A Cross-Cultural Comparison of Colour Emotion for Two-Colour Combinations

Li-Chen Ou,[1]* M. Ronnier Luo,[1] Pei-Li Sun,[2] Neng-Chung Hu,[3] Hung-Shing Chen,[4] Shing-Sheng Guan,[5] Andrée Woodcock,[6] José Luis Caivano,[7] Rafael Huertas,[8] Alain Treméau,[9] Monica Billger,[10] Hossein Izadan,[11] Klaus Richter[12]

[1]Department of Colour Science, University of Leeds, Leeds, United Kingdom

[2]Graduate Institute of Engineering, National Taiwan University of Science and Technology, Taiwan

[3]Department of Electronic Engineering, National Taiwan University of Science and Technology, Taiwan

[4]Graduate Institute of Electro-Optical Engineering, National Taiwan University of Science and Technology, Taiwan

[5]Department of Visual Communication Design, National Yunlin University of Science and Technology, Taiwan

[6]Coventry School of Art and Design, Coventry University, United Kingdom

[7]University of Buenos Aires and National Council for Research, SICyT-FADU-UBA, Argentina

[8]Departamento de Óptica, Universidad de Granada, Spain

[9]Université Jean Monnet, Laboratoire Hubert Curien UMR 5516, Saint Etienne, France

[10]Department of Architecture, Chalmers University of Technology, Sweden

[11]Department of Textile Engineering, Isfahan University of Technology, 84156-83111 Isfahan, Iran

[12]Bundesanstalt für Materialforschung und -prüfung (BAM), Germany

Received 17 April 2010; revised 7 June 2010; accepted 3 July 2010

Abstract: Psychophysical experiments were conducted in the UK, Taiwan, France, Germany, Spain, Sweden, Argentina, and Iran to assess colour emotion for two-colour combinations using semantic scales warm/cool, heavy/light, active/passive, and like/dislike. A total of 223 observers participated, each presented with 190 colour pairs as the stimuli, shown individually on a cathode ray tube display. The results show consistent responses across cultures only for warm/cool, heavy/light, and active/passive. The like/dislike scale, however, showed some differences between the observer groups, in particular between the Argentinian responses and those obtained from the other observers. Factor analysis reveals that the Argentinian observers preferred passive colour pairs to active ones more than the other observers. In addition to the cultural difference in like/dislike, the experimental results show some effects of gender, professional background (design vs. nondesign), and age. Female observers were found to prefer colour pairs with high-lightness or low-chroma values more than their male counterparts. Observers with a design background liked low-chroma

This article was published online on 12 November 2010. An error was subsequently identified in Table II. This notice is included in the online and print version to indicate that both have been corrected.

*Correspondence to: Li-Chen Ou (e-mail: l.ou@leeds.ac.uk).

colour pairs or those containing colours of similar hue more than nondesign observers. Older observers liked colour pairs with high-lightness or high-chroma values more than young observers did. Based on the findings, a two-level theory of colour emotion is proposed, in which warm/cool, heavy/light, and active/passive are identified as the reactive-level responses and like/dislike the reflective-level response. Col Res Appl, 37, 23–43, 2012; Published online 12 November 2010 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/col.20648

Key words: colour emotion; colour preference; cultural difference; gender difference; age effect; design training

INTRODUCTION

Colour plays a significant role in global marketing communication, as colour not only helps enhance brand recognition[1,2] but it also translates intended visual impressions into design elements of a product,[3,4] it can create desired atmosphere in a retail outlet[5,6] and influence consumers' behavior.[7–9] The impact of colour can be attributed to a wide range of colour associations[10] in which colour is seen as either a symbol[11] or a sign.[12] The former regards colour as representation of physical items or experiences (e.g., red is a symbol of "good luck" in China), whereas the latter uses colour to provide specific information in visual communication (e.g., red means "stop" as a traffic signal). In both cases, colour may serve either as an emotion elicitor[13] that creates emotional impact on the viewer[14–16] (e.g., Chinese people are happy to receive the "red envelopes" on Chinese New Year's Eve) or as an emotion messenger[17] sending a communicative signal describing affective quality of the colour itself or of the environment/product[18–20] (e.g., a vivid red is often regarded as a warm, exciting colour). Colour as an emotion messenger has attracted enormous interest from researchers in different disciplines, who have given various names to work in this area such as "colour meaning,"[21–24] "colour image,"[25,26] "colour emotion,"[27–33] and "expectations."[34–36] For consistency, and following the tradition of previous work by the authors, the term colour emotion will be used in this article to indicate the subject area described here.

There seems to be predictable trends in colour emotion responses. For instance, reddish colours are often regarded as "warm," and bluish colours "cool."[22,25,28,31,37] Such trends can be measured using psychophysical methods[38] in an approximately uniform colour space, such as the CIELAB system.[39] It has been shown that colours similar in appearance (e.g., hue, lightness, and chroma) will elicit similar responses in semantic terms. For instance, the "warm/cool" response has been modeled quantitatively as a function of hue and chroma by Sato *et al*.[27] and by Ou *et al*.[28] The perhaps most interesting finding in this area is the additivity property of colour emotion,[29,40,41] which states that the colour emotion value (e.g., the warm/cool response) of a colour pair can be determined by the average of the colour emotion values for the individual colours in that pair.

Another interesting trend found is that the colour emotion responses can be classified into three independent, orthogonal dimensions, correlated closely with the three colour appearance attributes, i.e., hue (corresponding to the warm/cool response), lightness (heavy/light), and chroma (active/passive).[25,27,28,31,37] It has been shown that a fourth dimension may exist, called "evaluation."[42–44] This fourth scale is related to liking and has been called by some researchers colour preference.[21,30]

Diverse findings have been shown with regard to the cultural difference in colour emotion. Some studies have suggested that colour preference is a universal, consistent psychophysical pattern across cultures. For instance, Eysenck[45] compared existing colour preference studies which he divided into two ethnic groups for the observers: Caucasian and other races. He found the same rankings of colour preferences for the two observer groups—blue being the most preferred, followed by red, green and purple; the least preferred colours being orange and yellow. Saito[46] studied colour preference using observers in Japan, Korea, and Taiwan. She found that white and blue were the most preferred colours for all observers. The two colours (as well as red) were also found to be the most preferred in China.[47] In addition to colour preference, studies have shown consistent colour association responses across cultures.[18,23,31] The consistency shown by the studies seems to support Humphrey's theory[48,49] which, based on his investigation of rhesus monkeys' behavior, asserts that human reactions to colour are innate, with a consistent, universal pattern shared by individuals, and that such consistency is a result of evolution for biological survival. Hurlbert and Ling[50] also took the biological view in explaining the gender difference in hue preferences as they found female observers preferred redder colours more than the male.

However, a number of studies have shown evidence of significant cultural differences in psychological responses to colour. For instance, Garth[51] studied colour preference using Indian, Caucasian, and the mixed-race of the two ethnic groups. The results showed that Indian observers preferred red the most and white the least; Caucasian preferred blue the most and white the least; the mixed-race preferred blue the most and yellow the least. Choungourian[52] studied colour preference using observers in the US, Lebanon, Iran, and Kuwait. Red was found to be the most preferred colour for the US observers, green for the Lebanese, and cyan for observers in Iran and Kuwait. Shoyama *et al*.[53] studied colour preference for clothing using Japanese and Korean women as the observers. The results showed that Japanese observers preferred black and light-greyish orange, whereas Korean observers preferred dark blue and light grey. It is unclear whether the results were affected by the context (i.e., clothing). In addition to colour preference, researchers also found significant cultural difference in observer responses to colour in both semantic[8,9,21] and affective terms.[54] Crozier argued[55] that the differences in colour preference are due to the fact that

TABLE I. Number of observers for each group taking part in this study.

	Male	Female	Age range (average)	Professional background	Date of experiment
British	9	3	19–31 (25.6)	Design	2003
Taiwanese-1	16	14	22–42 (27.7)	Visual Communication Design	2003
Taiwanese-2	16	16	23–69 (41.9)	Engineering	2008–2009
French	8	13	21–51 (28.8)	Vision Engineering	2003–2004
German	10	10	20–30 (25.5)	Material Science	2004
Spanish-1	10	10	19–36 (27.4)	Nondesign	2003–2004
Spanish-2	10	12	19–33 (27.0)	Nondesign	2009–2010
Swedish	11	9	21–42 (28.3)	13 Design + 7 Nondesign	2004
Argentinian	11	15	15–76 (45.1)	14 Design + 12 Nondesign	2009–2010
Iranian	10	10	22–43 (30.5)	Nondesign	2009
	111	112			

the human reactions to colour are conditioned by learned experiences. It has been demonstrated[56] that the evaluative responses to products are affected not only by professional training (design vs. nondesign) but also by duration of the training (1st- to 3rd-year design students and lecturers in design).

These diverse findings have raised several questions, e.g., why do some studies show significant cultural difference in colour emotion responses while others show little? Are the colour emotion responses learned or innate reactions? Is there a universal, predictable pattern of colour emotion?

This study intends to address the questions by analyzing psychophysical data of colour emotion obtained across different countries. In addition to the nationality/cultural factor, the effects of gender, professional background (design vs. nondesign), and age (young adults vs. older adults) will be examined. It must be acknowledged that many other factors are likely to influence colour emotion but are not investigated in this work, e.g., context, personality, lighting condition, or the size of colour stimulus. These variables will be considered in future studies.

METHODS

Four colour emotion scales, warm/cool, heavy/light, active/passive and like/dislike, were used to assess 190 colour pairs presented on cathode ray tube (CRT) displays by 223 observers in eight countries. The former three colour emotion scales have been recognized by earlier studies[25,27,28,31,37] as the three principal underlying dimensions of colour emotion. Like/dislike has been identified as an evaluative dimension of colour emotion,[42–44] also regarded as a colour preference scale.[21,30]

Observers

A total of 223 observers, comprising 12 British, 62 Taiwanese (from two institutions in Taiwan), 21 French, 20 German, 42 Spanish (consisting of two observer groups, one carrying out the experiment in 2003 and the other in 2009, with two observers participating in both sessions), 20 Swedish, 26 Argentinian, and 20 Iranian. All the observers were born and living in their native countries.

Table I summarizes the variation in the observer groups. Taiwanese-2 observers and Argentinian observers were the only ones that consisted of young (i.e., university students/staff) and older adults (local citizens aged over 60 years). The other observer groups included only young adults. British and Taiwanese-1 groups consisted solely of observers with training in design. Taiwanese-2, French, German, Iranian, and Spanish groups consisted of only those with a nondesign background. Swedish observers consisted of 13 with a design background and seven nondesign. Argentinian observers consisted of 14 with a design background and 12 nondesign.

Each observer passed Ishihara's Tests for Colour Deficiency[57] without showing any errors in the test. The four word pairs, warm/cool, heavy/light, active/passive, and like/dislike, were translated into the required languages by native speakers, as summarized in Table II. During the experiment, each observer assessed colour emotion using word pairs in native language.

In the experiment, the observer used the four colour emotion scales to assess 190 colour pairs (each colour was 7 $\times$ 7 cm^2 in size) presented individually in a random sequence on a calibrated 20-inch CRT monitor, situated in a darkened room. Figure A1 (in Appendix A)

TABLE II. Translations of colour emotion scales used in the experiment.

UK	Taiwan	France	Germany	Spain	Sweden	Argentina	Iran
Warm	溫暖的	chaud	warm	cálido	varm	cálido	گرم
Cool	寒冷的	froid	kalt	frío	kylig	frío	سرد
Heavy	重的	lourd	schwer	pesado	tung	pesado	سنگین
Light	輕的	léger	leicht	ligero	ljus	liviano	سبک
Active	主動積極的	actif	aktiv	activo	aktiv	activo	فعال
Passive	被動消極的	passif	passiv	pasivo	passiv	pasivo	غیرفعال
Like	喜歡	aimer	mögen	gusta	gilla	gusto	دوست داشتن
Dislike	不喜歡	ne pas aimer	nicht gern mögen	desagrada	ogilla	disgusto	دوست نداشتن

TABLE III. CIELAB specifications of the 20 colour samples.

	L^*	a^*	b^*	C_{ab}^*	h_{ab}
1	45.9	61.7	29.1	68.2	25
2	84.8	6.3	82.0	82.3	86
3	61.4	−49.7	17.8	52.8	160
4	49.6	−8.9	−33.2	34.4	255
5	38.0	13.8	−42.0	44.2	288
6	42.2	25.9	26.5	37.0	46
7	58.3	−3.2	40.3	40.4	94
8	39.3	−28.2	−5.8	28.8	192
9	41.4	5.0	−24.3	24.8	282
10	84.7	17.1	5.6	18.0	18
11	89.0	1.6	39.4	39.4	88
12	78.4	−26.7	−10.9	28.8	202
13	74.0	11.3	−23.7	26.3	296
14	64.2	−8.2	19.0	20.7	113
15	47.1	−22.0	−5.7	22.7	195
16	49.8	10.8	−11.9	16.1	312
17	15.7	0.3	−1.5	1.6	282
18	43.2	0.3	0.2	0.4	37
19	72.1	0.4	0.6	0.7	58
20	97.8	−2.1	0.4	2.1	168

shows the screen layout for the experiment. The viewing distance was ~50 cm, and the viewing angle for the colour pair as a whole was about 16°. The sequence for the ratings of the four colour emotion scales was randomized for each colour pair. Also randomized was the left-right order of the appearance of individual colours within each colour pair. Between presentations of each colour pair was a full-screen grey slide that lasted for one second to avoid any effect of afterimage that might arise from viewing the colour pair for too long.

