Seeing red? The Effect of Colour on Intelligence Test Performance

Emma Larsson & Sophie von Stumm

Later published as: Larsson, E. & von Stumm, S. (2015). Seeing red? The effect of colour on intelligence test performance. Intelligence, 48, 133-136.

Abstract

A series of recent studies reported that seeing red was associated with poor intelligence test performance in university students. Here we test for the first time the effect of colour on intelligence test scores in an adult sample and across a large battery of ability tests. Overall 200 British adults completed Raven's matrices without colour manipulation (i.e. baseline assessment); afterwards, they then either viewed red or green before completing six additional ability tests (i.e. word fluency, logical reasoning, vocabulary, syllogisms, verbal reasoning, and knowledge) and rating their self-perceived performance for each measure. We found no evidence for an association between colour and intelligence test scores or self-perceived performance, before and after adjusting for intelligence at baseline. The discrepancy with previous findings is likely to be due to testing adult rather than student samples, which in turn has implications for the recruitment and selection of study samples in future intelligence research.

Word count: 150

Keywords: colour; red; intelligence; self-perceived performance

Intelligence test scores result from the interaction between the characteristics of an individual, who is taking the test, and those of the situation, in which the test is administered. Thus, intelligence test scores vary as a function of individual differences in the ability to master an ability tests' items within the testing environment. Intelligence research concentrates traditionally on individuals' ability differences rather than on the properties of assessment settings; as consequence, our knowledge of situational factors that influence intelligence test performance is to date incomplete. Previous studies in this area have focused mainly on the role of auditory and sound environments (Bates & Rock, 2003; Pietschnig, Voracek, & Formann, 2010), but recently visual perceptions of or stimuli in the testing situation have been suggested to affect cognitive task performance. More specifically, the perception of specific colours is thought to influence performance outcomes in achievement-related contexts (Elliot & Maier, 2014). For example, viewing green prior to task performance enhances creative achievement (Lichtenfeld, Elliot, Maier & Pekrun, 2012), while viewing red impairs cognitive task performance, as indicated by lower intelligence test scores (Elliot, Maier, Moller, Friedman & Meinhardt, 2007). In previous studies, the effects of colour on cognitive task performance could not be attributed to individual differences in intelligence at baseline or gender (Elliot et al., 2007; Elliot, Payen, Brisswalter, Cury, & Thayer, 2011), suggesting a substantial relationship between colour and cognitive functioning. Elucidating the latter, Elliot and colleagues (2007) argued that colour communicates a deeper meaning that is either inferred through learned associations or through biological nativism, rather than merely serving aesthetic purposes. Specifically, people may have learnt to associate the colour red with poor intellectual performance outcomes, because teachers traditionally used red ink for marking school and course work (Elliot et al., 2007). Alternatively, the colour red may have an evolutionary history (Mollon, 1989), following Charles Darwin's (1872) observation that anger is associated with redness in the skin. Here, colour is thought to work as a subtle prime or danger warning, affecting psychological functioning, without the individual's conscious awareness (Elliot et al., 2007).

Previous studies on the negative effect of red on cognitive performance have employed various methods of colour manipulation. For example, high school students who were shown a red rectangle (12.7cm x 18.4cm) on the cover of a numeric subtest of the Intelligence Structure Test (IST; Amthauer, Brocke, Liepmann, & Beauducel, 1999) performed significantly worse than those who had viewed a grey rectangle (N = 20 overall; Maier, Elliot & Lichtenfeld, 2008). Furthermore, a study of 71 university students reported that those participants who had viewed their study identification or participant number in red (1.3cm x 1.9cm) performed significantly worse in an anagram test compared to students who had viewed their number in green or black (Elliot et al., 2007). In this study, the participants had been requested to verify that their participant number was written on all pages except the cover page to enforce the effectiveness of the colour manipulation. Even more impressive are the findings of Lichtenfeld et al. (2009) in this context, who showed that only reading the word "red" compared to reading a neutral colour word was significantly associated with lower IST analogy test scores in overall 49 high school students. The authors also showed that even the very small change of including the word “red” (Arial, 10 pt) in the copyright label’s text on the testing booklet’s introductory pages resulted in differential performance on the IST numeric test in overall 40 high school students (Lichtenfeld et al., 2009). Across all these studies, participants were unaware of the colour manipulation and its effects, supporting the theory that colour works as a subtle prime outside the individual’s conscious awareness (Elliot & Maier, 2007).

