

Effects of stimulus range on color categorization

Oliver Wright
Bilkent University, Turkey

This chapter reports three experiments detailing the influences of stimulus range on color categorization. The results show that both categorization and speed of categorization of colorful stimuli can be influenced by stimulus range. Two potential consequences are considered. First, the influences of stimulus range on color categorization can help explain inconsistencies in the color literature relating to the reliability of color categorization. Second, the same influences appear relevant to the interpretation of experiments investigating color categorical perception.

1. Introduction: Color categorization and range effects

This chapter reports three experiments addressing the issue of whether (and to what extent) changed range contexts influence judgments about color. A range effect would be shown if, for instance, a stimulus categorized most often as 'blue' when embedded in one stimulus range is categorized as 'green' most often when embedded in a different stimulus range. Differences in the time taken to categorize a stimulus contingent on its location in a stimulus range would be another kind of range effect.

Little work has been done investigating the influence of range on color judgments. Mitterer and de Ruiter (2008) provide data suggesting that range effects do not influence color categorization. This is somewhat surprising since numerous studies show other kinds of psychological judgments are influenced by stimulus range (e.g. Crawford, Huttenlocker & Engebretson 2000; Hollingworth 1910; Parducci 1965; Petzold 1982; Petzold & Haubensack 2004; Watson 1957).

Stimulus range might influence performance in several different ways. What Hollingworth (1910: 462) called the "central tendency of judgment" suggests that stimulus estimates are biased towards the center of a presented range. A related idea, proposed by Parducci (1965) and Petzold (1982), is that participant reference scales are adjusted to the endpoints of the subjective range. Poulton too (1973, 1974, 1989)

has given detailed consideration of range effects and their influence in within-subjects experimental designs.

In studies that show influences of stimulus range on categorical judgments, the categories used have usually been those for which participants must establish their own standards, such as size judgments (Parducci 1965; Petzold & Haubensak 2004). It is possible therefore that the lack of a range effect for color categorization reported by Mitterer and de Ruiter (2008) stems from differences between color categories and other categories, such as size, which have a more arbitrary basis.

Another reason to be interested in the possible effects of range on color categorization relates to a lack of clarity concerning the reliability of color categorization and in particular the 'sharpness' of category boundaries. Berlin and Kay (1969: 13) suggested that:

Category boundaries are not reliable, even for repeated trials for the same informant... In fact, in marked contrast to the foci, category boundaries proved to be so unreliable, even for an individual informant, that they have been accorded a relatively minor place in the analysis.

Rosch Heider (1972) makes a similar point, classing stimuli located away from focal regions of color space 'internominal', whilst Kay and McDaniel (1978) proposed that color categories be considered fuzzy sets.

On the other hand, much evidence suggests that color space is categorized more cleanly than the previous paragraph suggests (e.g. Boynton & Gordon 1965; and particularly Bornstein & Korda 1984). Similarly, Malkoç, Kay and Webster (2005) found that English speakers identified a boundary between blue and green categories more consistently than the best example of either color.

Such differences may be partially explicable in terms of differences in response categories available to participants and also by differences in the lightness and saturation of stimuli used in differing experiments. But variations in stimulus range used in differing studies of categorization might also help explain the different results obtained. Such an explanation, though, would only be valid if stimulus range can be shown to influence categorization.

Finally, the potential influence of stimulus range on color categorization relates to color categorical perception (CP). CP is the idea that, all other things being equal, discriminations between stimulus pairs belonging to different categories are made more reliably and/or faster than discriminations made between pairs belonging to the same category. Numerous studies have combined colorful stimuli and tasks such as target detection and visual search (e.g. Drivonikou, Kay, Regier, Ivry, Gilbert, Franklin & Davies 2007; Gilbert, Regier, Kay & Ivry 2006; Roberson & Davidoff 2000). Results show that, although performance of these tasks would not appear to require explicit categorization of the stimuli used, categorization nevertheless seems to influence performance. Better understanding of the factors which may influence color categorization should lead to a deeper understanding of color CP. This, in turn, has a bearing on

theories of linguistic relativity (e.g. Whorf 1956), which suggest that differences in language lead to differences in thought and perception.

