

Color and Categorical Claims

Alex O. Holcombe

School of Psychology, The University of Sydney

The effect of color on psychological functioning is the topic of a large literature. Published claims include that viewing blue causes calmness and that viewing red decreases test achievement. However, almost all these claims are made on the basis of testing just a single, or sometimes a few, hues. But colors like red are categories that comprise many perceptually distinct hues. Making a general claim about red on the basis of testing just one or two red hues may be akin to testing the reliability of one Toyota car and one Tesla car, finding that the Toyota is more reliable, and concluding that Toyotas are more reliable. This methodological issue was omitted from a recent literature review that was otherwise rather comprehensive. This article provides arguments for why this is a major issue and suggests ways to address it.

Keywords: Color, perception, methods, psychology, achievement, mood

In a recent issue of the *Review of General Psychology*, Elliot (2019) reviews our knowledge of the effect of color on psychological functioning. The relevant literature is large. Despite restricting itself to a subset of the topics studied, Elliot's review contains over 260 references. Elliot describes several methodological problems that are widespread in the literature, and concludes his review with a "blueprint for conducting a high-quality study on color and psychological functioning". Researchers in this area would do well to follow Elliot's blueprint. However, one important methodological problem was omitted from Elliot's review and from his guidelines for future studies.

In the reviewed literature on the effect of color, the majority of papers (perhaps the overwhelming majority, but I did not tally them all) test only one or two exemplars of a color category, and most individual experiments test only one. For example, the first experiment in one paper was designed to examine the influence of "perceiving the color red on the force and velocity of motor output" and compared the effects of using one particular kind of red pencil to one particular kind of gray pencil. The second and final experiment of the paper compared, between subjects, the effect of one particular red color to one particular gray and one particular blue (Elliot & Aarts, 2011). Thus, a total of two examples of red

were tested, and contrasted with a gray color in one experiment and a blue color in another.

Red is a color *category*, not a single hue. There are many perceptually distinct red hues, even restricting one's palette to a single brightness and saturation level. Unfortunately, I have not found a single paper in this literature that states its conclusions in terms of the very specific hue or hues tested. I did not do a systematic review of the literature, but it may be noteworthy that neither of the topic experts who reviewed this paper pointed one out. As an example, rather than referring to the specific two reds, two grays, and one blue tested, Elliot & Aarts (2011) concluded that the findings "clearly establish a link between red and basic motor action". This statement, the nature of which is typical in the literature, includes a claim about how the results generalize. The statement implies that the findings will hold for reds in general, not just the particular hue tested.

Before discussing why this claim is dubious, it should be pointed out that the claim is important. First, users of the results, be they designers interested in evoking a particular psychological effect with color, or just scientists interested in probing the finding further, would like to know how closely they should try to match the color used in the original study. Second, the extent to which a result generalizes to other hues and which particular hues it

does generalize to has implications for neural theories of the finding, as will become clear below.

Is it safe to assume that two instances of a color category will have similar psychological and behavioral effects?

Generalizing from a single hue to all the hues of a category may be akin to testing the reliability of one Toyota car and one Tesla car, finding that the Toyota is more reliable, and then concluding that Toyotas in general are more reliable than Teslas. To have the confidence in such a conclusion that would warrant the flat conclusion that Toyotas are more reliable than Teslas, one should test more than one Toyota and more than one Tesla.

Different models of cars made by the same manufacturer are known to differ in their reliability, sometimes markedly, making it clear that one must test more than one model of car. In the case of color, however, perhaps reds are more homogeneous in their effects on psychological functioning than are car models in their reliability. This may well be true, but is there evidence for it, or a good argument for it?

