Memory for colours: a reaction time experiment

V. Bonnardel and J. Herrero; University of Sunderland; Sunderland, United Kingdom

Abstract

We used simultaneous and delayed match to sample tasks to investigate memory for 5 colour tests (green, yellow, purple, pink and orange) in men and women. Stimuli were emulated Munsell colour samples displayed on a CRT monitor. Colour tests were presented with distracters that could vary either in hue or in saturation. Our results indicate that: 1) over the five colours, women were more accurate than men in remembering colours (p=0.025). This advantage was significant for pink and purple colours and, in the later case, could be explained by the high women agreement in their categorical perception; 2) better memory for hue than for saturation, or distortion of the remembered colour towards more saturated samples depend on the colour of the test; 3) support for category effect in memory is provided by faster response times to the green colour test when presented with its cross-category hue distracter (p=0.025); 4) the best remembered colours were yellow, green and purple and the worst remembered colours were orange and pink.

Introduction

Several factors have been reported as influencing performances of colour memory. First, Pérez-Carpinell *et al.* (1998) [1] reported gender differences in a study using Munsell samples with retention interval varying from 15 s to 24 hours. In overall, women were more accurate than men in remembering colours for 15 s and 15 min delays, although, there was no significant effect when colour tests were considered individually. Second, compared to simultaneous colour matching, in delayed matching task, the comparison involves the mnemonic representation of the colour and discrimination performances are deteriorated. Different degree of impairment can be observed depending on the colour attribute of hue, saturation or lightness. For instance, Newhall *et al.*, (1957) [2] reported an increase in saturation of remembered colours, and Ling & Hurlbert (2002) [3] found for three colour tests with different retention intervals a better memory for hue than for saturation. Finally, colour memory appears to be categorical. In a same/different task with a 10-second inter-stimuli interval, Boynton *et al*, (1989) [4] observed that the percentage of errors (omissions) increased as the categorical colour difference index between stimuli (OSA samples) was reduced, suggesting that categorisation occurs when colours must be remembered.

In the present study, we investigate gender differences by testing men and women in simultaneous and delayed match-to-sample tasks, using the simultaneous matching task as a base-line to account for colour discriminability. To each colour test corresponds 4 distracters; two of them vary in hue (in two opposite directions from the test) while the other two vary in saturation (with an increase or a decrease from the test). This design will allow us to examine differences in performances when the colour test is presented with a hue or a saturation distracter. To determine the category membership of the hue distracters, a subset of subjects was also requested to perform a colour naming task allowing us to compare performances between cross- and within-category hue distracters.

Method

Subjects

Ninety seven subjects (59 females, 38 males) participated to the experiment. Normal trichromacy was assessed with Ishihara pseudo-isochromatic plates and all participants had normal or corrected to normal vision. The majority of the subjects were first year psychology students from the University of Sunderland who received course credits for their participation. The remaining participation was from staff members who volunteered. Written informed consents were obtained in line with the Tenets of Helsinki, and the study had the approval of the Ethic Committee of the University of Sunderland.

Stimuli and material

Stimuli were twenty five emulated Munsell colour samples (Munsell conversion program – version 6.5.1, illuminant C, 1931) corresponding to five colour tests and 20 distracters (4 distracters each). Two distracters differed from the test in Hue and the others two in Chroma (i.e. saturation). Distracters were two Munsell units away from the tests providing approximately equivalent distances in the CIE 1976 u'v' uniform-chromaticity diagram with an average distance of 0.015 and a standard deviation of 0.0038 (figure 1). All stimuli had a Munsell Value of 6/ that corresponded to an average luminance of 30.6 cd/m^2 with a standard deviation of 0.92 cd/m^2. Munsell Hue references of the colour tests were: 10BG (green), 10Y (yellow), 10PB (purple), 5RP (pink) and 10R (orange). Apart from 10BG, for which the Chroma was /5, the four other tests had a Chroma of /6. Chroma distracters had the same Hue as their test with either a higher /8 (or /7) or a lower /4 (or /3) Chroma. Hue distracters had the same Chroma as their test with the Hue shifted in one (5B, 5GY, 5P, 10RP, 5YR) or the other direction (5BG, 5Y, 5PB, 10P, 5R) in the Munsell Hue circle.