The experiment was divided into six sessions. There were 64 colour pairs presented in Session 1, 64 in Session 2, and 62 in Session 3. These were then replicated in Sessions 4–6 to evaluate the intraobserver variability (i.e., observer repeatability). Each session took ~30 min to complete. The experimental instructions were shown to each observer in their native languages before the assessments. The English version of the full instructions is given in Appendix A.

The 190 colour pairs consisted of all possible two-colour combinations generated by 20 colours that covered a reasonably wide range of hue, lightness, and chroma in the CIELAB space. The 190 colour pairs have been studied in a previous colour emotion research.[29] Each colour pair was presented at the center of the screen against a medium grey background with $(x, y, Y) = (0.309, 0.332, 13.5)$. The luminance level for the display peak white was 70 cd/m^2. Table III shows CIELAB specifications of the 20 colours. Figures 1(a) and 1(b) show the distribution of the 20 colours in the CIELAB space.

Display Characterization

The experiments used a CRT display at each experimental site to present colour stimuli, and thus, it was important to ensure accurate and precise colour reproduction by the eight [now X-Rite] displays. This was achieved by use of a GretagMacbeth Chroma-4 colour sensor for conducting characterization of each display. The Chroma-4 was a tristimulus colourimeter designed for colour measurement in terms of tristimulus values under the CIE standard colourimetric observer conditions. During the display characterization, a Chroma-4 was attached to center of the screen via suction cups to perform colour measurement with a sensing area of 1.5 cm in diameter. The measured data were then analyzed to develop a GOG characterization model[58] that predicted monitor RGB from the CIE tristimulus values. The GOG model was tested for each display by means of the CMC (1:1) colour difference[59] using the top 12 colours (i.e., colours No. 1–12) of the colour-rendition chart developed by McCamy *et al.*[60] The average colour difference values for the displays were 0.62 for the UK, 0.34 for Taiwan, 0.56 for France, 0.65 for Germany, 0.20 for Spain,

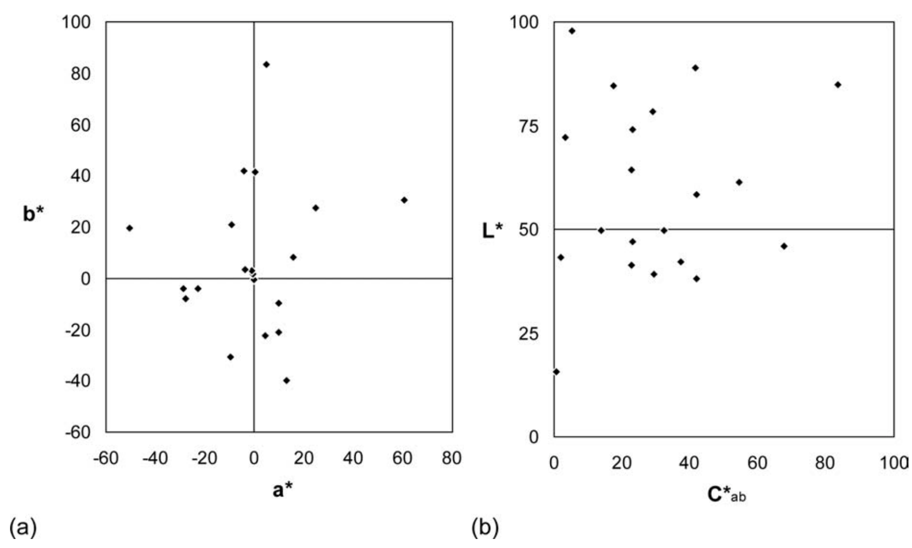


FIG. 1. Distribution of the 20 colour samples (for generating the 190 colour pairs used in the experiment) in (a) $a^* - b^*$ plane and (b) $L^* - C^*_{ab}$ plane.

TABLE IV. Intraobserver variability values (RMS) for the four colour emotion scales, with scale values ranging from 5 to 5.

	Warm/ Cool	Heavy/ Light	Active/ Passive	Like/ Dislike	Mean
British	2.20	2.28	2.40	2.43	2.33
Taiwanese	2.15	2.29	2.43	2.44	2.33
French	2.18	2.37	2.34	2.32	2.30
German	2.06	2.24	2.43	2.48	2.30
Spanish	2.14	2.48	2.56	2.60	2.45
Swedish	2.05	2.10	2.30	2.32	2.19
Argentine	2.09	2.27	2.48	2.52	2.34
Iranian	2.10	2.15	2.27	2.79	2.33
Mean	2.12	2.27	2.40	2.49	2.32

The lower the value is, the more repeatable the observers' responses.

0.61 for Sweden, 0.65 for Argentina, and 0.12 for Iran. These indicate a highly acceptable level of accuracy in colour reproduction by each display.

RESULTS

Torgerson's law of categorical judgement[61] was used to convert the raw data (with a scale range from −5 to 5) into *z*-score related scale values. A working example of the method is provided in Appendix G. The responses obtained from Taiwanese-1 and Taiwanese-2 observers were combined into one group (called "Taiwanese") as the two observer groups show close agreement in the colour emotion responses, with a correlation coefficient of 0.95 for warm/cool, 0.94 for heavy/light, 0.92 for active/passive, and 0.81 for like/dislike. The Spanish-1 and Spanish-2 were combined into one "Spanish" group also because of close agreement between the two observer groups in the colour emotion responses, with a correlation coefficient of 0.90 for warm/cool, 0.91 for heavy/light, and 0.91 for active/passive. The like/dislike data obtained from the Spanish-1 were excluded from the data analysis due to incorrect translations of the word pair. Thus, the like/dislike data for the combined Spanish group consisted of only those from the Spanish-2 observers. The intention of combining two groups into one group for the same nationality as described above was to create a dataset that better represents the country with better accuracy and higher reliability of the scale values due to larger number of observers involved. The resulting datasets were then used in cross-nationality comparisons as will be shown later. Data reliability was examined in terms of intra- and interobserver variability.

Intraobserver Variability

The intraobserver variability indicates how well an observer's responses can be repeated under the same viewing conditions. In this study, each observer did the entire visual assessments twice, and the two sets of responses were used to determine the intraobserver variability, by means of the root mean square (RMS):

$$\text{RMS} = \sqrt{\frac{\sum_i (x_i - y_i)^2}{N}} \quad (1)$$

where x_i and y_i are the first and second sets, respectively, of colour emotion responses to stimulus i; N is the total number of stimuli (i.e., 190).

The lower the RMS value, the more agreement between the replicated stimuli. Table IV summarizes the test results, showing that among the four colour emotion scales the warm/cool response was the most repeatable (with a mean RMS of 2.12) and like/dislike the least (2.49). Among the eight observer groups, the Swedish observers' responses were the most repeatable (2.19), and the Spanish observers the least (2.45). Compared with an earlier study[62] using the same scaling method (with an overall mean RMS value of 2.43 for 17 observers), the RMS values shown here seem to be reasonably low, indicating good repeatability of observers' responses.

Interobserver Variability

The interobserver variability indicates how well the observers' responses agree with each other within the group. This was determined by a revised version of RMS:

$$\text{RMS} = \sqrt{\frac{\sum_i (x_i - \bar{x}_i)^2}{N}} \quad (2)$$

where x_i represents an observer's colour emotion response to stimulus i; $\bar{x}_i$ represents the mean value for all observers within the group for stimulus i; N is the total number of stimuli (i.e., 190).

The lower the RMS value, the more the observers' responses agree with each other within the group, and thus, the more accurate the responses. Table V summarizes the test results, showing that among the four colour emotion scales, the heavy/light responses were the most accurate within the group (with a mean RMS of 1.94) and like/dislike the least (2.49). Among the eight groups, the Swedish observers' responses were found to agree most

TABLE V. Interobserver variability values (RMS) for the four colour emotion scales, with scale values ranging from 5 to 5.

	Warm/ Cool	Heavy/ Light	Active/ Passive	Like/ Dislike	Mean
British	2.04	1.88	2.16	2.41	2.12
Taiwanese	1.91	1.82	2.00	2.39	2.03
French	2.09	2.13	2.12	2.38	2.18
German	1.94	1.90	2.11	2.44	2.10
Spanish	2.05	2.05	2.30	2.53	2.23
Swedish	1.86	1.68	1.98	2.43	1.99
Argentine	2.17	2.10	2.26	2.64	2.29
Iranian	2.27	1.97	2.28	2.70	2.31
Mean	2.04	1.94	2.15	2.49	2.16

The lower the value is, the more consistent the observers' responses within an observer group.

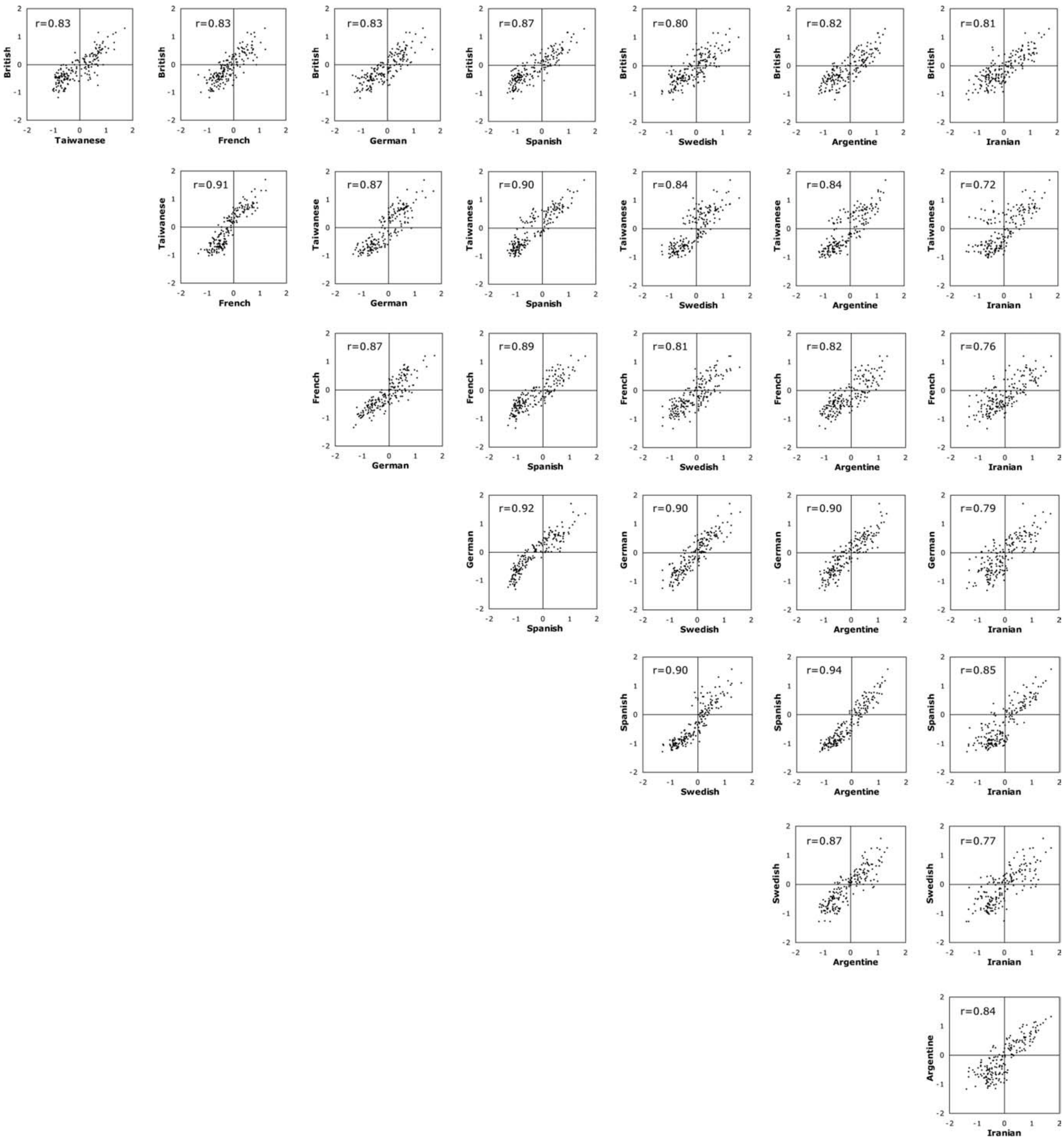


FIG. 2. Comparisons of the warm/cool responses between the eight observer groups: British, Taiwanese, French, German, Spanish, Swedish, Argentinian, and Iranian.

with each other (1.99), and the Iranian observers the least (2.31). Compared with an earlier study[62] (with an overall mean RMS value of 2.24 for 17 observers), these RMS values are reasonably low, indicating good agreement of observers' responses within each group.