Notwithstanding, previous studies suffer from three limitations. First, they have only tested school and university students, with overall sample sizes ranging from a minimum of 20 to a maximum of 71 participants *across* all study conditions (Elliot et al., 2007: Elliot et al., 2011; Lichtenfeld et al., 2009; Maier et al., 2008). No comparable data from adult samples have been reported, who may be less sensitive to the effects of red in achievement contexts because they are neither as often nor as recently exposed to feedback from a teacher’s red marking pen. Second, previous studies reported the effect of colour on selected performance measures, mainly abstract reasoning or fluid-type intelligence tests. It is unknown, however, if the performance impairing effect of red is specific to reasoning tests or if it holds also true for more information-, experience- and knowledge-based tests (i.e. crystallized intelligence) that are particularly relevant for typical intellectual performance, which are usual in the work place and educational settings. Third, no previous study has explored the extent to which the colour red may affect participants’ perception of their abilities, as a potential mediating variable. Self-estimated ability to perform in a task is an important determinant of task engagement and persistence (Dweck & Leggett, 1988) and thus, influences cognitive performance outcomes (von Stumm, 2013). Viewing red before completing a task may lower the participants’ expectation of their performance level and as a result, exert an indirect effect on the latter.

Overcoming these limitations, we report here data from 200 British adults, who completed a battery of ability measures and estimated how well they expected to perform in them before and after being exposed to colour manipulation.

Methods

Sample

A sample of 200 British adults (97 men and 103 women) was recruited in the London metropolitan area through online advertisements, flyers in public places (e.g. cafes), and research volunteer databases. Participants met these criteria: a) Native to the United Kingdom and having lived in the United Kingdom since age 10; b) at least 18 years old; and c) not dyslexic or colour-blind. The sample's age ranged from 18 to 69 years with a mean of 34.6 (SD 11.8). As their highest educational qualification, 14% of the participants reported GCSEs; 15% A-levels; 18% a vocational qualification or equivalent; 33.5 % an undergraduate degree and 19% a postgraduate degree. About half of the sample reported to earn less than £15.000 per annum, while about 8% declared to earn more than £35.000 per annum. Data from this sample have been previously reported elsewhere (e.g. von Stumm, 2012).

Measures

Ability tests. *Raven's Progressive Matrices (Set E; Raven, 1968):* Twelve items showed grids of 3 rows x 3 columns each with the lower right hand entry missing. Participants chose from eight alternatives the one that completed the 3 x 3 matrix figure. The test was timed at 4 min. *Verbal fluency (Ektroem et al., 1976):* Participants listed as many words as possible that started with one of two prefixes in

60 seconds. *Letter sets (Ektroem et al., 1976):* Participants identified the mismatching 4-letter set, inferring a rule underlying the composition of four other 4-letter sets. The test had 15 items and was timed at 6 min. *Vocabulary (Ektroem et al., 1976):* Participants identified the correct synonym for a given word out of five answer options. The test had 18 items and was timed at 4 min. The score was corrected for guessing (-.20 for every incorrect answer). *Nonsense syllogisms (Ektroem et al., 1976):* Participants judged if a conclusion that followed two preceding statements (premises) showed good (correct) reasoning or not. The test had 15 items and was timed at 4 min. The score was the number of items marked correctly minus the number marked incorrectly. *Verbal Reasoning (Ektroem et al., 1976):* Participants identified the correct pair of words from a five options to complete a comparison sentence, whose first and last words were missing. The test had 14 items and was timed at 7 min. The score was corrected for guessing (-.20 for every incorrect answer). *Knowledge (von Stumm, 2013)*. Overall 13 knowledge domains, including art, music, literature, geography, politics, history, finance, sports, fashion, film, medicine, science, food and health, and technology, were assessed with overall 132 open-ended questions. Domain scores mapped onto two knowledge factors, one based on popular culture and one on academic disciplines (von Stumm, 2013).