Stimuli were displayed on a Sony Trinitron GOM-F520 monitor as circular patches of 2.6° of visual angle on a uniform grey background with a luminance of 13 cd/m^2**.** Monitor calibration and colour measurements were made with an Ocean Optic spectrometer (SP2000) and luminance measurements with a Minolta chroma meter (CS100). A Matlab 6.5 program for Apple Macintosh G4 was developed to pilot the experiment.

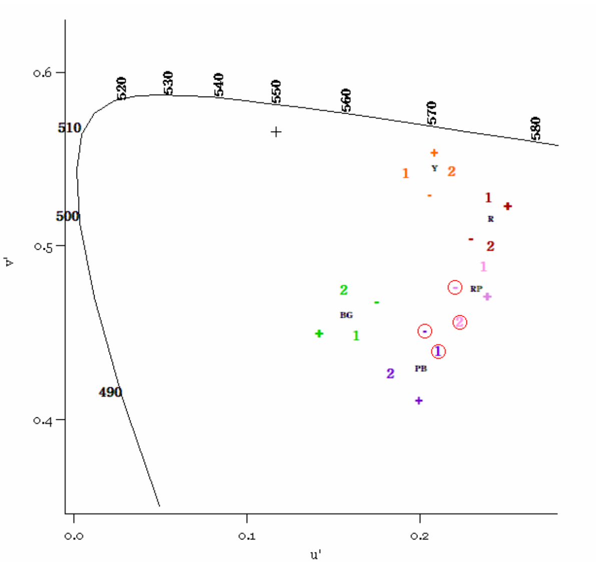

Figure 1*. Chromaticities of the 25 emulated Munsell samples in the 1976 u'v' uniform-chromaticity scale diagram. Chromaticities of the 5 colour tests are indicated by their Munsell Hue reference. Numbers '1' correspond to chromaticities of Hue distracters shifted in the anti-clockwise, and '2' in the clockwise direction. Plus symbols correspond to higher, and minus symbols to lower Chroma distracters. Circled chromaticities indicate stimuli that women remember significantly better than men (see text for explanation).*

Procedure

Each subject performed the two tasks, the order of which varied in a systematic way across participants. In the simultaneous match-to-sample or discrimination task (T1), test and test plus distracter were displayed in a triangular arrangement (figure 2). In the delayed match-to-sample or memory task (T2), the test was presented in the centre of the screen for 500 ms, and after 1 second interval during which the screen was uniform grey, test plus distracter were displayed side-by-side at the centre of the screen. Stimuli were displayed until the subject answered by pressing the appropriate key ('F' on the left or 'J' on the right) on the keyboard to indicate the test location. A feedback tone was given in case of error and reaction times superior to 2 seconds were considered as errors. A 500 ms delay separated subject's response from the next trial. Each of the 20 distracters was presented 6 times (3 on the right- and 3 on the left-hand side) providing a total of 120 experimental trials for each task. The trial presentation order was randomised across participants. Before each task, subjects performed 10 practice trials. The experiment was performed in a dark room and subjects were sitting at about 1 meter from the monitor in free viewing conditions. After the experiment, which lasted approximately 15 minutes, a subset of 55 subjects (33 women and 22 men) were asked to name the 25 stimuli presented in 5 arrays of 5 (the test and the 4 distracters) successively displayed. Colour names were restricted to 'red', 'orange', 'yellow', 'pink', 'blue', 'green' and 'purple' with the possibility of using 'pale' and 'dark' qualifiers.

