ABSTRACT

The present research constructed a complexity scale of color combinations and examined whether the scale can account for aesthetic judgments. Twenty-eight Japanese undergraduate students rated 96 color combinations on scales of complexity, pleasantness, and interestingness. In two factors of composing color combinations, complexity was defined by both factors of the color relations (similarity or contrast between component colors) and of the number of colors. This confirms that the subjective complexity scale in color combinations is well explained by compositions of these two factors. Pleasantness showed a relationship with complexity suggestive of an inverted-U function in mean ratings; however, the relationship between interestingness and complexity was linear. These relations suggest that the complexity scale is closely related to the aesthetic judgments.

The purpose of this research was to determine whether the pleasantness and interestingness of certain color combinations relate to their subjective complexity. Although the subjective complexity of an object has been found to be an important element in determining aesthetic judgments of that object, to date, research has focused more on object shapes than upon an object’s color properties. Attneave’s research in 1957, which provided a complexity scale for random shapes, established that the subjective complexity of random shapes could be explained, to a
certain extent, in terms of physical quantities, such as the number of independent
turns and the angular variability of shapes. Later, Munsinger and Kessen (1964)
found both inverted-U and linear relationships between perceived complexity
and preference, respectively, of random shapes. Little is known, however, about
the influence of complexity based on color combinations (i.e., color stimuli) on
aesthetic judgments. The present research was designed to construct a complexity
scale for color combinations and to evaluate its applicability to aesthetic judg-
ments of variously colored stimuli.

Past research has confirmed that evaluative judgments of aesthetic stimuli are
multidimensional in nature. For example, based upon results of factor analyses
from 17 investigations—dealing with both visual and auditory stimuli—
performed by various researchers, Berlyne (1974) identified five dimensions
of linguistic aesthetic responses: a hedonic tone factor, an uncertainty factor,
an arousal factor, interestingness, and arousal scales. Here, hedonic tone (an
inclusive term, proposed by Berlyne, which includes pleasantness, preference,
and beauty, among other judgments) and interestingness were considered as
different dimensions. Other research, which examined affective scales for color
combinations (Ohmi, Harada, Matsuo, & Ikeda, 1997), found that the scale of
interestingness was quite distinct from the scales of pleasantness, beauty and
likability. In traditional aesthetics, unity in variety has been considered one of
the most representative aesthetic form principles. Fechner, who was the original
proponent of experimental aesthetics, enumerated the concept of unity in variety
in one of the six principles concerning aesthetic impressions in his "Vorschule
der Ästhetik [Elementary Aesthetics]" (as cited in Berlyne, 1971, p. 127). Accord-
ing to this concept, it is possible to assume that pleasantness corresponds to
the notion of unity and interestingness to variety. Consistent with this idea,
Ohmi (1982), who examined the psychological structure of evaluation in chil-
dren’s paintings, found that two scales that measured interestingness and variety
formed a common factor that was quite distinct from a factor that included the
scales of harmony and stability. In fact, the terms harmony, pleasantness and
preference have often been used synonymously, and these three judgments have
been considered an evaluative factor in various semantic differential studies on
color combinations (e.g., Murakami, Akiyama, Asaka, Morimoto, & Ohmi, 1998;
Ohmi et al., 1997; Oyama, Soma, Tomiie, & Chijiwa, 1965).

Berlyne (1971, 1974) argued that when observing visual stimuli, such as
works of art, there is an inverted-U function between subjective complexity and
hedonic tone, and a positive linear relationship between complexity and inter-
estingness. Since that time, research concerned with the relationship between
complexity and aesthetic judgments has proliferated. For example, with regard
to hedonic tone, Nasar (2002) found an inverted-U relationship between attrac-
tiveness and complexity when using photographs of buildings as stimuli.
Hekkert and van Wieringen (1990) also found an inverted-U relationship
between beauty and complexity but only for abstract cubist paintings of human
figures. A similar result was reported by Russell and Milne (1997), although these researchers pointed out that the relationship between hedonic value and complexity exists only for relatively abstract (i.e., less representative) stimuli. The relationship between complexity and interestingness has also been explored by researchers (Avital & Cupchik, 1998; Berlyne, 1974; Russell & Milne, 1997) using abstract visual stimuli; they predominantly found a positive linear relationship.

However, findings that conflict with Berlyne’s idea has also been observed. For instance, Walker (1971) found a negative linear relationship with complexity to pleasantness, and an inverted-U relationship to interestingness in the case of printed word stimuli. Avital and Cupchik (1998) found that both pleasantness and interestingness of abstract paintings significantly correlated with complexity. Therefore, findings that address the effects of complexity on pleasantness and interestingness are generally mixed. One reason for this may be that too many factors have been included in previous studies. In an attempt to clarify the issue, the present research investigated only pleasantness and interestingness as aesthetic judgments, and used arrays of color combinations as stimuli because of their simpler structures compared with art works.