2. Experiments 1a, b and c

Participants in the three experiments reported here were native Turkish speakers. The experiments were 2-alternative-forced-choice tasks (2-AFC). In such experiments participants are required to decide which of two descriptors, such as 'green' or 'blue', most accurately describes a given stimulus. To manipulate stimulus range, each experiment used two conditions. Some stimuli were present in both conditions, others in only one or other of the conditions.

The three experiments differed primarily in the region of color space from which stimuli were drawn. Stimuli in Experiments 1a and b came from the region of color space named '*mavi*' and '*yesil*' by Turkish speakers. These correspond to English terms 'green' and 'blue' (Özgen & Davies, 1998). Stimuli in experiment 1c came from regions of color space named '*mavi*' and '*mor*', corresponding to English term 'purple' (Özgen & Davies 1998). For the sake of clarity, English color terms will be used when referring to stimuli used in the experiments.

2.1 Participants

There were a total of 131 participants divided almost equally among the three experiments. All had normal color vision, tested using pseudoisochromatic plates (Ishihara 2003). Table 1 gives details of the participant who took part in each of the experiments.

2.2 Experiment 1a: Apparatus, stimuli and design

The experiment was run using a personal computer and a sixteen inch LG 710S CRT monitor. Display was controlled using a GeForce 6200 LE graphics controller. Colorimetric measurements were made using a Cambridge Research Systems colorCAL.

Table 1. Response categories and participant details for experiments 1a, b and c. (Abbreviated column headings: Exp. = Experiment)

Exp.	Response Type: 1/2	Participants		
		Number in each condition	Mean age (S.D)	Sex
1a	green/blue	45 (C1 = 23, C2 = 22)	21 y. 9m. (33m.)	19F, 26M
1b	green/blue	45 (C1= 22, C2 = 23)	22y. 4m. (34 m.)	24F, 20M
1c	blue/purple	46 (23 in each condition)	21y. 1m. (37 m.)	26M, 20F

Conversions between CIE and Munsell color spaces were made using tables provided by Wyszecki and Styles (1982).

Stimuli were nine Munsell defined colors varying only in hue. Hues ranged from 7.5G to 7.5B in steps of 2.5 hue units and spanned the green–blue region of color space. Stimuli were of value 7 and chroma 8. Stimulus 7.5G was the greenest-appearing stimuli, 7.5B the bluest. Previous research has indicated that 7.5BG marks the boundary between blue and green categories (Bornstein & Korda 1984). Stimuli were presented as 36 mm squares in the center of the display corresponding to a visual angle of 4.1° from the viewing distance of 500 mm. Stimuli were shown singly against a gray background of identical luminance. Following the experiment the colorimetric properties of stimuli were rechecked. These measurements confirmed stimuli properties had remained stable.

2.2.1 *Experiment 1a: Procedure*

There were two conditions: 1 and 2. Each condition used seven stimuli. In condition 1 the stimuli were 7.5G, 10G, *2.5BG (1)*, *5BG (2)*, *7.5BG (3)*, *10BG (4)* and *2.5B (5)*. In condition 2 stimuli were *2.5BG (1)*, *5BG (2)*, *7.5BG (3)*, *10BG (4)*, *2.5B (5)*, 5B and 7.5B. Numbers in parentheses indicate labeling of stimuli in Tables 2 and 3. The five italicized stimuli were common to both conditions. Thus, in addition to the five stimuli common to both conditions, participants in condition 1 categorized two extra greener stimuli, those in condition 2 categorized two extra bluer stimuli.

Stimuli were presented for categorization singly and in a random order seven times each, making a total of forty-nine trials. Participants were instructed to categorize the stimuli as either ‘green’ (type 1 response), or ‘blue’ (type 2 response) using the left or right mouse buttons. Response keys were counterbalanced across participants. During stimulus presentation, two prompts, in size eighteen black font, appeared on the top left and right of display, indicating which mouse button corresponded to which response. Stimuli remained displayed until a response was recorded. A two second unfilled interval followed, after which the next stimulus was displayed. Participants sat approximately 500 mm from and at right-angles to the display in a small dark room. Participants were instructed to perform the task as rapidly and accurately as possible. For each trial, response type (1 or 2) and response time were recorded.