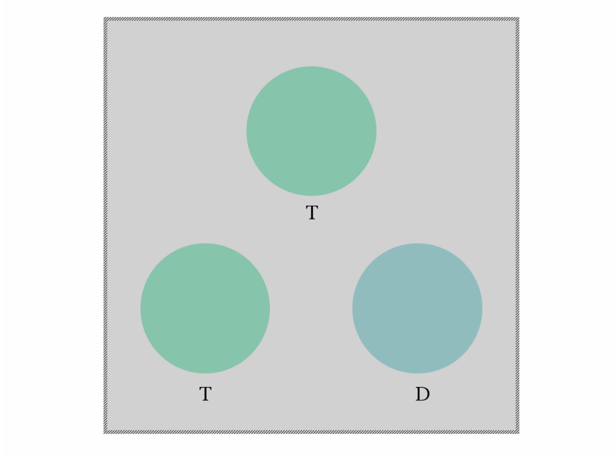

Figure 2*. Triangular arrangement of stimuli presentation in the discrimination task; top: test; bottom: test plus distracter.*

Results

Errors and reaction times were subjected to variance analyses. The discrimination task (T1) with a mean error rate (ER) of 0.171 (corresponding to 2.7% of errors) produced a significantly lower error rate than memory task (T2) with ER = 1.309 (i. e. 21%) [$F_{(1,95)} = 759.203$, $p < 0.005$][*]. The difference in reaction times (RT) between T1 (673 ms) and T2 (681 ms) was not significant.

Comparison between men and women

Error rate

Over the two tasks, women made significantly less errors (0.667) than men (0.813) [$F_{(1,95)} = 6.949$, $p = 0.01$]. The interaction between factors task and gender was significant [$F_{(1,95)} = 5.272$, $p = 0.024$], and a variance analysis performed for each task separately indicated that women were significantly more accurate than men in T2 [1.189 vs. 1.429, $F_{(1,95)} = 7.195$, $p = 0.009$] but not T1 (0.146 vs. 0.196) (figure 3).

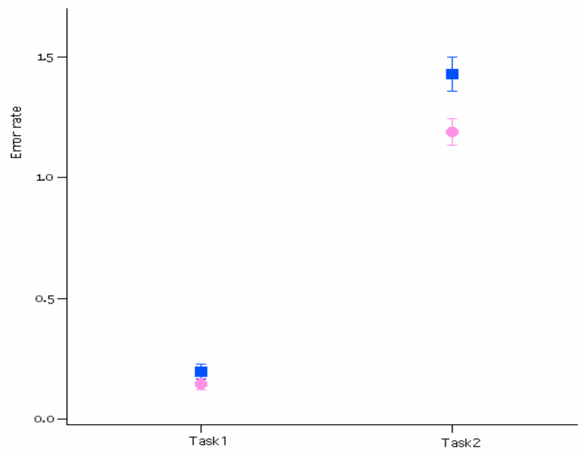

Figure 3*: Comparison between women (circles) and men (squares) error rates in discrimination (Task1) and memory (Task2) tasks. Error bars indicate the mean standard errors.*

[*] Unless noted otherwise, p values are given for two-tailed test.

To account for residual gender differences in discrimination performances, in a subsequent variance analysis, differences between T1 and T2 error rates (noted δER) was used as a dependent measure. In this second variance analysis, over all stimuli, women significantly outperformed men with a δER of 1.043 vs. 1.233 [$F_{(1, 95)} = 5.272$, $p = 0.024$].

Gender differences in accuracy were further investigated by considering distracters individually. Comparisons using *t*-tests showed that women made significantly less confusions than men on four specific distracters. Two distracters were associated to purple (10PB) and two to pink (5RP) colour tests. For each colour, one distracter corresponded to a lower Chroma (10PB 6/4 and 5RP 6/4) and the other to a Hue shift (5P 6/6 and 10P 6/6). The four distracters happened to be located in the same pink-purple area of the chromaticity diagram (figure1). Men and women δERs were respectively: 1) purple colour, Chroma distracter (10PB 6/4): 1.00 vs. 0.44 [$t = 2.471$, $df = 95$, $p = 0.0075$ one-tailed test], 2) purple colour, Hue distracter (5P 6/6): 1.37 vs. 0.49 [$t = 3.357$, $df = 50.976$, $p = 0.0005$ one-tailed test], 3) pink colour, Chroma distracter (5RP 6/4): 2.00 vs. 1.49 [$t = 1.823$, $df = 95$, $p = 0.035$, one-tailed test] and 4) pink colour, Hue distracter (10P 6/6): 1.58 vs. 1.12 [$t = 1.719$, $df = 95$, $p = 0.045$, one-tailed test]. Gender differences for purple, that were larger than for pink colour test, were due to a significantly lower than average δER in women for both distracters: Chroma [0.44 vs. 1.043, $t = 4.195$, $df = 58$, $p < 0.0005$] and Hue [0.49 vs. 1.043, $t = 5.322$, $df = 58$, $p < 0.0005$].