Effect of Culture

Comparisons of colour emotion responses were made between the eight observer groups. For each colour emotion scale, the responses obtained from the eight observer groups were plotted against each other. Figures 2–5 show the results for warm/cool, heavy/light, active/passive, and like/dislike, respectively. In each graph, each point represents a colour pair, with the coordinates for the two axes defined as the scale values (calculated using the categorical judgement method[61]; see Appendix G for a working example) for any two of the observer groups that are compared. Ideally, a perfect agreement is obtained if all points in the graph lie on a 45° straight line. The more

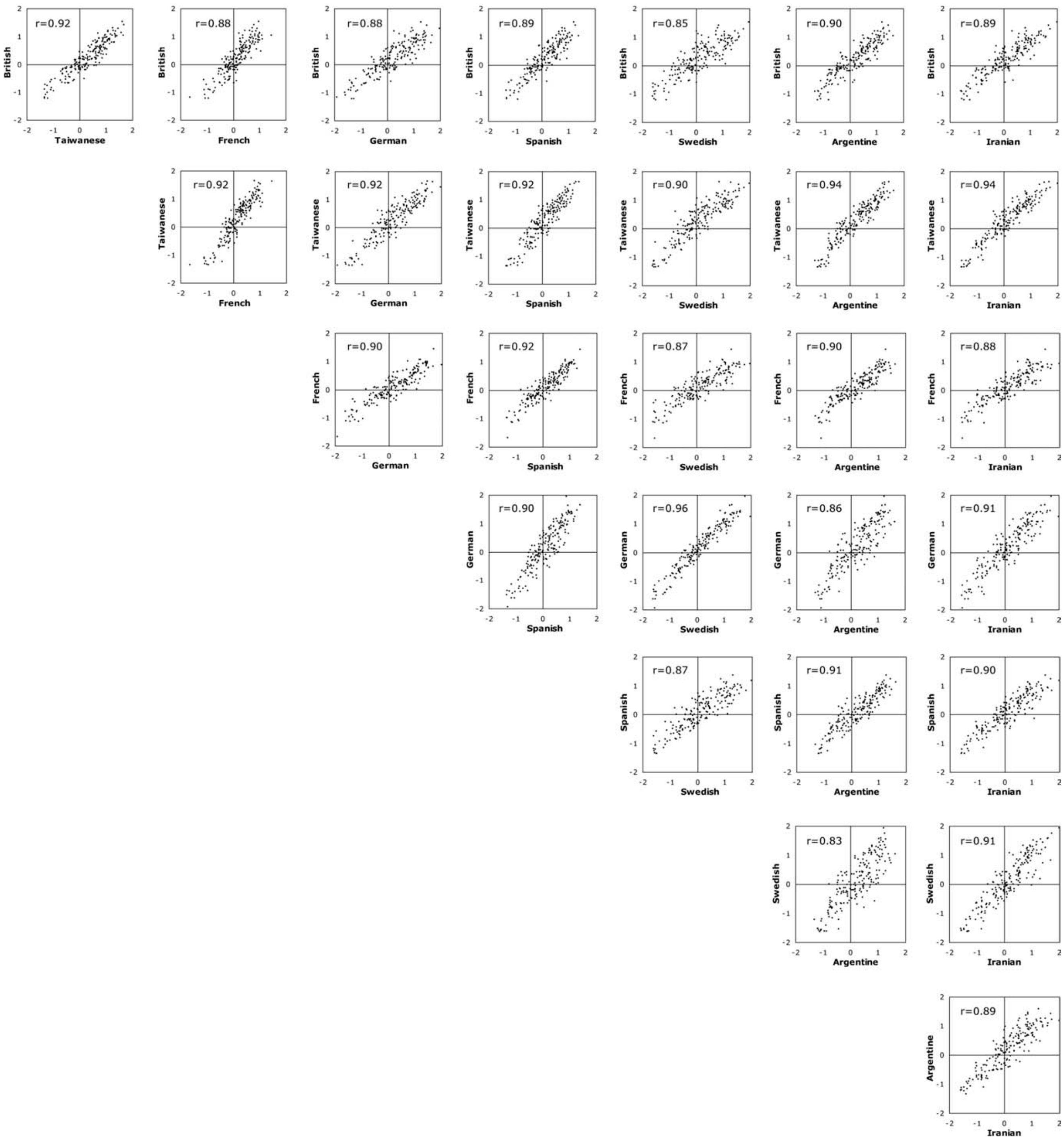


FIG. 3. Comparisons of the heavy/light responses between the eight observer groups: British, Taiwanese, French, German, Spanish, Swedish, Argentinian, and Iranian.

widely the points spread out, the less agreement between the two observer groups. Table VI summarizes the correlation coefficients obtained from comparisons between each observer group.

As a result, the responses for warm/cool, heavy/light, and active/passive all show reasonable agreement between the eight observer groups, with correlation coefficients ranging from 0.72 (for the warm/cool response between Taiwanese and Iranian) to 0.96 (for heavy/light between German and Swedish). It is interesting to note that the warm/cool response is consistent across the observer groups although the room temperature varied from one experimental site to another.

Although all correlation coefficients shown in Table VI were statistically significant (i.e., the p values obtained from the significance test were all under 0.01), only those for like/dislike contain correlation coefficients of less than 0.70, which corresponds to an R^2 value of 0.49, or about 50% of the total variance in one explained by the other. This suggests that the like/dislike response may have been

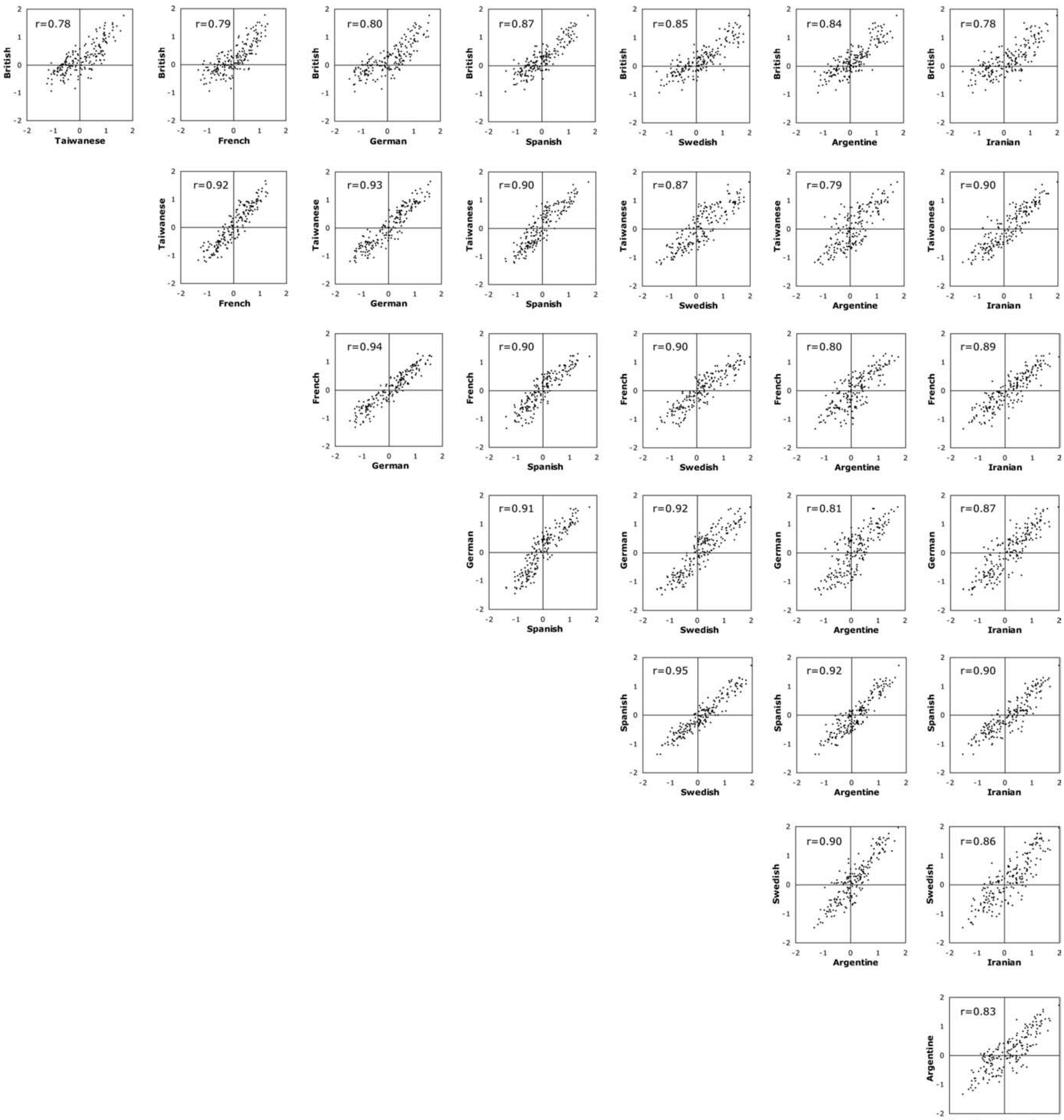


FIG. 4. Comparisons of the active/passive responses between the eight observer groups: British, Taiwanese, French, German, Spanish, Swedish, Argentinian, and Iranian.

influenced by the effect of culture. Most of the low correlation coefficients were found for comparisons between Argentinian group and the other groups, with the lowest value 0.27 for the comparison with French observers, followed by 0.32 for the comparison with German, and 0.35 for the comparison with British. Although the fact that Argentinian observers included both young and older adults may seem to contribute to the lower correlation coefficients, Taiwanese observers also consisted of both young and older adults but they show high correlation coefficients for comparisons with the other observer groups. Observer responses of the two age groups, i.e., young and older adults, will be compared in a later section of this article using the data obtained from Taiwanese and Argentinian groups.

To analyze the interrelationship between observers' responses for the four colour emotion scales, and that of the eight observer groups, factor analysis was carried out

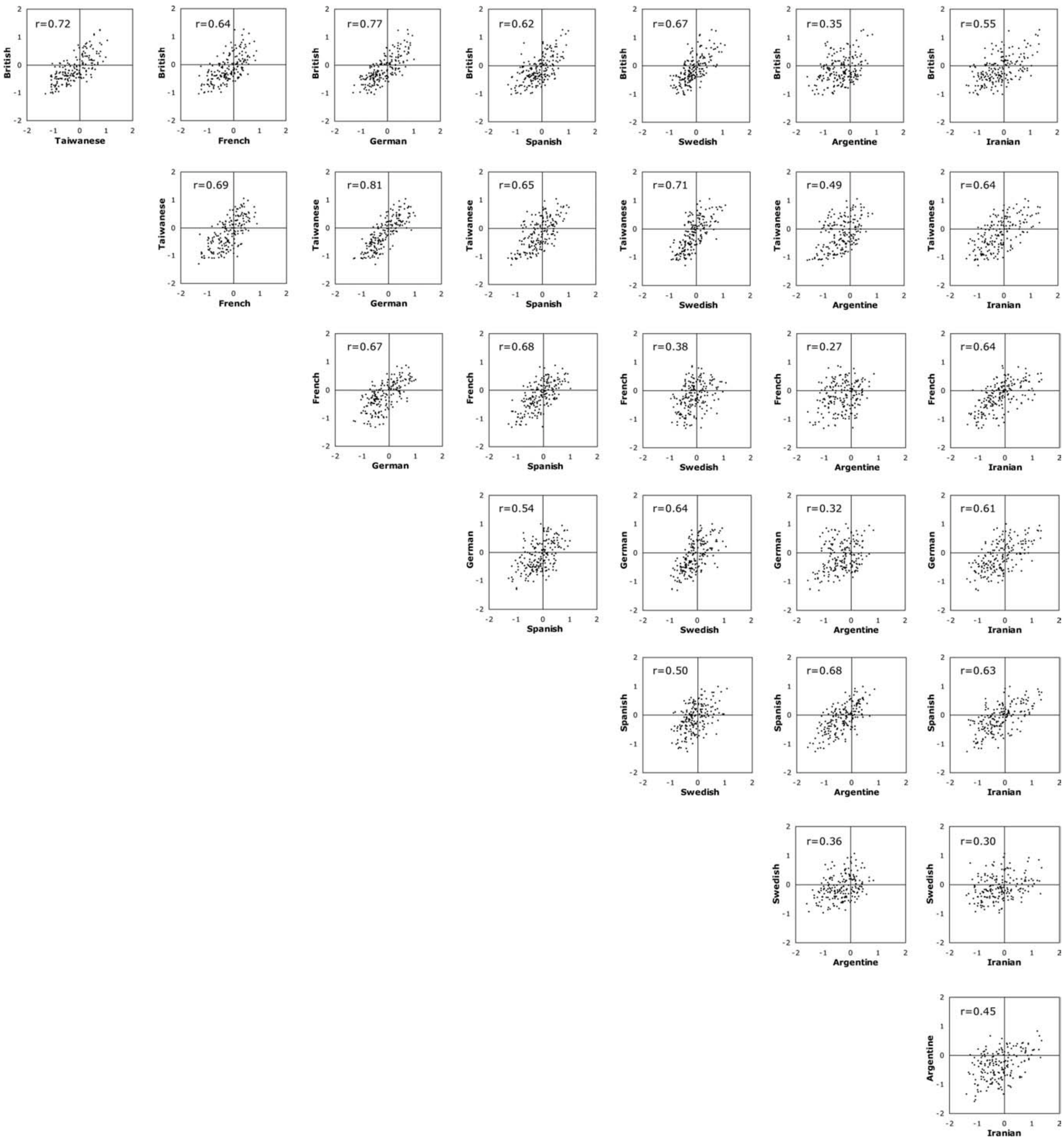


FIG. 5. Comparisons of the like-dislike responses between the eight observer groups: British, Taiwanese, French, German, Spanish, Swedish, Argentinian, and Iranian.

using the SPSS statistical analysis software. The input data were the scale values for the 190 colour pairs, with 32 variables defined by the combinations of the four colour emotion scales and the eight observer groups, i.e., the four colour emotion terms, "warm," "heavy," "active," and "like," each marked by a shortened country name for identifying each observer group: "uk" standing for British observers, "tw" for Taiwanese, "fr" for French, "de" for German, "es" for Spanish, "se" for Swedish, "ar" for Argentinian, and "ir" for Iranian.

