would be effects of red on self-ratings. This idea is in line with the repeatedly found effects of red on performance measures. However, we did not find convincing empirical evidence of such effects. Even more, we found evidence against such effects.

Considering that self-ratings are among the most often used measures in psychological research and that color is easy to apply and is frequently used as an element of style in online research, our results can be considered reassuring: Despite theoretical assumptions, red should not bias data collection whether intentionally or unintentionally applied.

Regarding the recently growing literature on effects of color on psychological functioning, our results suggest that caution is warranted concerning the assumption of broad and general effects of color on psychological functioning. In particular, the assumption that red leads to biases in evaluation contexts should be revised. Relevant contexts and outcomes need to be defined more precisely to make clear when effects are to be expected and when they are not.

Finally, our findings show the importance of high-powered replications. Our argument is also in line with the postulate to publish null results to allow for the accumulation of unbiased knowledge.

Context

Is there an effect of red on self-ratings? More than ever before, research is being conducted online, and self-ratings are one of the most frequently used methods in psychological research. Especially online research easily allows researchers to apply color stimuli intentionally or unintentionally, but so far there have been no publications on effects of red on self-ratings. Although the color-in-context theory and previous findings offer a theoretical framework that would have led us to expect such effects, we found no evidence of effects of any practical importance in a large data set. We conclude that there is no effect of red on self-ratings.

Data Accessibility Statement

All stimuli, presentation materials, participant data, and analysis scripts for this article are available online at <https://osf.io/x493s/>.

Supplemental Information

Supplemental information for this article is available online at the Open Science Framework (<https://osf.io/cmqfd/>).

Note

¹ Disregarding gender as a covariate did not lead to any noteworthy changes in significance levels for the remaining factors.

Acknowledgements

We strongly support transparency and openness in science and therefore “report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures” in our studies (Simmons, Nelson, & Simonsohn, 2012).

We thank Johannes Leder for providing helpful feedback on an early draft of the article and Jane Zagorski for language editing. Furthermore, we thank M. Brent Donnellan Richard E. Lucas and an anonymous reviewer for their valuable feedback on the submitted article.

Competing Interests

The authors have no competing interests to declare.

Author Contributions

- Conception and design: DW, AS
- Acquisition of data: DW, AS
- Analysis and interpretation of data: DW, AS
- Drafting and/or revising the article: DW, AS
- Final approval for publication: DW, AS

References

- Abramov, I., & Gordon, J.** (1994). Color appearance: On seeing red – or yellow, or green, or blue. *Annual Review of Psychology*, *45*(1), 451–485. DOI: <https://doi.org/10.1146/annurev.ps.45.020194.002315>
- Bäckström, M.** (2007). Higher-order factors in a five-factor personality inventory and its relation to social desirability. *European Journal of Psychological Assessment*, *23*(2), 63–70. DOI: <https://doi.org/10.1027/1015-5759.23.2.63>
- Bäckström, M., Björklund, F., & Larsson, M. R.** (2009). Five-factor inventories have a major general factor related to social desirability which can be reduced by framing items neutrally. *Journal of Research in Personality*, *43*(3), 335–344. DOI: <https://doi.org/10.1016/j.jrp.2008.12.013>
- Borghans, L., Meijers, H., & Ter Weel, B.** (2008). The role of noncognitive skills in explaining cognitive test scores. *Economic Inquiry*, *46*(1), 2–12. DOI: <https://doi.org/10.1111/j.1465-7295.2007.00073.x>
- Bornstein, M. H.** (1973). Color vision and color naming: A psychophysiological hypothesis of cultural difference. *Psychological Bulletin*, *80*(4), 257–285. DOI: <https://doi.org/10.1037/h0034837>
- Brooker, A., & Franklin, A.** (2016). The effect of colour on children's cognitive performance. *British Journal of Educational Psychology*, *86*(2), 241–255. DOI: <https://doi.org/10.1111/bjep.12101>
- Crozier, W. R.** (1999). The meanings of colour: Preferences among hues. *Pigment & Resin Technology*, *28*(1), 6–14. DOI: <https://doi.org/10.1108/03699429910252315>