In a review of the [first](#) version of this manuscript, Lakens (2019) provided such an argument. He suggests that there can be “strong theoretical reason to assume slightly different hues and chromas will not matter (because as long as a color is recognized as ‘red’ it will activate specific associations)”. The notion is that if exposure to red impairs performance on a test (Elliot et al. 2007), or boosts the magnitude of force exhibited in a physical task (e.g. Elliot & Aarts, 2011), or red clothing increases the sexual attractiveness of men (Elliot et al., 2010), then this occurs via activating a concept of “red”, one that is activated by all reds. But whether the categories attached to concepts are as broad as red or blue is not clear.

At least one study provides evidence that concepts that elicits color preferences can be more specific than the broad category of red. Schloss, Poggesi, & Palmer (2011) compared the color preferences of Stanford University students to University of California Berkeley students. They tested forty colors, presented one at a time, and found that Stanford students preferred the red of the Stanford logo more than did the Berkeley students. The issue of

generalization to other reds was not directly addressed in the original article but Schloss (Schloss, personal communication) informed me that the data show a significant interaction involving another shade of red, suggesting that Stanford students liked Stanford red more than another dark red, while Berkeley students had little to no preference for the Stanford red compared to the other dark red.

The expectation that an effect found for a particular red will generalize to the category of red is not only based on the questionable assumption that concepts are tied to the familiar broad color categories. It is also based on what one might term “the perceptual assumption.” This is the assumption that the same representations that underlie the conscious perception of color are the ones that drive the psychological and behavioral effects of that color. This is not a safe assumption due to the existence of other visual pathways to the brain.

Two lights that evoke the same conscious color experience can affect the brain in different ways. In addition to the three photoreceptor classes whose photopigment activations give rise to the conscious experience of color, the retina contains another photopigment called melanopsin. The cells that contain melanopsin are involved in synchronising our internal circadian clock to the external light-dark cycle given by the Earth’s rotation. Melanopsin stimulation may thus affect arousal – and arousal is a frequent topic of study in the literature reviewed by Elliot (2019).

There are different spectra (combinations of wavelengths, such as the light that might reflect from a given object in a given illuminant) that stimulate the three cones in the same way, but stimulate melanopsin differently. In other words, two different spectra that both look “red” may activate the same “red” concept and thus activate the same *cognitive* associations, but activate melanopsin quite differently. Indeed, two spectra that appear identical to humans (“metamers”) may stimulate melanopsin rather differently. This raises the possibility that differences in the effects of red and blue on arousal may not reflect the conscious categories of blue and red, but rather be caused by different levels of stimulation of melanopsin. This possibility can be empirically tested, for example with the “silent substitution” technique, pairs of lights are used that dif-

ferently activate melanopsin (or another photopigment) while keeping the activation of all others constant (Spitschan & Woelders, 2018).

Excitation of melanopsin is not the only way that two different spectra of similar or identical appearance can elicit different responses in the brain. Many cells of the superior colliculus, a midbrain structure, are driven strongly by visual input via a direct pathway from the retina. Unlike the circuits that underlie color experience, based on evidence from non-human primates, this pathway does not seem to include retinal cells that carry signals from S (short-wavelength) cones (Schiller & Malpeli, 1977; de Monasterio, 1978). As a result, two lights that appear identical or near-identical to humans may have substantially different effects on collicular neurons. These neurons are known to be involved in shifting attention, and some evidence suggests they are involved in distractibility (Gaymard et al., 2003) and attention deficit hyperactivity disorder (Brace et al., 2015; Clements et al., 2014). In addition, there is evidence that the superior colliculus provides visual signals for processing of emotions by the amygdala (Rafal et al., 2015).

It should now be clear that it is not safe to assume that two different colors affect behavior entirely as a result of the color experience they evoke. The collicular pathway is the dominant visual pathway in most mammals, and while it is less important in primates, it is not unlikely that it has psychological effects in humans.

How should color be statistically modeled?