2.3 Experiments 1b and 1c: Apparatus, stimuli, design and procedure

Experiments 1b and 1c differed from experiment 1a only in the stimuli used. In Experiment 1b, stimuli were of the same hue as those in Experiment 1a, but of lower value (6) and saturation (6). In Experiment 1c, stimuli were of the same value and saturation as in Experiment 1a, but of different hues: 7.5B, 10B, *2.5PB (1)*, *5PB (2)*, *7.5PB (3)*, *10PB (4)*, *2.5P (5)*, 5P and 7.5P. Numbers in parentheses indicate labeling of stimuli in Tables 2 and 3. The five italicized stimuli were common to both conditions. Stimuli 7.5B and 10B were present in condition 1 only; stimuli 5P and 7.5P were

present in condition 2 only. Previous research suggests that the boundary between blue and purple categories lies around 7.5PB (Franklin & Davies 2004) or 10PB (Roberson, Davidoff & Braisby 1999).

In Experiment 1b, the forced choice response categories were the same as those used in Experiment 1a. In Experiment 1c, the response categories were 'blue' (type 1) and 'purple' (type 2), as shown in Table 1.

2.4 Results of experiments 1a, 1b and 1c

No participants reported being unable to categorize stimuli using the categories provided.

The analyses below are confined to data produced by the five stimuli that occurred in both conditions within each experiment, labeled 1 to 5 in Tables 2 and 3. Stimulus 3 represents the centrally located stimulus within the entire stimulus range used in each experiment. In Experiments 1a and 1b, this is the stimulus with hue of 7.5BG; in Experiment 1c, the corresponding stimulus hue is 7.5PB.

Table 2. Proportion type 2 responses made in Experiments 1a, b and c for stimuli 1 to 5. For each experiment, the average proportion of type 2 responses made by participants in conditions 1 and 2 is shown, as is the combined responses of participants across both conditions. The lowest three rows give data combined across all experiments, for conditions 1 and 2 separately (corresponding to Figure 1) and, in the bottom row, combined across all experiments and conditions. Standard deviations are in parentheses. (Abbreviated column headings: Exp. = experiment, Cond. = condition, Comb. = combined.)

Exp.	Cond.	Stimulus					
		1	2	3	4	5	All
a	1	0.03 (0.09)	0.29 (0.31)	0.84 (0.18)	0.97 (0.12)	0.97 (0.10)	0.62 (0.43)
	2	0.02 (0.06)	0.02 (0.09)	0.30 (0.36)	0.83 (0.21)	0.99 (0.03)	0.43 (0.46)
	Comb.	0.02 (0.08)	0.15 (0.26)	0.57 (0.40)	0.90 (0.23)	0.98 (0.07)	0.53 (0.45)
b	1	0.02 (0.05)	0.42 (0.34)	0.85 (0.19)	0.99 (0.04)	1.00 (0.00)	0.65 (0.42)
	2	0.02 (0.06)	0.06 (0.13)	0.29 (0.32)	0.79 (0.32)	0.96 (0.07)	0.42 (0.44)
	Comb.	0.02 (0.05)	0.23 (0.31)	0.56 (0.39)	0.89 (0.25)	0.98 (0.06)	0.54 (0.45)
c	1	0.00 (0.03)	0.06 (0.13)	0.69 (0.39)	0.94 (0.13)	0.98 (0.05)	0.42 (0.45)
	2	0.02 (0.05)	0.04 (0.06)	0.27 (0.34)	0.81 (0.34)	0.97 (0.09)	0.54 (0.47)
	Comb.	0.01 (0.04)	0.05 (0.10)	0.48 (0.42)	0.87 (0.27)	0.98 (0.08)	0.48 (0.46)
All	1	0.02 (0.07)	0.26 (0.31)	0.80 (0.26)	0.96 (0.11)	0.98 (0.07)	0.60 (0.45)
	2	0.02 (0.05)	0.04 (0.10)	0.29 (0.33)	0.81 (0.33)	0.98 (0.09)	0.43 (0.45)
	Comb.	0.02 (0.06)	0.14 (0.25)	0.48 (0.40)	0.89 (0.25)	0.98 (0.07)	0.51 (0.45)

Table 3. Response times (in milliseconds) made in Experiments 1a, b and c for stimuli 1 to 5. For each experiment, average participant response times in conditions 1 and 2 are shown separately, as are combined response times across both conditions. The lowest three rows give data combined across all experiments, for conditions 1 and 2 separately (corresponding to Figure 2) and, in the bottom row, combined across all experiments and conditions. Standard deviations are in parentheses. (Abbreviated column headings: Exp. = experiment, Cond. = condition, Comb. = combined)