- Duckworth, A. L., Quinn, P. D., Lynam, D. R., Loeber, R., & Stouthamer-Loeber, M.** (2011). Role of test motivation in intelligence testing. *Proceedings of the National Academy of Sciences*, *108*(19), 7716–7720. DOI: <https://doi.org/10.1073/pnas.1018601108>
- Elliot, A. J.** (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist*, *34*(3), 169–189. DOI: https://doi.org/10.1207/s15326985ep3403_3
- Elliot, A. J.** (2015). Color and psychological functioning: A review of theoretical and empirical work. *Frontiers in Psychology*, *6*(368), 1–8. DOI: <https://doi.org/10.3389/fpsyg.2015.00368>
- Elliot, A. J., & Aarts, H.** (2011). Perception of the color red enhances the force and velocity of motor output. *Emotion*, *11*(2), 445–449. DOI: <https://doi.org/10.1037/a0022599>
- Elliot, A. J., & Maier, M. A.** (2012). Color-in-context theory. In M. P. Zanna, P. Devine, J. M. Olson, & A. Plant (Eds.), *Advances in Experimental Social Psychology* (Vol. 45, pp. 61–125). DOI: <https://doi.org/10.1016/B978-0-12-394286-9.00002-0>
- 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: <https://doi.org/10.1146/annurev-psych-010213-115035>
- 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*(1), 154–168. DOI: <https://doi.org/10.1037/0096-3445.136.1.154>
- Fairchild, M. D.** (2015). Color models and systems. In A. J. Elliot, M. D. Fairchild, & A. Franklin (Eds.), *Handbook of Color Psychology* (1st ed., pp. 9–26). Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9781107337930.003>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A.** (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. DOI: <https://doi.org/10.3758/BF03193146>
- Feingold, A.** (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin*, *116*(3), 429–456. DOI: <https://doi.org/10.1037/0033-2909.116.3.429>
- Freund, P. A., & Holling, H.** (2011). Who wants to take an intelligence test? Personality and achievement motivation in the context of ability testing. *Personality and Individual Differences*, *50*(5), 723–728. DOI: <https://doi.org/10.1016/j.paid.2010.12.025>
- Glaesmer, H., Grande, G., Braehler, E., & Roth, M.** (2011). The German version of the Satisfaction with Life Scale (SWLS): Psychometric properties, validity, and population-based norms. *European Journal of Psychological Assessment*, *27*(2), 127–132. DOI: <https://doi.org/10.1027/1015-5759/a000058>
- Gnambs, T., Appel, M., & Batinic, B.** (2010). Color red in web-based knowledge testing. *Computers in Human Behavior*, *26*(6), 1625–1631. DOI: <https://doi.org/10.1016/j.chb.2010.06.010>
- He, J., van de Vijver, F. J. R., Dominguez Espinosa, A., Abubakar, A., Dimitrova, R., Adams, B. G., Villieux, A., et al.** (2015). Socially desirable responding: Enhancement and denial in 20 countries. *Cross-Cultural Research*, *49*(3), 227–249. DOI: <https://doi.org/10.1177/1069397114552781>
- Imhof, M.** (1998). Erprobung der deutschen Version der Adjektiv-Checkliste nach Thayer (1989) zur Erfassung der aktuellen Aktiviertheit. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *19*(3), 179–186.
- Jacobs, G. H.** (2008). Primate color vision: A comparative perspective. *Visual Neuroscience*, *25*(5–6), 619–633. DOI: <https://doi.org/10.1017/S0952523808080760>
- Körner, A., Czajkowska, Z., Albani, C., Drapeau, M., Geyer, M., & Braehler, E.** (2015). Efficient and valid assessment of personality traits: Population norms of a brief version of the NEO Five-Factor Inventory (NEO-FFI). *Archives of Psychiatry and Psychotherapy*, *17*(1), 21–32. DOI: <https://doi.org/10.12740/APP/36086>
- Körner, A., Geyer, M., Roth, M., Drapeau, M., Schmutzer, G., Albani, C., Brähler, E., et al.** (2008). Persönlichkeitsdiagnostik mit dem NEO-Fünf-Faktoren-Inventar: Die 30-Item-Kurzversion (NEO-FFI-30). *Psychotherapie, Psychosomatik, Medizinische Psychologie*, *58*(6), 238–245. DOI: <https://doi.org/10.1055/s-2007-986199>
- Krohne, H. W., Egloff, B., Kohlmann, C.-W., & Tausch, A.** (1996). Untersuchungen mit einer deutschen Version der “Positive and Negative Affect Schedule” (PANAS). *Diagnostica*, *42*(2), 139–156. DOI: <https://doi.org/10.1037/t49650-000>
- Kuhbandner, C., & Pekrun, R.** (2013). Joint effects of emotion and color on memory. *Emotion*, *13*(3), 375–379. DOI: <https://doi.org/10.1037/a0031821>
- Lang, F. R., Lüdtke, O., & Asendorpf, J. B.** (2001). Testgüte und psychometrische Äquivalenz der deutschen Version des Big Five Inventory (BFI) bei jungen, mittelalten und alten Erwachsenen. *Diagnostica*, *47*(3), 111–121. DOI: <https://doi.org/10.1026//0012-1924.47.3.111>
- 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*(6), 1273–1276. DOI: <https://doi.org/10.1016/j.jesp.2009.06.003>
- Maier, M. A., Hill, R. A., Elliot, A. J., & Barton, R. A.** (2015). Color in achievement contexts in humans. In A. J. Elliot, M. D. Fairchild, & A. Franklin (Eds.), *Handbook of Color Psychology* (pp. 568–584). DOI: <https://doi.org/10.1017/CBO9781107337930.028>
- Moller, A. C., Elliot, A. J., & Maier, M. A.** (2009). Basic hue-meaning associations. *Emotion*, *9*(6), 898–902. DOI: <https://doi.org/10.1037/a0017811>