With infinite resources, one could test all possible reds and determine whether the same result is found for each one. With finite resources one must test a limited number of reds and make some assumptions to generalize to the others. Generalizing to a group from a sample, the problem of induction, is the subject of an enormous literature in both philosophy and statistics. It is perhaps most familiar to psychologists in the context of sampling human participants – the purpose of most psychology studies is to make a claim not solely about the individual persons tested, but rather to generalize to a population. Psychologists customarily sample participants from a population and often enter them into statistical models as a random factor. The reason is that a random factor, as Judd, Westfall, & Kenny

(2012) put it, is one “whose levels are sampled from some larger population of levels across which the researcher wishes to generalize, whereas fixed factors are those whose levels are exhaustive” (see also Clark, 1973). The same strategy could be used to generalize to the category red, if multiple examples of the category red were tested.

There is an important difference, however, between sampling reds and sampling people which suggests that treating hue as a random factor may not be very appropriate. For red, one aspect of the similarity of red hues is fairly well understood – perceptual similarity (although color similarity is not simple, and certainly not fully understood – see Witzel & Gegenfurtner, 2018). Capitalizing on knowledge of perceptual similarity, to create some confidence in generalizing to the entire category of reds, one can choose a set of reds evenly spaced in perceptual color space along the range of possible reds. The effect of hue might then be modelled by linear regression or a polynomial. This is only valid, however, if perceptual representations and thus perceptual similarity mediate the effect of color. A non-perceptual representation such as those involving melanopsin or the superior colliculus might instead mediate the effect of color.

Our understanding of the melanopsin and superior colliculus pathways, although limited, may be sufficient to provide non-perceptual similarity metrics that can be used to guide alternative models to fit to the data. Model comparison can then indicate which pathway is most likely to underlie the investigated effect.

So far we have only discussed the issue of generalizing within a color category. An additional issue is the need to justify the *specificity* aspect of the conclusions typical of the color psychology literature. Elliot & Aarts (2011), for example, concluded that there was a link between red and motor action, although the only other colors they tested, were a shade of gray and blue. Thus there was no evidence in the study to address whether the link is specific to red, rather than extending to green, yellow, purple, and brown as well. The solution is, of course, to test those colors.

Moving forward

It is easy to suggest that multiple color stimuli should be tested. It is less easy to actually perform such a study – it may be very expensive to conduct a well-powered study that investigates the effect of several hues. Elliot (2018) suggests that studies should be adequately powered to detect an effect size of $d = 0.35$, with a sample size of 130 participants per condition for between-participants designs. In fact many of the topics of interest in this literature require a between-participant design, with each participant being exposed to a single color to examine their subsequent performance on e.g. a test. Thus, to test five different stimuli would require six hundred fifty participants, and adequate control of the color each participant is exposed to typically requires testing them in the lab rather than online. Thus, a study that adequately justifies the claims that are frequently issued in this literature would be quite expensive.

As Elliot (2019) explains in his review paper, very little can currently be concluded about the effect of color on psychological functioning. This is a truly dismal outcome for a century's-worth of research. In the case of the popular hypothesis that red results in excitement or stimulation, for example, supportive evidence was found in approximately forty different studies, but nearly twenty studies did not find that link – and that is only the published studies. In psychology, publication bias is rife (Ferguson & Brannick, 2012; Ferguson & Heene, 2012), raising the possibility that many more studies were conducted and found evidence of no difference. To address this as well as to reduce the analytical flexibility that yields many false positives, preregistration is an absolute must for this literature going forward (Nosek et al., 2018), as are the color-related methodological prescriptions laid out by Elliot (2019).

The problem pointed out in this commentary of unjustified generalizations made from testing a single hue may be a substantial contributor to the weak replication record of the literature. Perhaps as a result of the unwarranted generalizations made in the literature, typically studies do not use the same exact hue as any previous study, and at this point there is little idea whether this accounts for discrepancies

in results. Given the existence of this as well as the multiple other methodological issues detailed by Elliot (2019), researchers in the area have a lot of work ahead of them if the literature is to provide reliable and useful results.

Open Science Practices

This article contained no relevant data, materials or analysis to be shared. The entire editorial process, including the open reviews, are published in the online supplement.