Exp.	Cond.	Stimulus					
		1	2	3	4	5	All
a	1	844 (538)	1435 (865)	1292 (658)	760 (354)	717 (293)	965 (861)
	2	670 (105)	691 (217)	1189 (1292)	1462 (1252)	812 (436)	1010 (842)
	Comb.	755 (389)	1055 (722)	1239 (985)	1119 (984)	765 (373)	987 (761)
b	1	724 (231)	1331 (768)	1092 (1079)	700 (189)	693 (186)	950 (568)
	2	686 (122)	820 (517)	1339 (879)	1059 (456)	844 (370)	908 (656)
	Comb.	705 (183)	1069 (694)	1218 (979)	883 (392)	771 (301)	929 (619)
c	1	687 (162)	1203 (545)	1109 (672)	840 (448)	712 (125)	859 (526)
	2	645 (118)	650 (112)	1133 (850)	1085 (648)	780 (181)	911 (568)
	Comb.	666 (141)	915 (694)	1122 (761)	968 (568)	748 (159)	884 (554)
All	1	752 (352)	1323 (847)	1164 (819)	767 (346)	708 (210)	942 (628)
	2	667 (115)	719 (331)	1219 (990)	1200 (860)	812 (340)	923 (667)
	Comb.	708 (261)	1012 (702)	1192 (908)	990 (695)	761 (288)	933 (647)

For each participant, for each stimulus, the proportion of type 2 responses ('blue' for Experiments 1a and b, 'purple' for Experiment 1c) was calculated. Data generated were combined across participants within each experimental condition as shown in Table 2. Similarly, for each participant, median response times to each stimulus were calculated and then combined across participants within each experimental condition, as shown in Table 3. Figure 1 shows, combined across experiments, the mean number of type 2 responses made in separate conditions. Figure 2 gives corresponding data for response times. Examination of these figures suggests that, overall, there are different patterns of performance between conditions.

Data for categorical response times and categorical responses were analyzed separately in 2 (Condition: 1, 2) X 3 (Experiment: 1a, 1b, 1c) X 5 (Stimulus: 1, 2, 3, 4, 5) mixed ANOVAs. Bonferroni corrected *t*-tests were used for subsequent comparisons. The analyses below report only significant main effects and interactions.

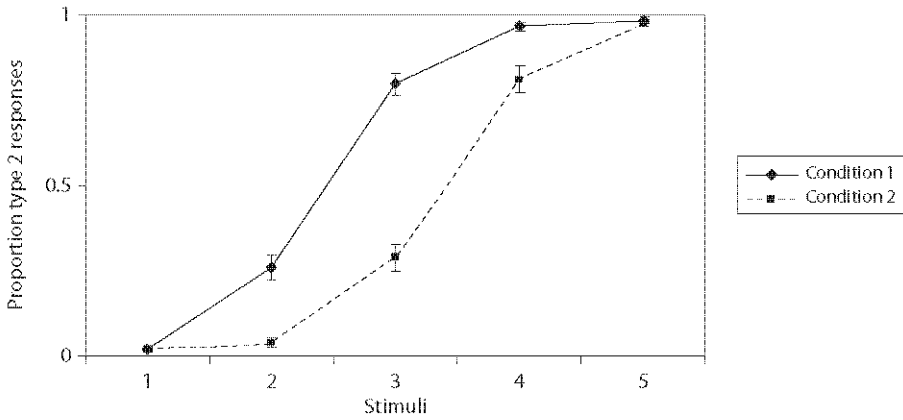


Figure 1. Combined data from experiment 1a, b and c showing proportion of type 2 ('blue' in experiments 1a and b, 'purple' in experiment 1c) responses (± 1 S.E.) in conditions 1 and 2. Data correspond to the numbers in bold in Table 2