Four principal components were extracted, PC-1 to PC-4, accounting for 89.2% of the total variance. Table VII shows component loadings of the 32 variables for each principal component. As the table demonstrates, for all observer groups, the responses of the same colour emotion term show high loadings for one principal component but low loadings for the others, e.g., the "heavy" responses for all the eight observer groups have high loadings for PC-1 only. Thus, the four principal components were named by the corresponding colour emotion

TABLE VI. Correlation coefficients between the eight observer groups for the four colour emotion scales warm/cool, heavy/light, active/passive and like/dislike.

	Taiwanese	French	German	Spanish	Swedish	Argentine	Iranian	Mean
Warm/cool								
British	0.83	0.83	0.83	0.87	0.80	0.82	0.81	0.83
Taiwanese	–	0.91	0.87	0.90	0.84	0.84	0.72	0.85
French		–	0.87	0.89	0.81	0.82	0.76	0.84
German			–	0.92	0.90	0.90	0.79	0.87
Spanish				–	0.90	0.94	0.85	0.90
Swedish					–	0.87	0.77	0.84
Argentine						–	0.84	0.86
Iranian							–	0.79
								0.85
Heavy/light								
British	0.92	0.88	0.88	0.89	0.85	0.90	0.89	0.89
Taiwanese	–	0.92	0.92	0.92	0.90	0.94	0.94	0.92
French		–	0.90	0.92	0.87	0.90	0.88	0.89
German			–	0.90	0.96	0.86	0.91	0.91
Spanish				–	0.87	0.91	0.90	0.90
Swedish					–	0.83	0.91	0.89
Argentine						–	0.89	0.89
Iranian							–	0.91
								0.90
Active/passive								
British	0.78	0.79	0.80	0.87	0.85	0.84	0.78	0.82
Taiwanese	–	0.92	0.93	0.90	0.87	0.79	0.90	0.87
French		–	0.94	0.90	0.90	0.80	0.89	0.88
German			–	0.91	0.92	0.81	0.87	0.88
Spanish				–	0.95	0.92	0.90	0.91
Swedish					–	0.90	0.86	0.89
Argentine						–	0.83	0.84
Iranian							–	0.86
								0.87
Like/dislike								
British	0.72	0.64	0.77	0.62	0.67	0.35	0.55	0.62
Taiwanese	–	0.69	0.81	0.65	0.71	0.49	0.64	0.67
French		–	0.67	0.68	0.38	0.27	0.64	0.57
German			–	0.54	0.64	0.32	0.61	0.62
Spanish				–	0.50	0.68	0.63	0.62
Swedish					–	0.36	0.30	0.51
Argentine						–	0.45	0.42
Iranian							–	0.55
								0.57

terms, i.e., "Heaviness" for PC-1, "Activity" for PC-2, "Warmth" for PC-3, and "Liking" for PC-4.

Figures 6(a) and 6(b) illustrate the interrelationships between each colour emotion term by plotting them in a semantic space based on the component loadings shown in Table VII. Each colour emotion term is represented by a dot, marked by one of the four colours: black for "heavy," red for "warm," yellow for "active," and blue for "like." Next to each dot is a shortened country name for identifying observer group.

The closer any two dots are located in the diagram, the more similar the semantic meanings of the two colour emotion terms. As shown in Fig. 6(a), dots in the same colour (except the blue dots) tend to be located next to each other. This tendency was also found in Fig. 6(b) for dots in black, red and yellow, but there is no such tendency for the blue dots. Although Fig. 6(a) does show the red "ir" dot being located relatively far away from the remaining red dots and toward to the right side (meaning that the Iranian observers were more likely than the other observers to regard heavier colour pairs as "warmer"), the overall result indicates that except for "like" (i.e., the blue dots), the colour emotion responses obtained from the eight observer groups are similar for each scale, suggesting there is little cultural difference in observers' responses for warm/cool, heavy/light, and active/passive.

Regarding the like/dislike response, however, both Figs. 6(a) and 6(b) show a larger spread between the eight blue dots, in particular the "ar" dot (i.e., the "like" response by Argentinian observers) is located furthest away from the other seven blue dots. The blue "ar" dot is located on the lower side of the diagram, whereas the other blue dots are located on the upper side or in the middle area. This suggests a high correlation between Argentinian observers' "like" response and the "passive" response, meaning that Argentinian observers tended to prefer "passive" colour pairs more than the other observers. According to an existing model of active/passive [see Eq. (A3) in Appendix B], "passive" colours are those located near medium grey (with a lightness value of 50) in the CIELAB space. In addition, the results seem to suggest that both Spanish and Argentinian observers liked "cooler" colour pairs

TABLE VII. Component loadings for the eight observer groups, British (uk), Taiwanese (tw), French (fr), German (de), Spanish (es), Swedish (se), Argentinian (ar), and Iranian (ir), for the four principal components PC-1–PC-4, which account for 89.2% of the total variance.

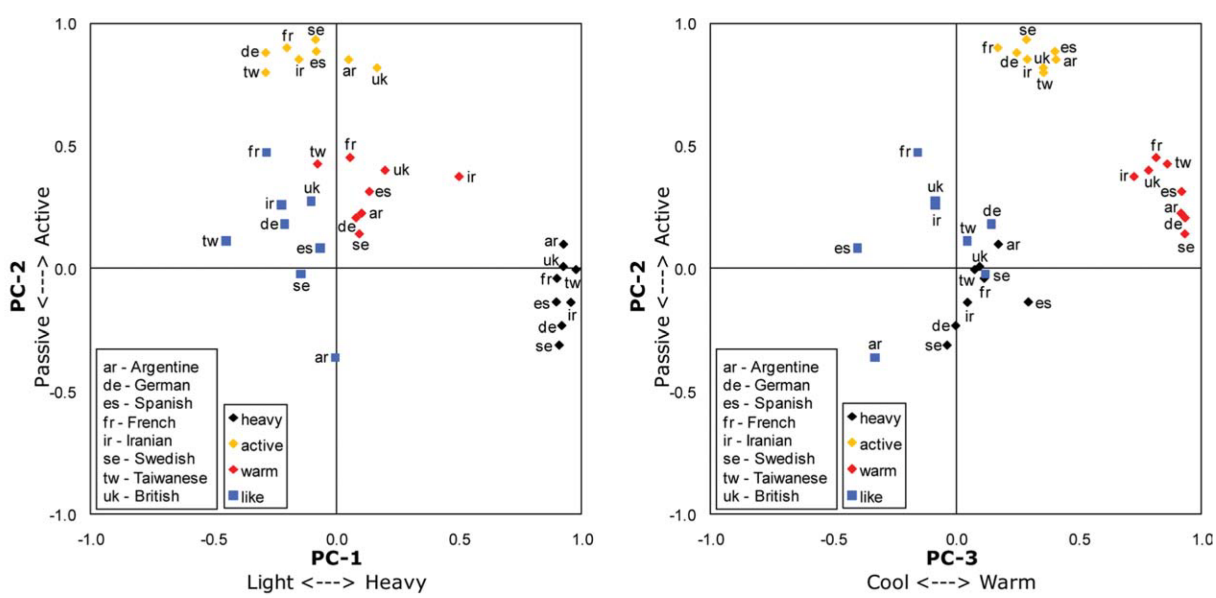
	PC-1 (heaviness)	PC-2 (activity)	PC-3 (warmth)	PC-4 (liking)
heavy_tw	0.98	−0.01	0.07	−0.09
heavy_ir	0.96	−0.14	0.04	−0.06
heavy_ar	0.93	0.10	0.17	−0.20
heavy_uk	0.93	0.01	0.09	−0.16
heavy_de	0.92	−0.24	0.00	−0.24
heavy_se	0.91	−0.31	−0.04	−0.17
heavy_fr	0.90	−0.04	0.11	−0.28
heavy_es	0.90	−0.14	0.30	−0.20
active_se	−0.08	0.94	0.29	−0.04
active_fr	−0.20	0.90	0.17	0.26
active_es	−0.08	0.89	0.40	0.08
active_de	−0.29	0.88	0.25	0.17
active_ir	−0.15	0.85	0.29	0.22
active_ar	0.05	0.85	0.40	−0.03
active_uk	0.17	0.82	0.35	0.17
active_tw	−0.29	0.80	0.35	0.33
warm_de	0.08	0.21	0.93	−0.07
warm_se	0.10	0.14	0.93	−0.12
warm_es	0.14	0.31	0.92	−0.07
warm_ar	0.10	0.22	0.92	−0.12
warm_tw	−0.07	0.43	0.86	0.08
warm_fr	0.06	0.45	0.81	0.06
warm_uk	0.20	0.40	0.78	−0.09
warm_ir	0.50	0.38	0.72	−0.12
like_tw	−0.45	0.11	0.05	0.83
like_de	−0.21	0.18	0.14	0.82
like_es	−0.06	0.08	−0.40	0.80
like_uk	−0.10	0.28	−0.08	0.78
like_ir	−0.22	0.26	−0.09	0.73
like_se	−0.15	−0.02	0.12	0.70
like_ar	0.00	−0.36	−0.33	0.68
like_fr	−0.28	0.47	−0.16	0.66

more than the other observers, as in Fig. 6(b) the blue "ar" and blue "es" dots are both located more on the left side of the graph than the other blue dots. Appendix D shows the ranking of colour pairs in the order of like/dislike for the eight observer groups.

Effect of Gender

To see whether there was any effect of gender on the colour emotion responses, the experimental data for observers of the same gender were combined and converted into *z*-score related scale values (see Appendix G for an example of calculation). The two sets of scale values, representing responses of the two gender groups (111 males and 112 females), were then compared for the four scales warm/cool, heavy/light, active/passive, and like/dislike, respectively. The results show good agreement between the two gender groups, with a correlation coefficient of 0.98 for warm/cool, 0.98 for heavy/light, 0.98 for active/passive, and 0.84 for like/dislike.

The correlation coefficient for like/dislike (0.84) stands out as the lowest among the four scales, given the fact that the other three scales all show nearly perfect correlation (0.98) between the two gender groups. This suggests that like/dislike may have been influenced by the effect of gender, despite the high correlation between the two gender groups for like/dislike (0.84). To see whether such effect was related to the appearance of each colour pair, the like/dislike response given by female group were first subtracted from those by male group, i.e., male response minus female response; the resulting values are called "score difference." The higher the score difference is for a colour pair, the more likely it is that the colour pair is



(a) (b)

FIG. 6. Component plots (a) PC-2 versus PC-1 and (b) PC-2 versus PC-3 for the observer responses, in terms of "heavy" (black dots), "warm" (red dots), "active" (yellow dots), and "like" (blue dots), given by the eight observer groups, British (uk), Taiwanese (tw), French (fr), German (de), Spanish (es), Swedish (se), Argentinian (ar), and Iranian (ir).

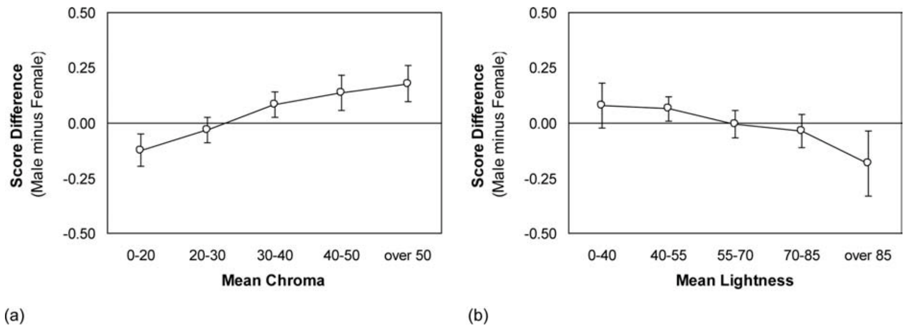


FIG. 7. Difference in the like/dislike responses between male and female observers (male score minus female score) plotted against (a) mean chroma and (b) mean lightness, which show the most significant trends among other colour appearance attributes. The error bars indicate the 95% confidence intervals.

preferred by male observers more than by female observers. The score difference values were then compared with various CIELAB values for each colour pair, such as mean and difference values of an appearance attribute (e.g., lightness) between a colour pair.

The comparison results seem to show some trends in score difference with regard to mean chroma and mean lightness. Figure 7(a) illustrates that as mean chroma value increases, the score difference becomes higher and higher. This suggests that female observers tended to like low-chroma colour pairs more than male observers. Figure 7(b) shows that as mean lightness increases, the score difference becomes lower and lower. This suggests that female observers liked high-lightness colour pairs more than male observers did. Figure A4(a) in Appendix E shows the ranking of colour pairs in the order of like/dislike for the two gender groups.

Effect of Design Training

As previously mentioned (see Table I), some observers had a first degree in a design related subject (including British, Taiwanese-1 and part of Swedish and Argentinian observers). To see whether formal training in design had any impact on the observers' responses to colour, the colour emotion scale values were calculated for the two groups of observers (i.e., design and nondesign), and the two sets of scale values were compared in terms of correlation coefficient. As a result, close correlations were found between the two sets of scale values for each colour emotion scale, with a correlation coefficient of 0.98 for warm/cool, 0.98 for heavy/light, 0.97 for active/passive, and 0.83 for like/dislike.