Self-estimated intelligence. (1) A bell curve of IQ scored was shown with a mean of 100 and `3 SD of 15. Participants estimated their IQ with reference to it. An IQ of 55 was labelled as ‘mild retardation’, and IQ of 75 as ‘borderline retardation’, and IQ of 100 as ‘average ability’, an IQ of 115 as ‘higher intellect’, and an IQ of 145 as ‘gifted ability’. (2) For each intelligence test (see above), participants rated on a 1-5 point Likert scale from very poor to very well how well they thought they had done on the test.

Personality (NEO-FFI, Costa & McCrae, 1992). A 60-item measure of personality assessed the Big Five Personality traits, including Neuroticism, Extraversion, Openness to Experience, Agreeableness and Conscientiousness. Each trait was assessed by 12 items on a 5-point Likert scale, ranging from *strongly disagree, disagree, somewhat agree, agree, to strongly agree.* Internal consistencies approximate .80 across samples (Costa & McCrae, 1992).

Procedure

Testing sessions with participant groups of up to 20 took part in lecture theatres at a London university and were supervised by a trained research assistant. Testing booklets were laid out in a fashion that participants sat at least 3 seats away from each other. Each booklet had an individual identification code in the top right corner, consisting of letters and numbers, that was hand-written into a 1.4cm x 4.8cm box, labelled as "Table Number", with a black-ink felt pen (average width 1.5mm). Participants opened the booklet and completed Raven's Matrices, timed by the research assistant. Participants were asked to turn over to the instruction page of the next ability test (i.e. word beginnings), which showed the same identification code on the top hand right corner, now in either red or green (respective ink felt pens with average with of 1.5mm, size of 1.4cm x 4.8cm), depending on the study condition that participants had been randomly allocated to. Participants were asked to check that both codes (i.e. black and red/ green) were identical, which they were in all cases (Figure 1). Participants then completed five ability tests (in order: word beginnings, letter sets, vocabulary, nonsense syllogisms and verbal reasoning), rating after each how well they thought they had done. Afterwards, participants took a 5 minute break, during which the original booklets were collected and replaced with new ones that showed again the identification code on the top right corner in the same colour as before (i.e. red or green) and included the knowledge test. After completing the knowledge test and self-perceived performance, the booklet was exchanged for a new one, again with a respectively coloured identification code in the top right corner of the first page, assessing self-estimated intelligence, personality and demographic background information. Testing sessions lasted overall for two hours, and participants received monetary compensation.

INSERT FIGURE 1 HERE

Statistical analysis

ANCOVA models tested if colour (i.e. red versus green) had an effect on (1) ability test performances, (2) self-rated ability test performance and (3) self-estimated overall intelligence. For all models the first step tested for a main effect of colour. In the second step, participants' Raven's score, which was obtained under the control (i.e. black) condition, was added to the model to adjust for individual differences in ability at baseline, as well as an interaction term between Raven's and colour. In a third step, the models were adjusted for gender and age. In a fourth and final step, the personality dimensions Neuroticism, Openness, and Conscientiousness were added, to control for core dimensions of individual differences that may confound associations between colour and cognitive performance (von Stumm, Chamorro-Premuzic, & Ackerman, 2011).

Results

Out of 200 participants, 13 were excluded from the analysis because the colour manipulation was wrongly implemented (i.e. they were not reminded to compare the first (i.e. black) with the second (i.e. red/ green) identification code. Other missing data points resulted from participants either not completing a measure or not following the correct instructions. None of the participants guessed the working hypothesis of the study.

All study variables were normally distributed. Table 1 shows the means and SD for the study variables across experimental groups (i.e. red and green condition). The ANCOVA models showed no significant effect of colour on any of the dependent variables ($p > .05$ in all cases). This finding did not change after adjusting in stepwise models for the covariates. In summary, the colour manipulation was here not associated with ability test performance, self-rated ability test performance and self-estimated intelligence.