Research on color pleasantness has a long history and much is already known based on investigations of color preference and color harmony theory (see Whitfield & Wiltshire, 1990). Typically, color preference is ascertained using ratings of likeability, pleasantness and beauty, etc. of color samples. Therefore, we may assume that such ratings are measuring the intensity of hedonic tone for each color. According to Guilford (1940), color preference depends on hue, tint, and chroma. In addition, harmony models of two-color combinations have been proposed on the basis of the distance between each component color in the color solid (e.g., Nayatani, Tsujino, Ikeda, & Namba, 1969; Ou & Luo, 2006). Formulas developed to predict color harmony are, however, complicated and they fail to simulate the underlying psychological process of evaluation. Moon and Spencer’s (1944) theory of harmony adopted Birkhoff’s (1933) formula which defined the above-mentioned concept of unity in variety as \( \text{aesthetic merit} = \frac{\text{order}}{\text{complexity}} \); however, they used physical indices such as the number of colors and difference in color attributes between component colors to estimate complexity. Thus, Moon and Spencer’s theory does not necessarily reflect subjective complexity. Furthermore, this kind of research typically has utilized stimuli consisting of two- or three-color combinations, and rarely more than four-color combinations. In addition, past research has not used color combination stimuli to examine interestingness. Therefore, the present research utilized stimuli composed of multi-color combinations, which were expected to shed light on the contradictory results of past research.

In summary, the present research addressed the following three research questions. First, what kind of physical or perceptual factors influence a subjective complexity scale of color combinations? Past research has—when defining a
complexity scale of abstract shapes—typically been based in large part on the number of elements and irregularities (i.e., variety) in the shapes’ structural organization. Specifically, the present research examined two factors, namely, the number of colors contained in a given color combination (corresponding to the number of elements in a visual display) and a color relations factor involving the similarity or contrast among the component colors in a color combination (corresponding to irregularities). We predicted that these two factors could, as is the case with random shapes, explain the ratings on a complexity scale of color combinations. Second, how are color combinations judged as pleasant or interesting and what kinds of relationships exist between these two types of judgments? Finally, what role does complexity as measured by a complexity rating scale play in aesthetic judgments? That is, can a complexity scale for color combinations account for judgments of pleasantness and interestingness?

METHOD

Participants

Twenty-eight Japanese undergraduate students (6 male, 22 female, mean age 22.0 years, SD = 5.67) untrained in visual arts, participated in the experiment.

Stimuli

There were 96 color combinations utilized as stimuli. Each color combination was based upon a square that was divided into 36 equi-area cells (see Figure 1a). Each cell was filled with one of 48 chromatic colors systematically and exhaustively extracted from the Practical Color Co-ordinate System (PCCS) (Japan Color Research Institute, 1982; see Nayatani, 2003). This color order system consists of the two dimensions of 24 hues for 12 different chromatic tones. In this experiment, 12 hues by 4 tones were used (see Figures 1b and 1c).

The composition of each of the 96 stimuli was based on a combination of two factors: a number of colors factor and a color relations factor. The number of colors factor was consisted of six levels (2, 4, 6, 8, 10, and 12 colors) with 16 stimuli at each level. The color relations factor consisted of the following four levels: (a) different hues but with the same tone combination dominated by a tone (same tone combination, including 4 tones, viz., 4 categories: light, vivid, dull, and dark); (b) different tones but similar hues combination dominated by a hue attribute (similar hues combination, including 4 hue attributes, viz., 4 categories: Red, Yellow/Green, Blue and Purple/Violet); (c) different tones and different hues combination, centering on contrast combination (contrast combination, 1 category); and, (d) random combination in which the component colors were randomly selected via random number table (1 category). Table 1 describes levels of the color relations factor. All together there were six levels
Figure 1. Examples of the stimuli and the conceptual diagrams of Tone and Hue.
of the number of colors factor and four levels (including 10 categories) of the color relations factor.

A random number table determined the colors used in each color combination under the above-mentioned conditions and its arrangements. In order to remove any artificial arrangement effect, each color was allocated to each cell at random within each of the color relations and the frequency of use of each of the 48 colors was made approximately equal.

Rating Scales

Three 9-point bipolar rating scales were used:

Pleasantness: unpleasant—pleasant (fukai-na—kokoroyoi)
Interestingness: uninteresting—interesting (omoshiroku-nai—omoshiroi)

Procedure

Participants rated the complexity of 100 color combination stimuli (including four identical stimuli) in a lecture room at the same time. Stimuli were projected on a screen in random order by the experimenter and remained visible until all participants finished rating a given stimulus. The visual angle for participants was approximately 25 degrees. Participants completed complexity ratings for all stimuli, and then they rated pleasantness and lastly rated interestingness in

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same tone</td>
<td>The tone was kept constant at light, vivid, dull, or dark tone in Figure 1b. Hues were chosen randomly within each applicable tone.</td>
</tr>
<tr>
<td>Similar hues</td>
<td>Hue was kept constant at Red, Yellow/Green, Blue, or Violet/Purple in Figure 1c. Tones were chosen randomly within each applicable hue.</td>
</tr>
<tr>
<td>Contrast</td>
<td>Colors were chosen from nonadjacent hues and tones in Figures 1b and 1c.</td>
</tr>
<tr>
<td>Random</td>
<td>Colors were chosen randomly from 48 color samples.</td>
</tr>
</tbody>
</table>

Note: Each level of composition consisted of 24 stimuli (including 4 stimuli in each number of colors).
the same experimental session. For each scale, the stimuli were presented in the same random order. Participants required about 20 minutes to complete ratings for each scale.