The correlation coefficient for like/dislike stands out again for being much lower than the other three scales, the latter three all showing nearly perfect correlation between the two professional-background groups. This suggests that like/dislike may have been influenced by the effect of professional background, despite the high correlation coefficient for like/dislike (0.83). The score difference values were again generated, calculated by subtracting the like/dislike responses of observers with nondesign backgrounds from those by observers with design backgrounds (i.e., design responses minus nondesign responses). The score difference values were then compared with CIELAB values of each colour pair.

The results seem to show some trends in score difference with regard to mean chroma and hue difference (i.e., absolute difference in the metric hue, ΔH^*_{ab}). As shown in Fig. 8(a), as mean chroma increases, the score difference becomes lower. This indicates that observers with design backgrounds liked low-chroma colour pairs more than those with nondesign backgrounds. Figure 8(b) shows that as hue difference increases, the score difference becomes lower (until the hue difference reaches around 35). This indicates that observers with design backgrounds liked

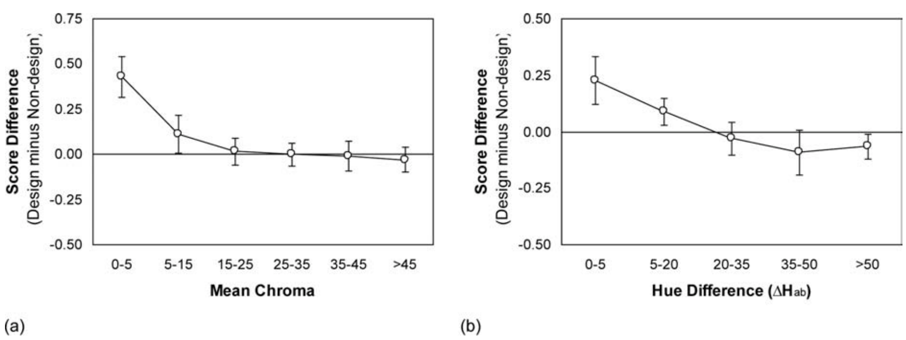


FIG. 8. Difference in the like/dislike responses between observers with and without a design background ("design" minus "nondesign") plotted against (a) mean chroma and (b) hue difference, which show the most significant trends among other colour appearance attributes. The error bars indicate the 95% confidence intervals.

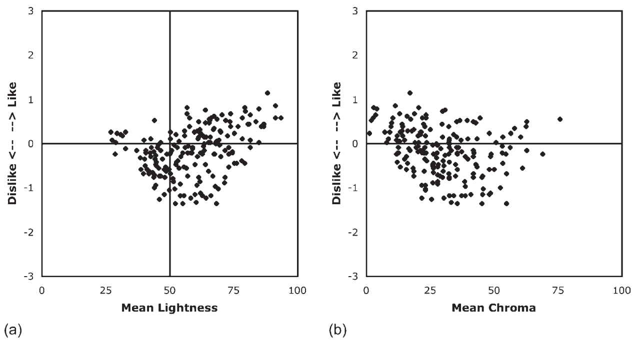


FIG. 9. Taiwanese young observers' like/dislike response to colour pairs plotted against (a) mean lightness and (b) mean chroma.

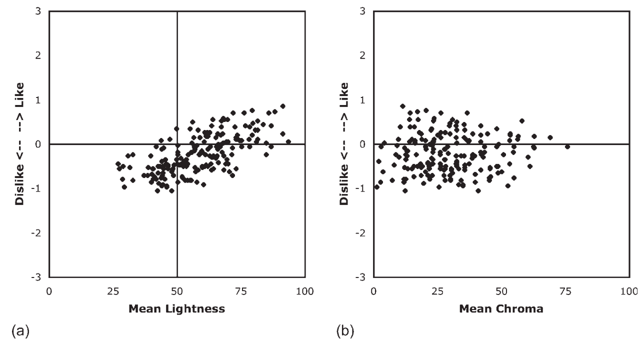


FIG. 10. Taiwanese older observers' like/dislike response to colour pairs plotted against (a) mean lightness and (b) mean chroma.

colour pairs containing colours similar in hue more than those with nondesign backgrounds did. The experimental results do not show any trend of the score difference when plotted against either lightness difference or chroma difference. Figure A4(b) in Appendix E shows the ranking of colour pairs in the order of like/dislike for the design and nondesign observer groups.

Effect of Age

It has been shown that as one ages, there is a decline in transmittance of human crystalline lens for short-wavelength light[63,64]; there is also a decline in the neural system and brain functioning for older adults.[64–66] As a result, older adults tend to have issues in the discrimination of saturation and hue.[67–70] It has been shown that older adults need more chroma to perceive the same chromatic content as perceived by young adults[71] due to the decline of colour perception intensity.[72]

As mentioned previously, Taiwanese-2 and Argentinian observer groups both consisted of young and older adults (see Table I). To see whether age had an impact on the observers' colour emotion responses, the experimental results for the young and older adults were calculated separately and were then compared in terms of correlation coefficients. Note that the two observer groups not only had differences in age but also in educational level, as the young observers were all university students, whereas the older adults were the local citizens with various educational levels.

The comparison results for Taiwanese-2 show a correlation coefficient of 0.84 for warm/cool, 0.90 for heavy/light, 0.79 for active/passive, and 0.52 for like/dislike; the Argentinian groups show a correlation coefficient of 0.81 for warm/cool, 0.82 for heavy/light, 0.81 for active/passive, and 0.60 for like/dislike. Although all these correlation coefficients were statistically significant (with the p values all under 0.01), only those for like/dislike show correlation coefficients of lower than 0.70 (corresponding to about 50% of the total variance). This suggests that like/dislike may have been influenced by the effect of age. To see whether the effect was related to the appearance of the colour pairs, the like/dislike response was plotted against the mean and difference values in terms of colour appearance attributes including lightness, chroma, and hue (in CIELAB system) for the constituent colours in each colour pair. As a result, mean lightness and mean chroma were found to show some trends related to the age effect, as summarized in Figs. 9–11 for Taiwanese observers and Figs. 12–14 for Argentinian observers.

As Figs. 9(a) and 10(a) demonstrate, both young and older observers in Taiwan preferred higher-lightness colour pairs to lower-lightness ones, but according to Fig. 11(a), the older observers showed a greater preference for the higher-lightness colour pairs, as the score difference in like/dislike (i.e., young minus older) decreases when mean lightness becomes larger. In terms of mean chroma, Figs. 9(b), 10(b), and 11(b) show that young observers tended to prefer lower-chroma colour pairs to higher-chroma ones, and this tendency was not found in the older observers' response.

For Argentinian observers, according to Fig. 12(a), the young observers seem to prefer lower-lightness colour pairs to higher-lightness ones, a trend opposite to that for young observers in Taiwan. Nevertheless, Fig. 13(a) shows that the older Argentinian observers preferred higher-lightness colour pairs to lower-lightness ones, a trend that agrees with the older observers in Taiwan. It is interesting to note that despite the different trends for

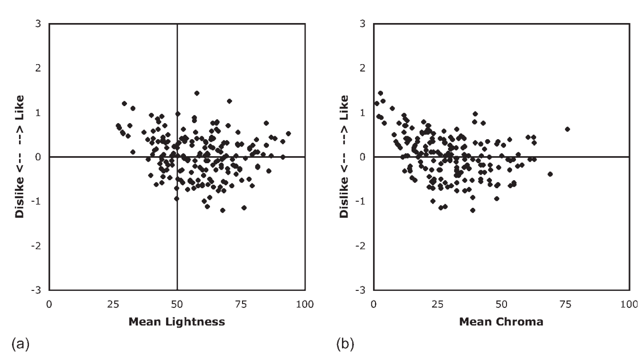


FIG. 11. Difference in the like/dislike responses to colour pairs between Taiwanese young and older observers (young minus older) plotted against (a) mean lightness and (b) mean chroma.

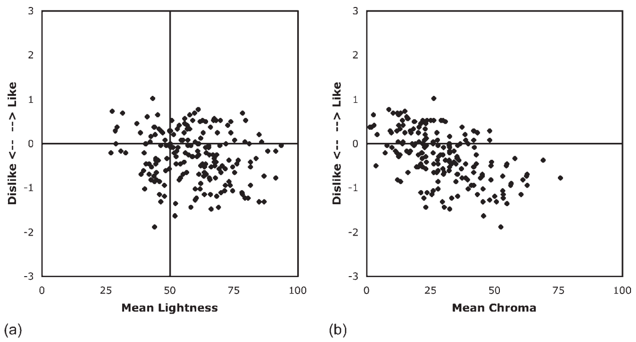


FIG. 12. Argentinian young observers' like/dislike response to colour pairs plotted against (a) mean lightness and (b) mean chroma.

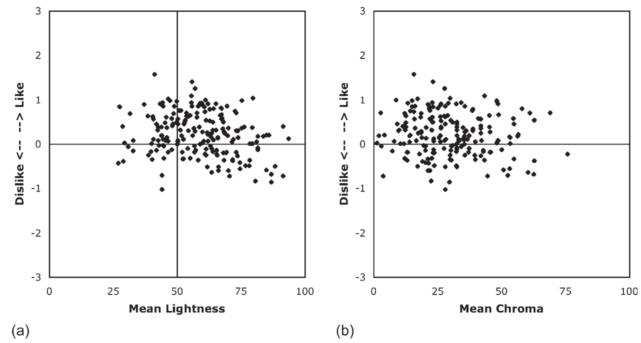


FIG. 14. Difference in the like/dislike responses to colour pairs between Argentinian young and older observers (young minus older) plotted against (a) mean lightness and (b) mean chroma.

young Taiwanese and young Argentinians [Figs. 9(a) and 12(a), respectively], the score difference between young and older observers show similar trends for the two observer groups: the score difference decreases as the mean lightness becomes larger. This means that for both Taiwanese and Argentinian, the older observers tended to prefer lighter colour pairs.

In terms of mean chroma, the comparison results also show that for both Taiwanese and Argentinian, the older observers tended to prefer higher-chroma colour pairs more than the young observers did [see Figs. 11(b) and 14(b)]. Appendix F shows the ranking of colour pairs in the order of like/dislike for the young and older observer groups using the experimental results obtained from Taiwan and Argentina.

Testing Colour Emotion Models

Ou *et al.*'s predictive models of colour emotion and preference were developed for single colours[28,30] and for colour pairs.[29] Each model was based on the CIELAB system for calculation. First, the CIELAB values for the colour concerned were used to determine the single-colour emotion values using Eqs. (A1)–(A4) as shown in Appendix B. The single-colour emotion values for component colours within each colour pair were then averaged using Eq. (A5) to predict colour emotion responses for colour pairs. The latter model illustrates the additivity of colour emotion, i.e., the colour emotion value for a colour combination can be determined by the average of all colour emotion values for individual colours in that combination.

The present experimental data were used to test how well the models work for colour stimuli viewed under different viewing conditions, as the models were derived on the basis of reflecting colour patches presented in a viewing cabinet. This study, however, used CRT colours and investigated different groups of observers, while the models were based on psychophysical responses of British and Chinese observers.

Figures 15(a)–15(c) show test results for warm/cool, heavy/light, and active/passive, respectively. The diagrams indicate good predictive performance of the models for the three scales, with a correlation coefficient of 0.94 for warm/cool, 0.90 for heavy/light, and 0.90 for active/passive. The results seem to support previous findings regarding the additivity of colour emotion.[29,40,41]

Figure 16(a) shows the test result for like/dislike using Eqs. (A4) and (A5), with a correlation coefficient of 0.64. Note that Eq. (A4) was a sinlge-colour preference model, based purely on experimental data for single colours[30]; note also that the additivity of colour emotion as illustrated by Eq. (A5) has already been found not to work well for the like/dislike response.[29] The relatively poorer correlation shown in Fig. 16(a) for like/dislike seems to support previous findings that colour preference for colour pairs cannot be predicted well by simple calculations such as arithmetic mean of single-colour emotion values,[29,73,74] i.e., the additivity of colour emotion.

It has been shown that the visual response of like/dislike is closely correlated with the harmonious/disharmonious response.[30] It would be interesting to see how well a colour harmony model predicts the like/dislike response. The model tested here was developed recently by Ou and Luo[62] as summarized in Appendix C. As a result, Fig. 16(b) shows significantly better agreement [compared with Fig. 16(a)] between the observer response (like/dislike) and the predicted value (harmonious/disharmonious), with a correlation coefficient of 0.74. Note that the colour

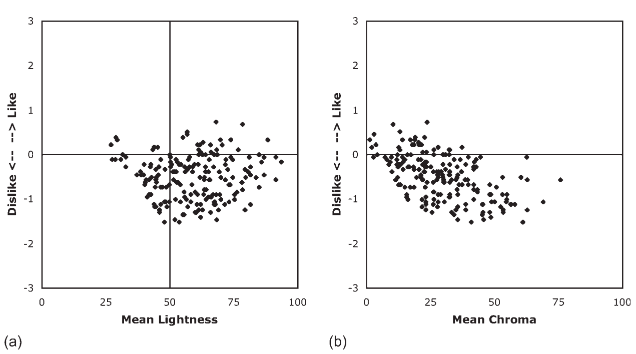


FIG. 13. Argentinian older observers' like/dislike response to colour pairs plotted against (a) mean lightness and (b) mean chroma.

harmony model tested here takes into account the colour contrast effect between the constituent colours,[62] an effect not considered in Eqs. (A4) and (A5). This suggests that to model colour-combination preference, the colour contrast effect should be included.