INSERT TABLE 1 HERE

Discussion

In the current study, the colour in which the participants' identification number was written on the testing booklets was neither associated with their intelligence test performances nor with their perception of the latter or their overall self-estimated intelligence. Thus, the negative effect of colour on cognitive performance was not observed here in contrast to previous studies (Elliot et al., 2007: Elliot et al., 2011; Lichtenfeld et al., 2009; Maier et al., 2008; Mehta & Zhu, 2009).

The discrepancy in findings is likely to be due to differences in this and pervious studies' size and nature of samples. Previous reports on colour-intelligence associations were based on small groups of high school and university student samples but here data from 200 British adults were reported, who have seen more time elapse since associating red with negative performance feedback in educational settings than for high school or university students. It is possible then that the negative effect of viewing red on intelligence test performance varies as a function of age or schooling experience and hence, was not detected here. Alternatively, previous studies in this area that were often underpowered may have been affected by Type I errors. By comparison, our study had the power of .80 to detect a mean difference between both groups of a third of a SD. While accept that, despite its power, a sample size of 200 is insufficient to definitely accept or reject the hypothesis of colour-related effects on cognitive test performance, our findings appear to challenge the notion that viewing red works as a subtle prime because of evolutionary reasons - if that was the case, such effects should be detectable in student and adult samples alike. Future research will elucidate if the association between viewing red and cognitive performance is replicable across samples of different natures and sizes.

Of course, it is also possible the colour manipulation we used in this study was not effective, perhaps because the colour displays were not big or refulgent enough. In this context, Elliot and Maier (2014) recently stressed the importance of hue, lightness and chroma when studying the effects of colour on performance, which we did not control for in the current study. However, previous studies in this area also paid little attention to hue, lightness and chroma (see Elliot et al., 2011, for an exception), used even more subtle colour primes (e.g. copyright label alteration; Lichtenfeld et al., 2009), and still observed a negative effect of red on performance. It is important to note that the present study is a close replication of the manipulation performed in Elliot et al.'s (2007) widely cited paper. Here and in Elliot et al.'s work, the colour manipulation consisted of altering the colour of a participant's identification number that was shown on the testing booklets. Thus, differences in hue, lightness and chroma or the colour manipulation in general are unlikely to explain our null-findings. Similarly, one might argue that our order of stimulus presentation prevented observing the expected effects: presenting participants first with a neutral (i.e. black) colour may have "washed out" any effects of subsequent colour displays but it seems unlikely, as our procedure closely resembles those of previous studies (Elliot et al., 2007; Lichtenfeld et al., 2009).

In conclusion, we found no evidence for a negative effect of red on test performance or on the self-estimated ability to perform in such tests. We attribute this finding largely to the nature of our sample, which consisted of adults instead of students, who have been typically investigated in research on colour-intelligence associations. As a result, our report supports the notion that intelligence research based on student samples is of limited validity, even if the study follows a lab-based, quasi-experimental design.

References:

Amthauer, R., Brocke, B., Lipmann, D., & Beauducel, A. (1999). *Intelligenz-Struktur-Test 2000.* Göttingen, Germany: Hogrefe

Bates, T. C., & Rock, A. (2004). Personality and information processing speed: Independent influences on intelligent performance. *Intelligence, 32(*1), 33-46.

Costa, P. T., Jr., & McCrae, R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI): Professional Manual.* Odessa, FL: Psychological Assessment Resources Inc.

Darwin, C. R. 1872. *The expression of the emotions in man and animals*. London: John Murray.

Dweck, C. S., & Leggett, E. L. (1988). A Social-Cognitive Approach to Motivation and Personality. *Psychological Review, 95*(2), 256-273.

Ekström, R. B., French, J. W., Harman, H., & Derman, D. (1976). *Manual for kit of factor-referenced tests.* Princeton, NJ: Educational Testing Service.

Elliot, A. J., & Maier, M. A. (2007). Color and Psychological Functioning. *Current Directions in Psychological Science, 16*(5), 250-254.

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.

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, 136*(1)*,* 154-168.

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.

Lichtenfeld, S., Elliot, A. J., Maier, M. A., & Pekrun, R. (2012). Fertile Green: Green Facilitates Creative Performance. *Personality and Psychology Bulletin, 38*(6), 784-797.