To further explore the origin of these differences, we considered the possibility of category effect explanation. If distracters were perceived as from a different category than that of the test, then linguistic label could have been used in addition to visual code. Under these circumstances, the memory task would be facilitated resulting in a decrease in error rate and/or reaction time. However, a categorical effect might have operated differently across individuals as it depends on stimuli category membership phenomenological judgments. In this study, category membership judgements were assessed from the naming data obtained from a subset of 55 subjects. These data indicated that 10PB test was named ‘purple’ by all subjects (33 women, 22 men). Its Hue distracter (5P 6/6) was called ‘pink’ by all but one woman (who named it ‘purple’) and by 13 men (2 named it ‘red’ and 7 ‘purple’). Likewise, its Chroma distracter (10PB 6/4) was called ‘pink’ by 30 women (3 named it ‘purple’), whereas 11 men named it ‘pink’ (2 ‘red’ and 9 ‘purple’). Overall, 94% of women and 64% men judged these two distracters to be in a different colour category from that of the test. If we assume that similar percentages exist in our full sample of subjects, then the higher proportion of women sharing the same categorical perception (i.e., a purple test presented with pink distracters), could explain their improved accuracy as a group.

This explanation would not however account for the smaller women advantage observed for the 5RP colour test. This test was named ‘pink’ by 32 of the 33 women (1 named it ‘orange’), and by 16 men (1 named it ‘orange’ and 5 ‘red’). Its Hue distracter (10P 6/6) was also named ‘pink’ by 20 women (12 named it ‘purple’) and by 5 men (1 named it ‘red’ and 16 ‘purple’). Likewise, its Chroma distracter (5RP 6/4) was named ‘pink’ by 29 women (3 named it ‘red’ and 1 ‘purple’) and by 16 men (3 named it ‘purple’ and 3 ‘red’). Only 24% of women judged the distracters as from a different category than that of the test compared to 52 % of men, yet memory advantage for pink was still in favour of women.

Reaction times

Men mean reaction time, which was identical for the two tasks (674 ms), was only marginally faster than that of women (681 ms). Women were slightly faster in T1 (673 ms) compared to T2 (689 ms); this difference was not significant. No gender difference in RTs over the twenty distracters and the two tasks was significant.

Within-subject comparisons

The emulated Munsell samples did not provide strictly equal distances between test and distracters in the u’v’ diagram. Distances in a uniform chromaticity scale diagram correspond to perceptual distances. Perceptual distances reflect colour differences that are greater than the just noticeable colour differences and do not relate to discrimination thresholds in a simple way. It was therefore necessary to verify how these distances affected colour discriminability as measured, in our experiment, by errors and reaction times. There was a significant and negative correlation between each dependant variables and u’v’ distances in the discrimination task (table 1). Thus u’v’ distances could be a potential confounding variable with distracter types in the discrimination task. Although not significant, negative correlations were also obtained in the memory task (table 1) and to account for stimuli discriminability δER was used as the dependent variable in the following comparison unless mentioned otherwise.

	Task1		Task2	
	R	*p**	R	*p**
ER	-0.46	0.02	-0.326	0.08
RT	-0.42	0.0325	-0.29	0.104

*one-tailed test

Table 1: Spearman coefficient of correlations and p-values computed between test-distracter u’v’ distances and the two dependant variables error rates (ER) and reaction times (RT) for the two tasks.

Comparison between colour tests

The three best remembered colours were yellow (δER = 0.947), green (δER = 0.984) and purple (δER = 1.051) and the worst remembered colours were orange (δER = 1.291) and pink (δER = 1.417), this latter colour had also the longest RTs in T2. The δER difference between purple and orange was significant [$F_{(1,96)} = 10.68$, $p = 0.002$].