DISCUSSION

The experimental results show common, consistent responses across cultures for warm/cool, heavy/light, and active/passive dimensions, the three principal scales of colour emotion.[25,27,28,31,37] The finding agrees well with a recent cross-cultural study of colour emotion for single colours[31] which also found consistent responses from observers of different regions: Japan, Thailand, Hong Kong, Taiwan, Italy, Sweden, and Spain.

For like/dislike response, however, the present experimental results show strong effect of culture, especially for the comparisons between Argentinian group and the other observer groups. Factor analysis [Figs. 6(a) and 6(b)] reveals high correlation between Argentinian response for like/dislike with their active/passive response, i.e., Argentinian observers tended to prefer passive colour pairs [i.e., containing more greyish colours, see Eq. (A3) in Appendix B] than the other observers. Although the above findings are interesting, it is unclear why there are such distinct differences. Schloss and Palmer[75] argued that colour preferences "reflect people's cumulative emotional responses to objects, institutions, and events associated with those colours." It will be interesting to look into this in a broader scope by conducting tests based on images of various objects/contexts.

Effects of gender, professional background (design vs. nondesign), and age (young adults vs. older adults) were also found to be strong on the like/dislike response (Figs. 7, 8, 11, and 14). Data analysis reveals that male observers liked higher-chroma colour pairs more than female observers, a finding shared by Schloss and Palmer's study using American observers.[75] In addition, male observers were found to prefer lower-lightness colour pairs more than female observers, a result that agrees with Guilford and Smith's study[76] but disagrees with Schloss and Palmer's; the latter showed little gender difference in lightness preference.[75] In terms of hue preference, the present result shows little difference between the two gender groups, a finding that agrees with Schloss and Palmer[75] but disagrees with Hurlbert and Ling[50] who showed that female observers preferred redder colours (presented against a grey background) more than male observers. As it appears, there is a wide diversity of findings regarding gender difference in colour preference. A more thorough review of various reports will be needed to reach a consensus.

For the effect of design training, the present results show that observers with a design background liked lower-chroma colour pairs or those containing colours of similar hue more than observers with a nondesign background. It has been shown that colours that are either low in chroma or similar in hue tend to harmonize.[62,77,78] We assume that the observers of a design background are more sensitive to harmony in colour than those without any design training, and that it is, thus, easier for the former to identify harmonious colour schemes and to give higher scores for them. To test this assumption will require more empirical evidence.

For the age effect, older adults tended to like higher-lightness or higher-chroma colour pairs more than young observers did. The result agrees well with previous findings that older adults have less colour perception intensity and thus for them to perceive the same chromatic content in an image as perceived by young adults, more chroma is needed in the image for older adults.[71,72]

Test results for the existing colour emotion models, including the additivity of colour emotion [Eq. (A5) in Appendix B], show good predictive performance for warm/cool, heavy/light, and active/passive dimensions. The results not only demonstrate that the three colour emotion scales can be predicted accurately using colour-appearance-based models but they also lend support to the additivity of colour emotion previously revealed.[29,40,41]

The satisfactory predictive performance of the colour harmony model (Appendix C) for the like/dislike response suggests that although colour harmony and preference seem to be different concepts,[30] responses for the two scales are highly correlated and can be predicted well from each other. Further investigation into the relationship between colour harmony and preference will be interesting.

Cupchik[79] argued that emotional responses to aesthetic stimuli can be divided into two levels: reactive and reflective. The reactive level of emotions referred to bodily feelings of pleasure and arousal, attributed to configurations of the stimulus. The reflective level, on the other hand, referred to blends of emotions (e.g., a mixture of sadness and anger), attributed to contextual meanings of the stimulus. On the basis of the present experimental results, we propose that the two-level theory also applies to colour emotion, i.e., the responses to colour can be divided into reactive and reflective levels, and that each level has distinct properties. It is further proposed that colour emotion responses at the reactive level, such as warm/cool, heavy/light, and active/passive, are universal, culture-independent and work well with the additivity for colour combinations [Figs. 15(a)–15(c)]. The essential attributes for this level of emotion are local configurations of the stimulus,[79] which in the context of colour refer to perceptual colour appearance attributes such as hue, lightness, and chroma.

The reflective level of colour emotion responses such as like/dislike, however, is culture-dependent and does not work well with the additivity for colour combinations [Fig. 16(a)], as it involves contextual meanings of the stimulus, interrelationship between the stimulus and the viewer, and is affected by the colour contrast effect [Fig.

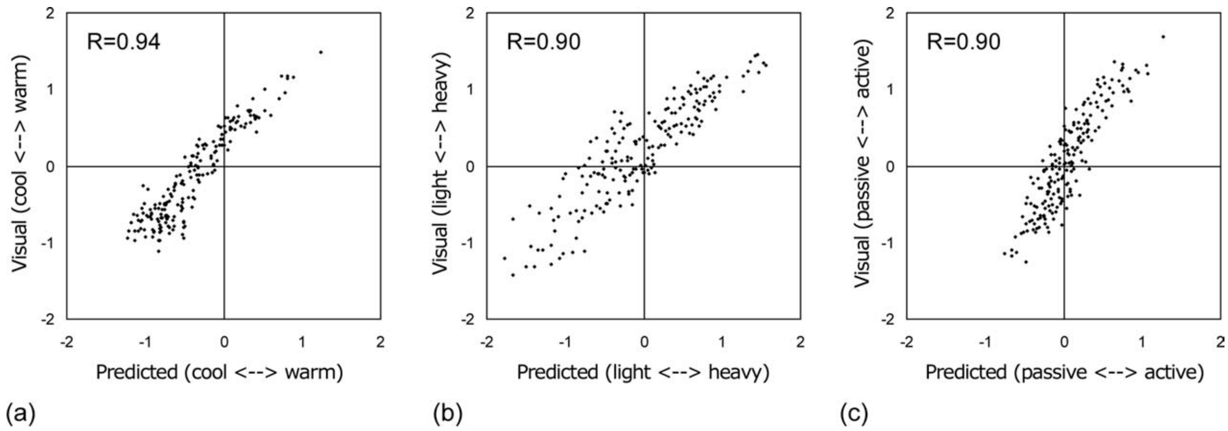


FIG. 15. Average observers' responses from all observer groups for (a) warm/cool, (b) heavy/light, and (c) active/passive plotted against the predicted values.

16(b)]. The present experimental results show large differences in observers' like/dislike response between cultures (Fig. 5), genders [Fig. 7], professional backgrounds [Fig. 8], and age groups [Figs. 11 and 14]. It is, thus, proposed here that the reflective-level colour emotion such as like/dislike is a learned,[55,75] rather than innate,[48–50] reaction.

It has been pointed out that context[79,80] and circumstance[81] play a significant role in emotional responses. Although this study is based on "contextless" judgement about colour, we believe the effect of context/circumstance on colour emotion responses would be strong. Stahre *et al*.[32] have found large differences of the reactive-level responses to textile chips and to full-scale rooms (with similar colour schemes); however, it is unclear whether the differences were due to the context (i.e., textile chips vs. real rooms), the size effect (each textile chip being 1.5 × 1.0 cm^2 in size and each room 4.2 × 2.9 m^2), or both. To clarify this will require further research in various contexts/circumstances.

There are many other factors that are likely to be influential in the colour emotion responses, such as lighting conditions,[33,82] texture effect, size effect, personality, or physiology. These will be considered in the future work.

ACKNOWLEDGMENTS

The authors thank Dr Guihua Cui for developing the experimental software used in this study. They also thank the reviewers of this article for their valuable comments and suggestions.

APPENDIX A: INSTRUCTIONS TO OBSERVERS (ENGLISH VERSION)

Thank you for participating in the visual assessment. There will be six sessions for you, each taking about 30 min. In each session, you will be presented with about 70 colour pairs shown individually on the display, as demonstrated in Fig. A1.

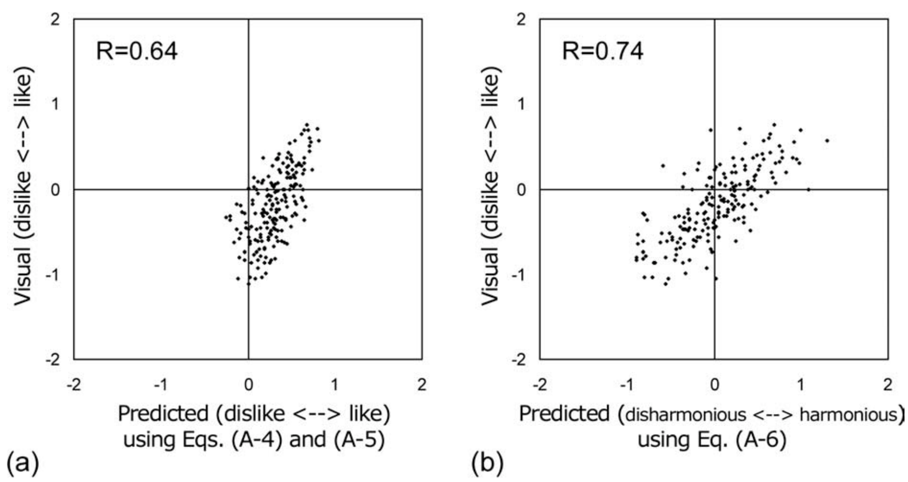


FIG. 16. Average observers' like/dislike response from all observer groups plotted against the predicted values by models for (a) like/dislike and (b) colour harmony.

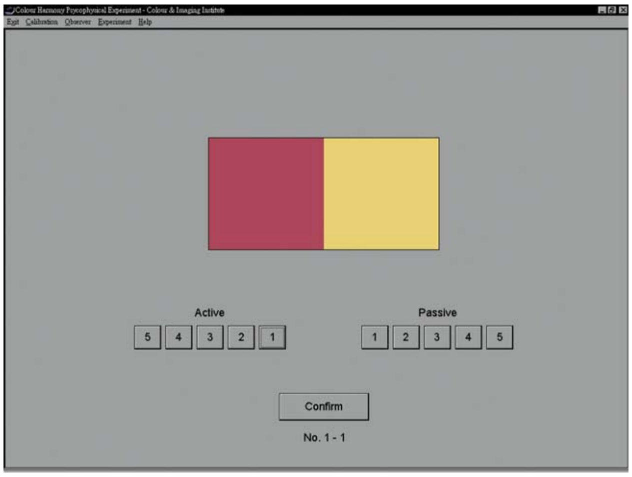


FIG. A1. Screen layout of the experiment.

Please describe each colour pair (or your impression of them) using the following word pairs:

warm–cool
heavy–light
active–passive
like–dislike

For each colour pair, give your answer by pressing one of the 10 buttons shown on the screen on the basis of the scale shown in Fig. A2 (e.g., ''active vs. passive'').

APPENDIX B: COLOUR EMOTION MODELS

Ou *et al*.[28,30] developed CIELAB-based predictive models of single-colour emotions, for scales warm/cool (WC), heavy/light (HL), active/passive (AP), and like/dislike (LD):

$$WC = -0.5 + 0.02(C^*_{ab})^{1.07}\cos(h_{ab} - 50^\circ) \qquad \text{(A1)}$$

where C^*_{ab} is CIELAB chroma and h_{ab} is CIELAB hue angle.

$$\mathrm{HL} = -2.1 + 0.05(100 - L^*) \qquad \text{(A2)}$$

where L^* is the CIE lightness.

$$\mathrm{AP} = -1.1 + 0.03\left[(C^*_{ab})^2 + (L^* - 50)^2\right]^{\frac{1}{2}} \qquad \text{(A3)}$$

where C^*_{ab} is CIELAB chroma and L^* is the CIE lightness.

$$\mathrm{LD} = -0.65 + 0.03\left[(L^* - 50)^2 + \left(\frac{a^* + 8}{2}\right)^2 + \left(\frac{b^* - 30}{1.7}\right)^2\right]^{\frac{1}{2}} \qquad \text{(A4)}$$

where L^* is the CIE lightness; a^* and b^* are CIELAB redness/greenness and yellowness/blueness, respectively.

Ou *et al*.[29] verified the additivity of colour emotion originally identified by Hogg.[40] The additivity property can be illustrated by the following equation:

$$E = (E_1 + E_2)/2 \qquad \text{(A5)}$$

where E is the colour emotion value for a colour pair; E_1 and E_2 are the colour emotion values for the two colours in that pair.