Acknowledgments

I thank all five reviewers, whose comments led to major changes and additions to this work (see the associated peer report). I also thank Manuel Spitschan for comments and Geoffrey Aguirre for discussion

Declaration of Interests

The author declares no competing interests.

References

- Brace, L. R., Kraev, I., Rostron, C. L., Stewart, M. G., Overton, P. G., & Dommett, E. J. (2015). Altered visual processing in a rodent model of Attention-Deficit Hyperactivity Disorder. *Neuroscience*, 303, 364–377. <https://doi.org/10.1016/j.neuroscience.2015.07.003>
- Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12(4), 335–359. [https://doi.org/10.1016/S0022-5371\(73\)80014-3](https://doi.org/10.1016/S0022-5371(73)80014-3).
- Clements, K. M., Devonshire, I. M., Reynolds, J. N. J., & Overton, P. G. (2014). Enhanced visual responses in the superior colliculus in an animal model of attention-deficit hyperactivity disorder and their suppression by d-amphetamine. *Neuroscience*, 274, 289–298.

- <https://doi.org/10.1016/j.neuroscience.2014.05.054>
- Elliot, A. J. (2019). A Historically Based Review of Empirical Work on Color and Psychological Functioning: Content, Methods, and Recommendations for Future Research. *Review of General Psychology*, gpr0000170. <https://doi.org/10.1037/gpr0000170>
- Elliot, A. J., & Aarts, H. (2011). Perception of the color red enhances the force and velocity of motor output. *Emotion*, 11(2), 445–449. <https://doi.org/10.1037/a0022599>
- Ferguson, C. J., & Brannick, M. T. (2012). Publication bias in psychological science: prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods*, 17(1), 120.
- Ferguson, C. J., & Heene, M. (2012). A Vast Graveyard of Undead Theories: Publication Bias and Psychological Science's Aversion to the Null. *Perspectives on Psychological Science*, 7(6), 555–561. <https://doi.org/10.1177/1745691612459059>
- Gaymard, B., François, C., Ploner, C. J., Condy, C., & Rivaud-Péchoix, S. (2003). A direct prefrontotectal tract against distractibility in the human brain: Prefrontotectal Tract. *Annals of Neurology*, 53(4), 542–545. <https://doi.org/10.1002/ana.10560>
- Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, 103(1), 54–69. <https://doi.org/10.1037/a0028347>
- de Monasterio, F.M. (1978). Properties of ganglion cells with atypical receptive-field organisation in the retina of macaques. *J. Neurophysiol.* 41, 1435–1449.
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11), 2600–2606.
- Rafal, R. D., Koller, K., Bultitude, J. H., Mullins, P., Ward, R., Mitchell, A. S., & Bell, A. H. (2015). Connectivity between the superior colliculus and the amygdala in humans and macaque monkeys: virtual dissection with probabilistic DTI tractography. *Journal of Neurophysiology*, 114(3), 1947–1962. <https://doi.org/10.1152/jn.01016.2014>
- Schiller, P.H., and Malpeli, J.G. (1977). Properties and tectal projections of monkey retinal ganglion cells. *J. Neurophysiol.* 40, 428–445.
- Schloss, K. B., Poggesi, R. M., & Palmer, S. E. (2011). Effects of university affiliation and “school spirit” on color preferences: Berkeley versus Stanford. *Psychonomic Bulletin & Review*, 18(3), 498–504. <https://doi.org/10.3758/s13423-011-0073-1>
- Spitschan, M., & Woelders, T. (2018). The Method of Silent Substitution for Examining Melanopsin Contributions to Pupil Control. *Frontiers in Neurology*, 9. <https://doi.org/10.3389/fneur.2018.00941>
- Witzel, C., & Gegenfurtner, K. R. (2018). Are red, yellow, green, and blue perceptual categories? *Vision Research*, 151, 152–163. <https://doi.org/10.1016/j.visres.2018.04.002>