Comparison between Hue and Chroma distracters

Over the five colour tests, Hue distracters gave rise to significantly less errors than Chroma distracters (2.082 vs. 2.384 [$F_{(1,96)} = 8.045$, $p < 0.004$]) and to significantly faster reaction times in T2 (675 ms vs. 690 ms, [$F_{(1,96)} = 5.682$, $p = 0.019$]). However, interaction between factors distracters and colours was significant [$F_{(4,384)} = 17.408$, $p < 0.0005$]. Analyses made for each colour test independently indicated a better accuracy in presence of a Hue distracter for yellow (1.392 vs. 2.309, [$F_{(1,96)} = 14.116$, $p < 0.0005$]) and purple (1.268 vs. 2.804, [$F_{(1,96)} = 53.37$, $p < 0.0005$]), an absence of difference between Hue and Chroma distracters for green and a significantly larger δER for Hue distracters in case of pink

(3.041 vs. 2.557, [$F_{(1,96)}$ = 4.092, p = 0.046]) and orange colours (2.722 vs. 2.351, [$F_{(1,96)}$ = 3.975 p = 0.049]). Distracters that produced lower δERs were systematically associated with significantly faster reaction times in T2.

Comparison between distracters varying in Chroma

Over the five colour tests, distracters with higher Chroma produced significantly more errors (1.464 vs. 0.924, [$F_{(1,96)}$ = 14.362, p < 0.0005]) and slower reaction times in T2 (740 ms vs. 640 ms, [$F_{(1,96)}$ = 53.962, p < 0.0005]) compared to distracters with lower Chroma. The interaction between factors distracters and colours was significant [$F_{(4,384)}$ = 34.358, p < 0.0005], and the described pattern of results was obtained for green (1.6 vs. 0.31, [$F_{(1, 96)}$ = 56.452, p < 0.0005]), yellow (1.649 vs. 0.66, $F_{(1,96)}$ = 16.435, p < 0.0005]) and purple colours (2.144 vs. 0.66, [$F_{(1, 96)}$ = 44.207, p < 0.0005]) with higher Chroma distracters producing in all cases significantly slower reaction times in T2. However, for orange and pink colours, lower Chroma gave rise to more errors than higher Chroma distracters, and in the later case, the difference was significant (0.866 vs. 1.691, [$F_{(1,96)}$ = 15.994, p < 0.0005]); no significant difference in RT was observed for these two colours.

Comparison between distracters varying in Hue

Everything being equal otherwise, in our experiment, category effect will be reflected by better performances when the test is presented with its hue cross-category compared to its within-category distracter. To test for category effect, we had to unsure that the colour test and one of its distracter were reported as from the same category (within-category distracter) while the other distracter belonged to a different category (cross-category distracter). From the naming data, this criterion was fulfilled for the green colour test. This test (10BG) together with its 5BG distracter were named ‘green’ whereas 5B was named ‘blue’ by 46 (18 men, 28 women) of the 55 subjects. The other colour tests gave a poor naming agreement. Hence, the next analysis has been limited to the 46 subjects for the green colour test.

As categorical effect could also be observed at the discrimination level, differences between the two Hue distracters were considered for each task separately. It should be noted that the cross-category distracter (5B) was at a slightly smaller distance from the test (0.0132) than the within-category distracter (5BG) (0.0146) and would be at disadvantage (assuming the validity of the u’v’ spacing and assuming the difference in distance being of noticeable magnitude). In the discrimination task, a lower but non-significant error rate (0.073 vs. 0.181) associated to marginally longer reactions times (728 ms vs. 718 ms) was observed for the cross-category distracter. In the memory task, a non-significant lower error rate (1.099 vs. 1.236), this time, associated to a significantly faster reaction time was observed for cross-category distracter (681 vs. 749 ms, [$F_{(1,44)}$ = 5.379, p=0.025]).

Discussion

With distracters two Munsell steps away from the test, the discrimination task provided a small percentage of errors (2.7%). In the memory task, deterioration of performances was reflected by a higher percentage of errors (21%), whereas response times were unafected. Discrimination was generally more difficult with increase of errors and reaction times for smaller u’v’ distances between test and distracters and discriminability performances as measured by error rates were taken into account when memory performances per se were considered.