- Musch, J., Brockhaus, R., & Bröder, A.** (2002). Ein Inventar zur Erfassung von zwei Faktoren sozialer Erwünschtheit. *Diagnostica*, *48*(3), 121–129. DOI: <https://doi.org/10.1026//0012-1924.48.3.121>
- Olson, C. L.** (1976). On choosing a test statistic in multivariate analysis of variance. *Psychological Bulletin*, *83*(4), 579–586. DOI: <https://doi.org/10.1037/0033-2909.83.4.579>
- Osorio, D., Smith, A. C., Vorobyev, M., & Buchanan-Smith, H. M.** (2004). Detection of fruit and the selection of primate visual pigments for color vision. *The American Naturalist*, *164*(6), 696–708. DOI: <https://doi.org/10.1086/425332>
- Paulhus, D. L.** (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49–69). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Paulhus, D. L., & Reid, D. B.** (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology*, *60*(2), 307–317. DOI: <https://doi.org/10.1037//0022-3514.60.2.307>
- Payen, V., Elliot, A. J., Coombes, S. A., Chalabaev, A., Brisswalter, J., & Cury, F.** (2011). Viewing red prior to a strength test inhibits motor output. *Neuroscience Letters*, *495*(1), 44–48. DOI: <https://doi.org/10.1016/j.neulet.2011.03.032>
- Pointer, M. R., & Attridge, G. G.** (1998). The number of discernible colours. *Color Research and Application*, *23*(1), 52–54. DOI: [https://doi.org/10.1002/\(SICI\)1520-6378\(199802\)23:1<52::AID-COL8>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1520-6378(199802)23:1<52::AID-COL8>3.0.CO;2-2)
- Quirin, M., & Bode, R. C.** (2014). An alternative to self-reports of trait and state affect. *European Journal of Psychological Assessment*, *30*(3), 231–237. DOI: <https://doi.org/10.1027/1015-5759/a000190>
- Rushton, J. P., & Erdle, S.** (2010). No evidence that social desirability response set explains the general factor of personality and its affective correlates. *Twin Research and Human Genetics*, *13*(2), 131–134. DOI: <https://doi.org/10.1375/twin.13.2.131>
- Rutchick, A. M., Slepian, M. L., & Ferris, B. D.** (2010). The pen is mightier than the word: Object priming of evaluative standards. *European Journal of Social Psychology*, *40*, 704–708. DOI: <https://doi.org/10.1002/ejsp.753>
- Schermer, J. A., & MacDougall, R.** (2013). A general factor of personality, social desirability, cognitive ability, and the survey of work styles in an employment selection setting. *Personality and Individual Differences*, *54*(1), 141–144. DOI: <https://doi.org/10.1016/j.paid.2012.08.012>
- Schermer, J. A., & Vernon, P. A.** (2010). The correlation between general intelligence (g), a general factor of personality (GFP), and social desirability. *Personality and Individual Differences*, *48*(2), 187–189. DOI: <https://doi.org/10.1016/j.paid.2009.10.003>
- Shi, J., Zhang, C., & Jiang, F.** (2015). Does red undermine individuals' intellectual performance? A test in China. *International Journal of Psychology*, *50*(1), 81–84. DOI: <https://doi.org/10.1002/ijop.12076>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U.** (2012). A 21 word solution. *Dialogue: The Official Newsletter of the Society for Personality and Social Psychology*, *26*(2), 4–12. Retrieved from http://spsp.org/sites/default/files/dialogue_26%282%29.pdf