APPENDIX C: COLOUR HARMONY MODEL

Ou and Luo[62] developed a predictive model of colour harmony for two-colour combinations:

$$CH = H_{\mathrm{C}} + H_{\mathrm{L}} + H_{\mathrm{H}} \qquad \text{(A6)}$$

in which

$$H_C = 0.04 + 0.53\tanh(0.8 - 0.045\Delta C)$$

$$\Delta C = [(\Delta H^*_{ab})^2 + (\Delta C^*_{ab}/1.46)^2]^{\frac{1}{2}}$$

$$H_L = H_{\mathrm{Lsum}} + H_{\Delta L}$$

$$H_{\mathrm{Lsum}} = 0.28 + 0.54\tanh(-3.88 + 0.029L_{\mathrm{sum}}) \text{ in which } L_{\mathrm{sum}} = L^*_1 + L^*_2$$

$$H_{\Delta L} = 0.14 + 0.15\tanh(-2 + 0.2\Delta L) \text{ in which } \Delta L = |L^*_1 - L^*_2|$$

$$H_H = H_{\mathrm{SY1}} + H_{\mathrm{SY2}}$$

$$H_{\mathrm{SY}} = E_{\mathrm{C}}(H_{\mathrm{S}} + E_{\mathrm{Y}})$$

$$E_{\mathrm{C}} = 0.5 + 0.5\tanh(-2 + 0.5C^*_{ab})$$

$$H_{\mathrm{S}} = -0.08 - 0.14\sin(h_{ab} + 50^\circ) - 0.07\sin(2h_{ab} + 90^\circ)$$

$$E_{\mathrm{Y}} = [(0.22L^* - 12.8)/10] \times \exp\{(90^\circ - h_{ab})/10 - \exp[(90^\circ - h_{ab})/10]\}$$

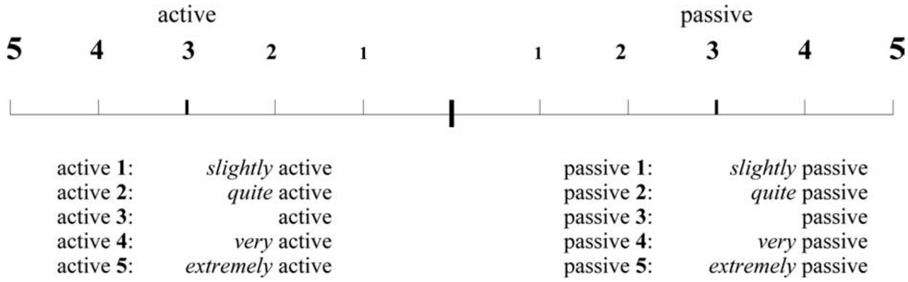


FIG. A2. The 10-step scale for measuring colour emotion responses.

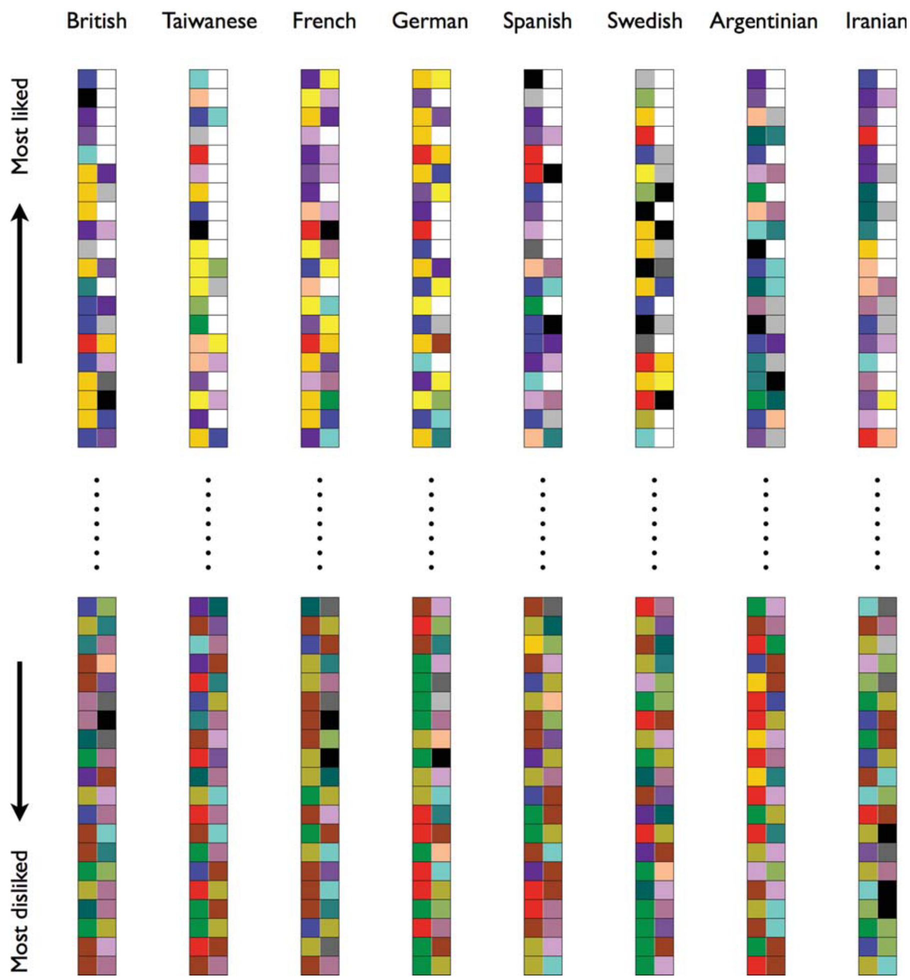


FIG. A3. Ranking of colour pairs in the order of the like/dislike response obtained from the eight observer groups.

Here L^*, C^*_{ab}, and h_{ab} represent lightness, chroma, and hue angle in CIELAB system, respectively.

APPENDIX D: RANKING RESULTS FOR THE EIGHT OBSERVER GROUPS

Figure A3 shows the top-20 and bottom-20 colour pairs ranked in the order of the like/dislike response obtained from the eight observer groups.

APPENDIX E: RANKING RESULTS FOR DIFFERENT GENDERS AND PROFESSIONAL BACKGROUNDS

Figure A4 shows the top-20 and bottom-20 colour pairs ranked in the order of the like/dislike response for (a) the two genders and (b) the two professional backgrounds (i.e., design and nondesign).

APPENDIX F: RANKING RESULTS FOR DIFFERENT AGE GROUPS

Figure A5 shows the top-20 and bottom-20 colour pairs ranked in the order of the like/dislike response for the two age groups obtained from (a) Taiwanese and (b) Argentinian observers.

APPENDIX G: WORKING EXAMPLE OF CATEGORICAL JUDGEMENT METHOD

Torgerson's Law of Categorical Judgement[61] indicates that the difference between a category boundary value (B_i) and the scale value (x_j) for a stimulus is a random variable, whose probability density function forms a normal distribution. This can be illustrated by Eq. (A7).

$$B_i - x_j = z_{ij}\sqrt{\sigma_i^2 + \sigma_j^2 - 2\rho_{ij}\sigma_i\sigma_j} \quad \text{(A7)}$$

where z_{ij} is the normal deviate (z-score) corresponding to the proportion of times stimulus j is sorted below B_i; σ_i and σ_j are standard deviation values for category boundary i and stimulus j, respectively; ρ_{ij} is correlation coefficient for category boundary i and stimulus j.

In Condition D,[61] all stimuli are assumed to share the same standard deviation among observations, i.e., $\sigma_i = \sigma_j$

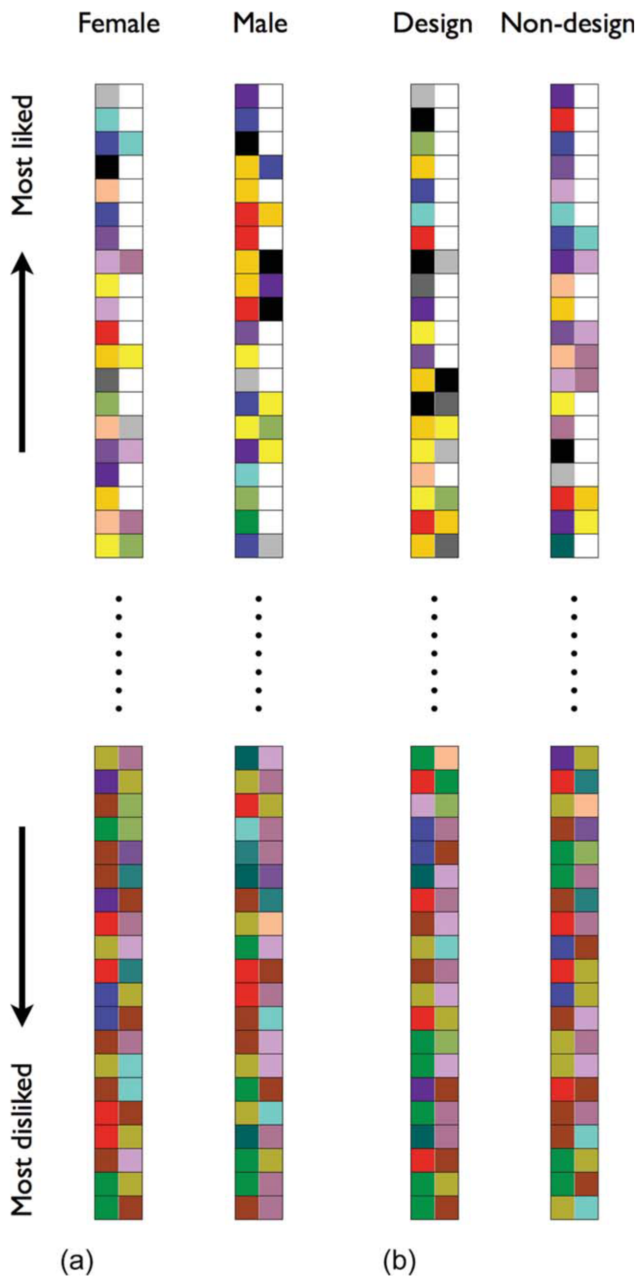


FIG. A4. Ranking of colour pairs in the order of observers' like/dislike response for (a) two genders and (b) two professional backgrounds.

$= \sigma$, and that stimulus scale values are independent of category boundaries, i.e., $\rho_{ij} = 0$. Thus Eq. (A7) can be simplified into

$$B_i - x_j = z_{ij}\sigma\sqrt{2} \qquad \text{(A8)}$$

Table AI shows a working example of the categorical judgement method using a simplified dataset:

a. Establish a frequency matrix for m categories and n stimuli. Each entry shows the frequency of a stimulus being classified into a category. From this a cumulative frequency matrix, a cumulative probability matrix and a z-score matrix are generated. For probability values of 0 or 1, the z-scores can be estimated using the Maxwell function.[83]
b. The difference matrix shows the z-score difference between category i and category $i+1$, which is equal to the difference between category boundaries B_i and B_{i+1}, as illustrated by Eq. (A9). The mean value of $B_{i+1} - B_i$ for all stimuli represents the range of the category.

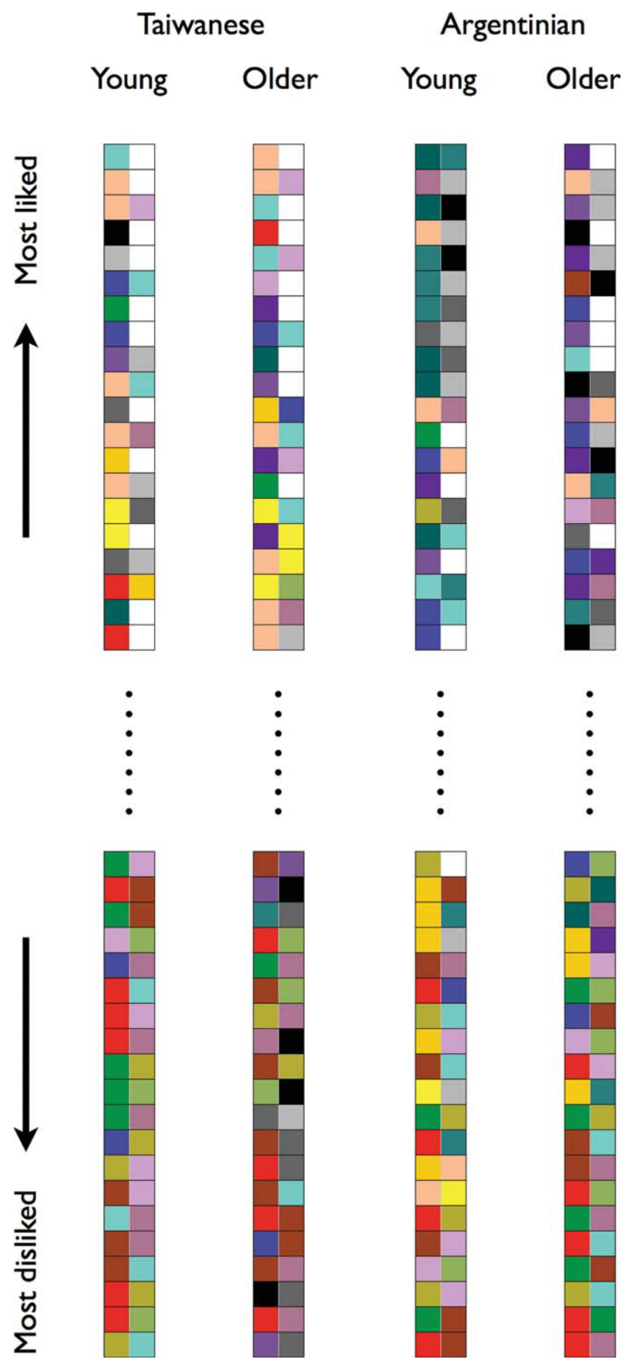


FIG. A5. Ranking of colour pairs in the order of the like/dislike response for the two age groups obtained from (a) Taiwanese and (b) Argentinian observers.

TABLE AI. A working example of the categorical judgement method.