As reported by Pérez-Carpinell *et al.* (1998), women were more accurate than men in remembering colours. This advantage was significant for the purple-pink range of colours, and resulted from a women improvement rather than a men impairment in accuracy for the purple colour test. If we accept to generalise the naming data obtained from the subset of 55 subjects to our full sample of participants, then women memory advantage for purple could partly be explained by their larger agreement in categorical perception of test and distracters. For the pink colour, women agreement on the categorical perception of test and distracters was lower than that of men, and their advantage calls for alternative explanation. It is worth noting that gender differences for pink colour have repeatedly been observed in colour preference tasks. Hurlbert *et al.* (2003) [5], for instance, reported a specific consensual preference in women for lilac-pink colours. The authors attributed the origin of this difference to differences in the relative prominence of men and women colour mechanisms. The existence of perceptual mechanisms responsible for the differential salience of pink colour across genders could account for women memory advantage in absence of evidence of a verbal strategy. In the present study, since no evidence of gender difference was observed in the supra threshold discrimination task, it confirms that these mechanisms must occur at higher levels of colour processing.

Considering colour attributes, subjects had a more accurate hue memory for yellow and purple, but a more precise memory of saturation for pink and orange colours. These results do not suggest a systematic advantage in memory for one particular attribute. Likewise, a systematic distortion in memory towards more saturated colours is not supported by the present results. Indeed, the memory task was more difficult when the test was presented with its more saturated distracter for green, yellow and purple colours, but this was not the case for orange and pink colours, and for the later, the task was made more difficult in presence of a desaturated distracter.

Finally, for methodological reasons, test for category effect in memory was limited to the performance of 46 subjects for the green colour test. There was no clear evidence of category effect in the simultaneous match to sample task. However, even if not significant, the error rate was lower for the cross-category distracter despite its smaller u’v’ distance from the test. There is a possibility that category effect facilitation had actually compensated for the smaller u’v’ distance, and significant differences might have been observed with distracters of equal perceptual distances. Yet, when using emulated Munsell samples as stimuli, unless just noticeable differences (JND) are determined for each subject prior to the experiment, we must rely on validity of the u’v’ spacing and do not know how this spacing relates to JNDs of individual observers. Besides, although all the stimuli had the same Munsell Value there were subtle luminance variations that could have been used by the subjects. The lack of appropriated control of stimuli physical parameters for the discrimination task leads us to limit the evidence of a category effect to the memory task for which response times were faster when the green colour test was presented with its blue compared to its green distracter.

References

[1] J. Pérez-Carpinell, R. Baldoví, M. D. de Fez and J. Castro, "Color memory matching: Time effect and other factors. Colour Res. and App., 23, 234 (1998).

[2] S. M. Newhall, R. W. Burham, and J. R. Clark, "Comparison of successive with simultaneous colour matching. J. of Opt. Soc. of Am., 47, 43 (1957).

[3] Y. Ling and A. Hurlbert, Colour memory under changing illumination, Perception, pg. 131 (2002)

[4] R. M. Boyton, L. Fargo, C. X. Olson, and H. S. Smallman, "Category effects in colour memory. Colour Res. and App., 14, 229 (1989).

[5] A. Hurlbert, Y. Ling and J. Robinson, 'Real men don't like pink': Sex differences in colour preference, Perception, pg. 29. (2003).

Author Biographies

Valérie Bonnardel holds a PhD in Neurosciences from the University Pierre & Marie Curie-Paris VI (France). She worked as a research associate in the department of Experimental Psychology and then in the Physiological Laboratory at the University of Cambridge (UK). Since 2001, she is a Senior Lecturer in the Department of Psychology at the University of Sunderland (UK). She is a member of the Colour Group (Great-Britain) and of the International Colour Vision Society. Her research interest focuses on individual differences in human colour vision.

Jose Herrero received his BSc in Psychology in 2001 from the Pontifical University of Salamanca (Spain). He is currently in the final stage of his PhD in cognitive neuroscience at the University of Sunderland (UK) and has been a visiting scientist in 2004/05 in the Department of Perceptual Dynamics - Riken-Brain Science Institute (Japan). His work concerns the development of a new experimental paradigm in the study of visual working memory.