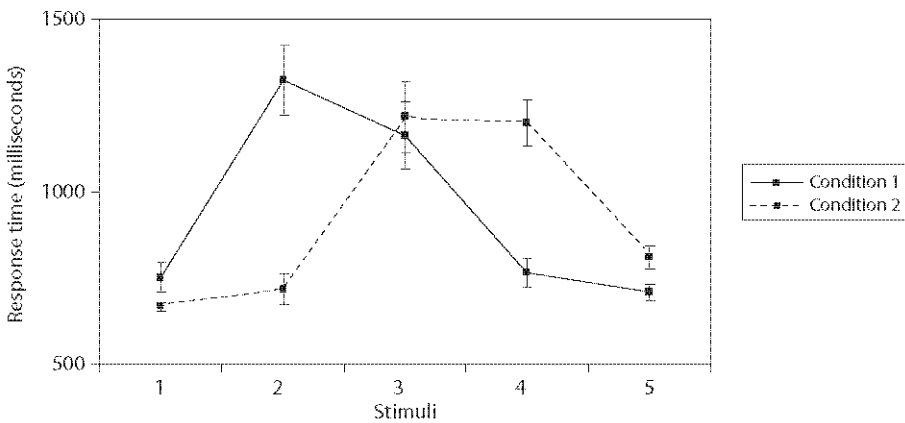


Figure 2. Combined data from Experiment 1a, b and c showing mean response times (± 1 S.E.) in conditions 1 and 2. Data correspond to the numbers in bold in Table 3

3. Categorical responses

3.1 Main effects

All the main effects were significant. A main effect of condition, $F(1, 130) = 84.49$, $MSE = 262.83$, $p < 0.001$, indicates participants in condition 2 gave significantly fewer type 2 responses than participants in condition 1. The main effect of experiment,

$F(2, 130) = 3.73$, $MSE = 11.59$, $p < 0.05$, indicates participants in Experiment 1c gave fewer type 2 responses than participants in Experiments 1a or 1b, min. $t(89) = 3.21$, max. $p < 0.05$. The main effect of stimulus, $F(4, 520) = 717.45$, $MSE = 1222.78$, $p < 0.001$, indicates significantly different response patterns between all stimuli pairs, min. $t(135) = 3.74$, max. $p < 0.01$, with stimulus 2 receiving more type 2 responses than stimulus 1, stimulus 3 receiving more type 2 responses than stimulus 2, and so on.

3.2 Condition by stimulus interaction.

A significant two-way interaction was found between condition and stimulus, $F(4, 520) = 42.16$, $MSE = 71.85$, $p < 0.001$. This reflects differences in responses to stimuli 2, 3 and 4 in different conditions, min. $t(134) = 3.86$, max. $p < 0.001$, with fewer type 2 responses to those stimuli being made by participants in condition 2.

3.3 Experiment by stimulus interaction.

A significant interaction found between experiment and stimulus, $F(4, 520) = 2.23$, $MSE = 3.80$, $p < 0.025$, appears due to stimulus 2 receiving type 2 responses less frequently in Experiment 1c than in either Experiment 1a or 1b, min $t(89) = 1.34$, max. $p < 0.001$.

4. Response times

4.1 Main effect

There was a significant main effect of stimulus, $F(4, 520) = 17.85$, $MSE = 5358618$, $p < 0.001$, with significant differences in response times between all stimuli, min. $t(135) = 2.83$, max. $p < 0.05$, except for the following stimulus pairs: 2 and 3, 2 and 4, and 1 and 5. For the significant differences, response times were always higher for stimuli closer to the center of the overall range used in the experiments, that is closer to the stimulus numbered 3 in Table 3.

4.2 Condition by stimulus interaction.

A significant two-way interaction between condition and stimulus, $F(4, 520) = 16.20$, $MSE = 4861948$, $p < 0.001$, reflects differences in response to stimuli 2 and 4 in different conditions, min. $t(134) = 4.18$, max. $p < 0.01$. Responses to stimulus 2 were slower in condition 1 (where that stimulus is closer to the center of the stimulus range), with the opposite pattern for stimulus 4.

5. Discussion

5.1 Summary of experiments and results

Experiments 1a, b and c were forced choice categorization experiments involving color stimuli. Each experiment involved two conditions, with participants in both conditions categorizing seven stimuli seven times each. In each experiment five stimuli were common to both conditions, the other two were unique to each condition.

Comparison of the stimuli common to both conditions in each experiment revealed differences between conditions in the patterns of categorization, visible in Figure 1. In Experiment 1a, for example, participants in condition 1 categorized the five stimuli common to both conditions as 'blue' more frequently than participants in condition 2. This difference in performance reflects the influence of the two extra greenish colored stimuli in the stimulus set named by participants in condition 1.