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.

Maier, M. A., Elliot, A. J., & Lichtenfeld, S. (2008). Mediation of the Negative Effect of Red on Intellectual Performance. *Personality and Social Psychology Bulletin, 34,* 1530-1539.

Mehta, R., & Zhu, R. (2009) Blue or Red? Exploring the Effect of Color on Cognitive Task Performance. *Science, 323,* 1226-1229.

Mollon, J. D. (1989). "Tho' she kneel'd in that place where they grow . . .": The uses and origins of primate colour vision. *Journal of Experimental Biology*, 146, 21–38.

Pietschnig, J., Voracek, M., & Formann, A. K. (2010). Mozart effect-Shmozart effect: A meta-analysis. *Intelligence, 38*(3), 314-323.

Raven, J. C. (1968). *Progressive matrices: A perceptual test of intelligence.* London: H. K.

von Stumm, S. (2012). Investment Trait, Activity Engagement, and Age: Independent Effects on Cognitive Ability. *Special Issue, Journal of Aging Research, doi:10.1155/2012/949837.*

von Stumm, S. (2013). Investment Traits and Intelligence in Adulthood. *Journal of Individual Differences, 34*(2), 82-89.

von Stumm, S., Chamorro-Premuzic, T., Ackerman, P. L. (2011). Re-visiting intelligence-personality associations: Vindicating intellectual investment. In T. Chamorro-Premuzic, S. von Stumm, & A. Furnham (eds.), *Handbook of Individual Differences*. Chichester, UK: Wiley-Blackwell.

Table 1

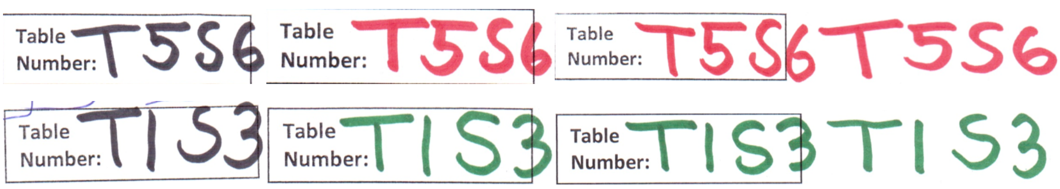
Means and SD for actual and self-perceived intelligence test performance across two experimental groups

	Red			Green				
	N	Mean	SD	N	Mean	SD	*W*	F
Raven Matrices	86	4.07	2.77	101	4.42	3.08	2.98	0.64
Word Beginnings (WB)	85	22.12	7.42	91	22.10	7.77	0.05	0.00
Lettersets (LS)	84	7.07	3.89	101	7.72	3.63	1.08	1.40
Vocabulary (V)	86	9.90	3.74	101	10.25	3.14	1.12	0.50
Nonsense Syll (NS)	86	1.58	3.87	101	1.76	4.44	0.96	0.09
Reasoning (R)	86	8.27	3.91	101	8.16	3.87	0.06	0.03
Knowledge (K)	86	60.15	26.44	101	57.89	26.50	0.28	0.34
Self-rated WB	83	2.76	0.85	92	2.83	0.96	0.16	0.24
Self-rated LS	86	2.92	0.81	98	2.92	0.89	0.08	0.00
Self-rated V	86	2.77	1.01	101	2.80	1.10	0.48	0.05
Self-rated NS	86	3.52	0.90	100	3.37	0.87	0.64	1.38
Self-rated R	86	2.77	0.86	100	2.78	0.94	0.67	0.01
Self-rated K	84	3.44	0.83	98	3.56	0.85	0.18	0.94
Self-est. Intelligence	74	108.46	16.88	90	109.36	13.79	0.65	0.14

Note. *W* refers to the Levene's test of the homogeneity of variance ($p > .05$ in all cases). F refers to the F-test as applied in ANOVA ($p > .05$ in all cases).

Figure 1

Identification or participant code in black and red/ green as used for the study's colour manipulation



Note. Participants first viewed their 'identification code' in black and subsequently in red or green, depending on their experimental condition, across overall four testing booklet covers.