- Soto, C. J., John, O. P., Gosling, S. D., & Potter, J.** (2011). Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample. *Journal of Personality and Social Psychology*, *100*(2), 330–348. DOI: <https://doi.org/10.1037/a0021717>
- Spangenberg, L., Romppel, M., Bormann, B., Hofmeister, D., Brähler, E., & Strauß, B.** (2013). Psychometrische Überprüfung einer Kurzform des Narcissistic Personality Inventory (NPI-15): Dimensionalität und psychometrische Eigenschaften des NPI-15 in einer repräsentativen Bevölkerungsschichtprobe. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, *63*(8), 341–347. DOI: <https://doi.org/10.1055/s-0032-1333288>
- Steyer, R., Mayer, A., Geiser, C., & Cole, D. A.** (2015). A theory of states and traits – revised. *Annual Review of Clinical Psychology*, *11*(1), 71–98. DOI: <https://doi.org/10.1146/annurev-clinpsy-032813-153719>
- Steyer, R., Schmitt, M., & Eid, M.** (1999). Latent state-trait theory and research in personality and individual differences. *European Journal of Personality*, *13*(5), 389–408. DOI: [https://doi.org/10.1002/\(SICI\)1099-0984\(199909/10\)13:5<389::AID-PER361>3.3.CO;2-1](https://doi.org/10.1002/(SICI)1099-0984(199909/10)13:5<389::AID-PER361>3.3.CO;2-1)
- Thompson, E. R.** (2007). Development and validation of an internationally reliable short-form of the Positive and Negative Affect Schedule (PANAS). *Journal of Cross-Cultural Psychology*, *38*(2), 227–242. DOI: <https://doi.org/10.1177/0022022106297301>
- von Collani, G., & Herzberg, P. Y.** (2003). Eine revidierte Fassung der deutschsprachigen Skala zum Selbstwertgefühl von Rosenberg. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *24*(1), 3–7. DOI: <https://doi.org/10.1024//0170-1789.24.1.3>
- Wuensch, K. L.** (2012). Using SPSS to obtain a confidence interval for Cohen's *d*. Retrieved November 19, 2018, from <http://core.ecu.edu/psyc/wuenschk/SPSS/CI-d-SPSS.pdf>
- Zhang, T., & Han, B.** (2014). Experience reverses the red effect among Chinese stockbrokers. *PLoS ONE*, *9*(2), 1–4. DOI: <https://doi.org/10.1371/journal.pone.0089193>

Peer review comments

The author(s) of this paper chose the Open Review option, and the peer review comments can be downloaded at:
<http://doi.org/10.1525/collabra.224.pr>

How to cite this article: Wolf, D., & Schütz, A. (2019). No Effect of Red on Personality Trait Self-Ratings: Testing for Effects of Font Color. *Collabra: Psychology*, 5(1): 57. DOI: <https://doi.org/10.1525/collabra.224>

Senior Editor: M. Brent Donnellan

Editor: M. Brent Donnellan

Submitted: 08 January 2019

Accepted: 15 November 2019

Published: 27 December 2019

Copyright: © 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



UNIVERSITY
of CALIFORNIA
PRESS

Collabra: Psychology

Collabra: Psychology is a peer-reviewed open access journal published by University of California Press.

OPEN ACCESS