Response times to individual stimuli also differed depending on condition, as shown in Figure 2. In general, responses took longer for stimuli located closer to the center of each condition's stimulus range.

5.2 Stimulus range and color categorization

Stimulus range influences performance of forced choice color categorization experiments of the kind described here. The effects were found to be broadly similar in three differing regions of color space. Stimulus range influences not only the frequency with which particular categorical responses are made, but also the time required to produce responses. These influences manifest themselves over the course of only forty-nine trials. It seems probable that in experiments involving larger numbers of trials, range effects would be larger.

However, there are also significant differences between response patterns made in experiments 1a, b and c. That such differences exist shows, unsurprisingly, that task performance is influenced by factors besides stimulus range.

In the introduction, three reasons for investigating the effects of stimulus range on color categorization were mentioned. One concerned the failure of a previous experiment (Mitterer & de Ruyter 2008) to find such an effect. The second reason relates to an apparent paradox in the literature on color categorization. Whilst some researchers have concluded that the divisions between color categories are clear cut (e.g. Bornstein & Korda 1984), others (e.g. Berlin & Kay 1969) have drawn different conclusions. Finally, it was suggested that understanding the influences of range on color categorization can help us understand more about color CP.

There are several potential explanations for the differences between the results of Mitterer and de Ruyter's (2008) study and this one. As is mentioned above, differences in performance between the three experiments described here suggest that factors besides stimulus range influence categorization. Mitterer and de Ruyter (2008) used a

wider range of stimuli. Their stimulus range also included more prototypical stimuli – that is, stimuli considered to be the best examples of the categories they represented. These factors may be sufficient to eliminate the effects of stimulus range on categorization. Further experiments could determine the issue. However the results reported here confirm that psychological judgments about color are, like many other kinds of psychological judgments, influenced by stimulus range.

Influences of stimulus range can also help explain disagreements over the categorization of regions of color space away from category centers. The results of experiments described here seem to show that, in a 2-AFC task, changing the location of stimuli within a range can influence the consistency with which they are categorized.

If during a forced-choice-categorization experiment, the suspected location of a category boundary is located at the center of the stimulus range, then range effects are likely to influence the results. In particular, if the boundary stimulus is the one at the center of the stimulus range, then it is likely that range effects will enhance the apparent sharpness of a category boundary. Alternatively, range effects might also lead to a particular stimulus being categorized less consistently than others, and hence identified as the stimulus closest to the boundary. Using a different range, a different stimulus may likewise be identified as that closest to the boundary. Thus a stimulus identified as being closest to a boundary may only be identified as such in virtue of the context in which it is presented. Range effects may also mask individual differences in color categorization.

Consequently, disagreements over the categorization of colors located away from category centers may be partially explicable in terms of differences in the range of stimuli used by different researchers, as well as differences in experimental techniques (especially multiple versus single trials involving individual stimuli) which influence participant assessment of stimulus range.

Finally, studies of color CP often include a 2-AFC categorization task prior to a main task assessing CP (e.g. Gilbert et al. 2006; Roberson & Davidoff 2000), using a stimulus range that straddles the suspected category boundary. Some potential consequences of this procedure have just been mentioned. In the experimental phase, something similar happens, with participants being exposed to equal numbers of stimuli belonging to each of the previously established categories.

If, as is claimed (Gilbert et al. 2006; Roberson & Davidoff 2000), CP is underpinned by active classification of stimuli, then in research using this methodology, the strength of CP might be enhanced by range effects – because stimuli either side of the center of the range are categorized less ambiguously than if they are located at the center of the stimulus range.

On the other hand, time taken to categorize stimuli increases as stimuli are located closer to the center of the range. If CP is underpinned by classification of stimuli then range effects might be expected to reduce the effect of CP in situations where the category boundary is located at the center of the stimulus range and the experimental measure is the speed of discrimination. The reason is as follows. A CP effect is found

when discriminations between stimulus pairs belonging to different categories are made more reliably and/or faster than discriminations made between pairs belonging to the same category. But if stimuli comprising the across-category trials are drawn from close to the center of the stimulus range, as is usually the case, categorizing these stimuli should take longer than if they were located towards the ends of the stimulus range. Conversely within-category stimulus pairs are usually drawn from the ends of the stimulus range, locations where categorization is faster.