	Category 1	Category 2	Category 3	Category 4	Category 5
Frequency matrix					
Stimulus 1	0	0	0	8	2
Stimulus 2	0	1	5	4	0
Stimulus 3	1	4	4	1	0
Cumulative frequency matrix					
Stimulus 1	0	0	0	8	10
Stimulus 2	0	1	6	10	10
Stimulus 3	1	5	9	10	10
Cumulative probability matrix					
Stimulus 1	0.0	0.0	0.0	0.8	1.0
Stimulus 2	0.0	0.1	0.6	1.0	1.0
Stimulus 3	0.1	0.5	0.9	1.0	1.0
Z-score matrix					
Stimulus 1	−2.11	−2.11	−2.11	0.85	2.11
Stimulus 2	−2.11	−1.28	0.25	2.11	2.11
Stimulus 3	−1.28	0.00	1.28	2.11	2.11
	Category 2-1	Category 3-2	Category 4-3		
Difference matrix					
Stimulus 1	0.00	0.00	2.96		
Stimulus 2	0.83	1.53	1.86		
Stimulus 3	1.28	1.28	0.83		
Mean	0.70	0.94	1.88		
Category boundary estimates					
	Category 1	Category 2	Category 3	Category 4	
	0.00	0.70	1.64	3.52	
Scale values					Mean
Stimulus 1	2.11	2.81	3.75	2.67	2.84
Stimulus 2	2.11	1.98	1.39	1.41	1.72
Stimulus 3	1.28	0.70	0.36	1.41	0.94

$$z_{(i+1)j} - z_{ij} = (B_{i+1} - x_j) - (B_i - x_j) = B_{i+1} - B_i \quad \text{(A9)}$$

c. Each scale value is then determined by averaging the difference between a z-score and the corresponding category boundary value.

1. Henderson P, Cote J. Guidelines for selecting or modifying logos. J Market 1998;62:14–30.
2. Hem L, Iversen N. How to develop a destination brand logo: A qualitative and quantitative approach. Scand J Hospitality Tourism 2004;4:83–106.
3. Tsai H, Chou J. Automatic design support and image evaluation of two-coloured products using colour association and colour harmony scales and genetic algorithm. Comput-Aided Des. 2007;39:818–828.
4. Hsiao S, Chiu F, Hsu H. A computer-assisted colour selection system based on aesthetic measure for colour harmony and fuzzy logic theory. Color Res Appl 2008;33:411–423.
5. Bellizzi JA, Crowley AE, Hasty RW. The effects of color in store design. J Retailing 1983;59:21–45.
6. Babin BJ, Hardesty DM, Suter TA. Color and shopping intentions: The intervening effect of price fairness and perceived affect. J Business Res 2003;56:541–551.
7. Gorn GJ, Chattopadhyay A, Yi T, Dahl DW. Effects of color as an executional cue in advertising: they are in the shade. Manag Sci 1997;43:1387–1400.
8. Madden TJ, Hewett K, Roth MS. Managing images in different cultures. J Int Market 2000;8:90–107.
9. Aslam MM. Are you selling the right colour? A cross-cultural review of colour as a marketing cue. J Market Commun 2006;12:15–30.
10. Dorcus RM. Color preferences and color associations. J Genet Psychol 1926;33:399–434.
11. Birren F. Color & Human Response. New York: Wiley; 1978. p 9–11.
12. Caivano JL. Color and semiotics: A two-way street. Color Res Appl 1998;23:390–401.
13. Lewis M, Haviland-Jones JM, Barrett LF, editors. Handbook of Emotions, 3rd ed. New York: Guilford Press; 2008. p 305–307.
14. Valdez P, Mehrabian A. Effects of color on emotions. J Exp Psychol Gen 1994;123:394–409.
15. Kwallek N, Lewis CM, Lin-Hsiao JWD, Woodson H. Effects of nine monochromatic office interior colors on clerical tasks and worker mood. Color Res Appl 1996;21:448–458.
16. Mehta R, Zhu R. Blue or red: Exploring the effect of color on cognitive task performances. Science 2009;323:1226–1229.
17. Ruys KI, Stapel DA. Emotion elicitor or emotion messenger: Subliminal priming reveals two faces of facial expressions. Psychol Sci 2008;19:593–600.
18. D'Andrade R, Egan M. The colors of emotion. Am Ethnol 1974;1:49–63.
19. Terwogt MM, Hoeksma JB. Colors and emotions: Preferences and combinations. J Gen Psychol 1995;122:5–17.
20. Clarke T, Costall A. The emotional connotations of color: A qualitative investigation. Color Res Appl 2008;33:406–410.
21. Oyama T, Tanaka Y, Chiba Y. Affective dimensions of colors: A cross-cultural study. Jpn Psychol Res 1962;4:78–91.
22. Wright B, Rainwater L. The meanings of color. J Gen Psychol 1962;67:89–99.
23. Adams FM, Osgood CE. A cross-cultural study of the affective meaning of color. J Cross-Cult Psychol 1973;4:135–156.
24. Taft C. Color meaning and context: Comparisons of semantic ratings of colors on samples and objects. Color Res Appl 1997;22: 40–50.
25. Kobayashi S. The aim and method of the color image scale. Color Res Appl 1981;6:93–107.
26. Kuo W. The feasibility of establishing new color image scales using the magnitude estimation method. Color Res Appl 2007;32:463–468.
27. Sato T, Kajiwara K, Hoshino H, Nakamura T. Quantitative Evaluation and categorising of human emotion induced by colour. Adv Color Sci Technol 2000;3:53–59.

28. Ou L, Luo MR, Woodcock A, Wright A. A study of colour emotion and colour preference. I. Colour emotions for single colours. Color Res Appl 2004;29:232–240.
29. Ou L, Luo MR, Woodcock A, Wright A. A study of colour emotion and colour preference. II. Colour emotions for two-colour combinations. Color Res Appl 2004;29:292–298.
30. Ou L, Luo MR, Woodcock A, Wright A. A study of colour emotion and colour preference. III. Colour preference modelling. Color Res Appl 2004;29:381–389.
31. Gao X, Xin JH, Sato T, Hansuebsai A, Scalzo M, Kajiwara K, Guan S, Valldeperas J, Lis MJ, Billger M. Analysis of cross-cultural color emotion. Color Res Appl 2007;32:223–229.
32. Stahre B, Hårleman M, Billger M. Colour emotions in larger and smaller scale. Proceedings of Interim Meeting of the Association Internationale de la Coleur (AIC) "Color and Paints," Porto Alegre, Brazil, 2004. p 27–30.
33. Hårleman M, Werner I, Billger M. Significance of colour on room character: Study on dominantly reddish and greenish colours in north- and south-facing rooms. Color: Des Creativity 2007;1:1–15.
34. Hutchings JB. The continuity of colour, design, art and science. I. The philosophy of the total appearance concept and image measurement. Color Res Appl 1995;20:296–306.
35. Hutchings JB. Expectations and the Food Industry: The Impact of Color and Appearance. New York: Kluwer/Plenum Publishers; 2003.
36. Wei S, Ou L, Luo MR, Hutchings JB. Quantification of the relations between juice colour and consumer expectations. Proceedings of the 11th Congress of the Association Internationale de la Coleur (AIC), Sydney, Australia, 2009.
37. Lee W, Luo MR, Ou L. Assessing the affective feelings of two- and three-dimensional objects. Color Res Appl 2009;34:75–83.
38. Gescheider GA. Psychophysics: The Fundamentals. New Jersey: Lawrence Erlbaum Associates; 1997.
39. Commission Internationale de l'Éclairage CIE Publ. 15:2004. Colorimetry, 3rd ed. Vienna: CIE Central Bureau; 2004.
40. Hogg J. The prediction of semantic differential ratings of color combinations. J Gen Psychol 1969;80:141–152.
41. Sueeprasan S, Srimork PH. An investigation of colour emotions using two-colour combinations. Proceedings of the 10th Congress of the Association Internationale de la Coleur (AIC), Granada, Spain, 2005. p 271–274.
42. Hogg J. A principal component analysis of semantic differential judgements of single colors and color pairs. J Gen Psychol 1969;80:129–140.
43. Hogg J, Goodman S, Porter T, Mikellides B, Preddy DE. Dimensions and determinants of judgements of colour samples and a simulated interior space by architects and non-architects. Br J Psychol 1979;70:231–242.
44. Sivik L. Research on the meanings of color combinations. Proceedings of the 6th Congress of the Association Internationale de la Coleur (AIC), Buenos Aires, Argentina, 1989. p 130–132.
45. Eysenck HJ. A critical and experimental study of colour preferences. Am J Psychol 1941;54:385–394.
46. Saito M. Comparative studies on color preference in Japan and other Asian regions: With special emphasis on the preference for white. Color Res Appl 1996;21:35–49.
47. Shen NC. The color preference of 1368 Chinese students, with special reference to the most preferred color. J Soc Psychol 1937;8: 185–204.
48. Humphrey N. Colour and brightness preferences in monkeys. Nature 1971;229:615–617.
49. Humphrey N. The colour currency of nature. In: Porter T, Mikelides B, editors. Colour for Architecture. London: Studio Vista; 1976. p 95–98.
50. Hurlbert AC, Ling Y. Biological components of sex differences in color preference. Curr Biol 2007;17:623–625.
51. Garth TR. The color preference of five hundred fifty-nine full blood Indians. J Exp Psychol 1922;5:392–418.
52. Choungourian A. Colour preferences and cultural variation. Percept Mot Skills 1968;26:1203–1206.
53. Shoyama S, Tochihara Y, Kim J. Japanese and Korean ideas about clothing colors for elderly people: Intercountry and intergenerational differences. Color Res Appl 2003;28:139–150.
54. Hupka RB, Zaleski Z, Otto J, Reidl L Tarabrina NV. The colors of anger, envy, fear, and jealousy: A cross-cultural study. J Cross Cult Psychol 1997;28:156–171.
55. Crozier WR. The psychology of colour preference. Rev Prog Color 1996;26:63–72.
56. Whitfield A, Wiltshire J. Design training and aesthetic evaluation: An intergroup comparison. J Environ Psychol 1982;2:109–117.
57. Ishihara S. Ishihara's Tests for Colour Deficiency, 38 Plates Edition. Tokyo: Kanehara Trading; 2003.
58. Berns RS. Methods for characterizing CRT displays. Displays 1996;16:173–182.
59. Berns RS, Billmeyer and Saltzman's Principles of Color Technology, 3rd ed. New York: Wiley; 2000.
60. McCamy CS, Marcus H, Davidson JG. A color-rendition chart. J Appl Photogr Eng 1976;2:95–99.
61. Torgerson WS. Theory and Methods of Scaling. New York: Wiley; 1958.
62. Ou L, Luo MR. A colour harmony model for two-colour combinations. Color Res Appl 2006;31:191–204.
63. Pokorny J, Smith VC, Lutze M. Aging of the human lens. Appl Opt 1987;26:1437–1440.
64. Weale RA. Age and transmittance of the human crystalline lens. J Physiol 1988;395:577–587.
65. Werner JS. Visual problems of the retina during aging: Compensation mechanisms and colour constancy across the life span. Prog Retin Eye Res 1996;15:621–645.
66. Werner JS, Peterzell DH, Scheetz AJ. Light, vision, and aging. Optom Vis Sci 1990;67:214–229.
67. Verriest G. Further studies on acquired deficiency of color discrimination. J Opt Soc Am 1963;53:185–195.
68. Smith VC, Pokorny J, Pass AS. Color-axis determination on the Farnsworth-Munsell 100-hue test. Am J Ophthalmol 1985;100:176–182.
69. Knoblauch K, Saunders F, Kusuda M, Hynes R, Podgor M, Higgins KE, de Monasterio FM. Age and illuminance effects in the Farnsworth-Munsell 100-hue test. Appl Opt 1987;26:1441–1448.
70. Cooper BA, Ward M, Gowland CA, McIntosh JM. The use of the lanthony new color test in determining the effects of aging on color vision. J Gerontol 1991;46:320–324.
71. Beke L, Kutas G, Kwak Y, Sung GY, Park D, Bodrogi P. Color preference of aged observers compared to young observers. Color Res Appl 2008;33:381–394.
72. Schefrin BE, Werner JS. Age-related changes in the color appearance of broadband surfaces. Color Res Appl 1993;18:380–389.
73. Guilford JP. The prediction of affective values. Am J Psychol 1931;43:469–478.
74. Lo C. The affective values of color combinations. Am J Psychol 1936;48:617–624.
75. Schloss KB, Palmer SE. An ecological valence theory of human color preferences. Proceedings of the 31st Annual Conference of the Cognitive Science Society, Amsterdam, Netherlands, 2009.
76. Guilford JP, Smith PC. A system of color-preferences. Am J Psychol 1959;73:487–502.
77. Szabó F, Bodrogi P, Schanda J. Experimental modeling of colour harmony. Color Res Appl 2010;35:34–49.
78. Lee Y, Kobayasi M, Sobagaki H. Metric quantities representing beauty of two-color combinations. Proceedings of the 11th Congress of the Association Internationale de la Coleur (AIC), Sydney, Australia, No. 78, 2009.
79. Cupchik GC. Emotion in aesthetics: Reactive and reflective models. Poetics 1994;23:177–188.
80. Norman RD, Scott WA. Colour and affect: A review and semantic evaluation. J Gen Psychol 1952;46:185–233.
81. Heise DR, Calhan C. Emotion norms in interpersonal events. Social Psychol Q 1995;58:223–240.
82. Tangkijviwat U, Rattanakasamsuk K, Shinoda H. Color preference affected by mode of color appearance. Color Res Appl 2010;35:50–61.
83. Maxwell AE. The logistic transformation in the analysis of paired-comparison data. Br J Math Stat Psychol 1974;27:62–71.