To investigate, experiments assessing CP could be carried out involving two conditions, one in which the suspected category boundary is located at the center of the stimulus range used, another in which the boundary is located away from the center of the stimulus range. Comparison of the results in each condition would aid assessment of the influence of range effects on color CP. This in turn can further understanding of the mechanisms driving CP.

References

- Berlin, Brent & Paul Kay. 1969. *Basic Color Terms: Their universality and evolution*. Berkeley: University of California Press.
- Bornstein, M. H. & N. Korda. 1984. "Discrimination and matching within and between hues measured by reaction times: Some implications for categorical perception and levels of information processing". *Psychological Research* 46.207–222.
- Boynton, R. M. & J. Gordon. 1965. "Bezold-Brücke hue shift measured by color-naming technique". *Journal of the Optical Society of America* 55.78–86.
- Crawford L. E., J. Huttenlocker & P. H. Engbretson. 2000. "Category effects on perception of stimuli: Perception or reconstruction". *Psychological Science* 11.280–284.
- Drivonikou, G. V., P. Kay, T. Regier, R. B. Ivry, A. L. Gilbert, A. Franklin & I. R. L. Davies. 2007. "Further evidence for lateralization of Whorfian effects to the right visual field". *Proceedings of the National Academy of Sciences* 104.1097–1102.
- Franklin, A. & I. R. L. Davies. 2004. "New evidence for infant color categories". *British Journal of Developmental Psychology* 22.349–377.
- Gilbert, Aubrey L., Terry Regier, Paul Kay & Richard B. Ivry. 2006. "Whorf hypothesis is supported in the right visual field but not the left". *Proceedings of the National Academy of Sciences* 103.489–494.
- Hollingworth, H. L. 1910. "The central tendency of judgment". *The Journal of Philosophy, Psychology and Scientific Methods* 7.461–469.
- Ishihara, S. 2003. *Test for Colour-Blindness*. Tokyo: Shuppan.
- Malkoc, G., P. Kay & M. A. Webster. 2005. "Variation in Normal Color Vision IV: Unique-binary hues and hue scaling". *Journal of the Optical Society of America A*. 22: 10.2154–2168.
- Mitterer, H. & J. P. de Ruiter. 2008. "Recalibrating color categories using world knowledge". *Psychological Science* 19.629–634.
- Kay, P. & C. McDaniel. 1978. "The linguistic significance of the meanings of basic color terms". *Language* 54.610–646.

- Özgen, E. & I. R. L. Davies. 1998. "Turkish color terms: Tests of Berlin and Kay's theory of color universals and linguistic relativity". *Linguistics* 36.919–956.
- Parducci, A. 1965. "Category judgement: A range frequency model". *Psychological Review* 72.407–418.
- Petzold, P. 1982. "The edge effect of discriminability in categorical judgements". *Journal of Experimental Psychology: Human Perception and Performance* 7.1371–1385.
- & G. Haubensak. 2004. "The influence of category membership of stimuli on sequential effects in magnitude judgment". *Perception & Psychophysics* 66.665–678.
- Poulton, E. C. 1973. "Unwanted range effects from using within-subjects experimental designs". *Psychological Bulletin* 80.113–121.
- . 1974. "Range effects are characteristic of a person serving in a within-subjects experimental design – A reply to Rothstein". *Psychological Bulletin* 81.201–203.
- . 1989. *Bias in Quantifying Judgments*. Hove: Erlbaum.
- Roberson, D., J. Davidoff & N. Braisby. 1999. "Similarity and categorisation: Neuropsychological evidence for a dissociation in explicit categorisation tasks". *Cognition* 71.1–42.
- & J. Davidoff. 2000. "The categorical perception of colours and facial expressions: The effect of verbal interference". *Memory & Cognition* 28.977–986.
- Rosch Heider, E. 1972. "Universals in color naming and memory". *Journal of Experimental Psychology* 93.10–20.
- Watson, W. A. 1957. "Contrast, assimilation, and the effect of central tendency". *American Journal of Psychology* 70.560–568.
- Whorf, B. L. 1956. *Language, Thought and Reality: Essays by B. L. Whorf* ed. by J. B. Carroll. Cambridge, Mass.: MIT Press.
- Wyszecki, G. & W. S. Styles. 1982. *Color Science: Concepts and Methods, Quantitative Data and Formula*, 2nd ed. New York: John Wiley & Sons.