RESULTS

To confirm whether the participants’ ratings were stable, the difference between points of the scales for the two ratings of each of the four identical stimuli (12 sets of ratings; the maximum possible disagreement was 108 points) were used to calculate an inconsistency ratings ratio for each participant. On the basis of these data, two participants were excluded due to insufficient rating stability (both of 19%; mean consistency ratio for all participants including ratings of the two participants in question was 10%). The following analyses were performed on the data for the remaining 26 participants, of whom 6 male and 20 female (M age of 22.2 years and SD of 5.76).

Stability and Discriminability

To assess the stability of each set of ratings (complexity, pleasantness, or interestingness), the ratio of inconsistency of ratings was calculated for each scale. It was determined by dividing the total number of disagreement points between the scales of the two ratings for each of four identical stimuli by 36-points of possible maximum disagreement points (9-points by four stimuli). The resulting ratio was lower for the complexity scale (7%) than for the two aesthetic judgment scales (10% and 14% for pleasantness and interestingness, respectively). Therefore, the judgment of complexity was quite stable; stability was somewhat less for pleasantness and interestingness.

The degree to which participants were able to differentiate stimuli along each of the three scales (complexity, pleasantness, and interestingness) is indicated by the means, standard deviations, and maximum and minimum values for these ratings across all stimuli; the values are $M_s = 5.68, 4.88,$ and $4.99$, $SDs = 1.78, 0.97, $ and $0.88$; $Max = 8.00, 6.62,$ and $7.19$; $Min = 1.88, 2.73,$ and $2.35$, respectively. These data suggest that the discriminability of the complexity judgment was high, whereas judgments of pleasantness and interestingness were, relative to complexity, lower.

Homogeneity

To confirm the homogeneity of the judgments of complexity, pleasantness, and interestingness, a principal components analysis was carried out on the data matrix of participants × stimuli for each set of ratings. The eigenvalues and the accumulation contribution ratios for the complexity, pleasantness, and interestingness judgments appear in Table 2. As for the complexity judgments, the eigenvalue for the first principal component was relatively larger than those for
the other two components. The eigenvalues for the first principle components for pleasantness and interestingness were lower than that for complexity, again confirming previous conclusions about the greater rating homogeneity of complexity ratings.

Physical or Perceptual Attributes of the Color Combinations and Judgments

In order to clarify the relation with the physical or perceptual attributes of color combinations to complexity, pleasantness, and interestingness, and to extract the structure of these judgments, a quantification theory type I analysis (Hayashi, 1952) was performed on each of the mean ratings using the number of colors factor (6 categories) and the color relations factor (10 categories) as predictor variables. This multivariate statistical technique realizes an analysis similar to that obtained using multiple regression analysis. A multiple regression analysis requires interval scaled (quantitative) measurement of both predictor and dependent variables, whereas this technique is applicable when predictor variables are qualitative (e.g., types of color relations) and dependent variables are quantitative (e.g., means of complexity ratings). This technique is similar to Guttman’s Scalogram.

Table 3 shows the coefficient of determination ($R^2$) and the partial correlation coefficients ($pr$, which correspond to partial regression coefficient in the case of regression analysis) for the number of colors factor and the color relations factor obtained by analysis in each of the three judgment scales.

Consider first, the complexity judgment. The coefficient of determination was quite high, confirming that the subjective complexity scale in color combinations is well explained by combinations of the two factors, number of colors, and color relations. The partial correlation coefficients were high and similar for

<table>
<thead>
<tr>
<th>Table 2. The Eigenvalues and the Accumulation Contribution Ratios of Complexity, Pleasantness, and Interestingness Obtained by the Principal Component Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating scale</td>
</tr>
<tr>
<td>-----------------------------------</td>
</tr>
<tr>
<td>Complexity</td>
</tr>
<tr>
<td>Pleasantness</td>
</tr>
<tr>
<td>Interestingness</td>
</tr>
</tbody>
</table>
both color factors, although the number of colors factor produced a slightly higher $pr$ value than did the color relations factor. In the number of colors factor, the level of two-color combinations received very low complexity ratings. Although complexity increased proportionately to the number of colors in a color combination, this tendency diminished logarithmically. When Tukey’s multiple comparison test was used to verify the significant differences between the levels of this factor, no significant differences emerged between the four highest levels which ranged from 6-color combinations to 12-color combinations. For the color relations factor, results indicated that the same tone combinations consisted of light tone and the similar hues combinations were judged less complex, whereas the contrast and the random combinations were judged more complex.

With regard to the pleasantness judgment, the coefficient of determination was high; however the partial correlation coefficient was extremely high only for the color relations factor, indicating that only the color relations defined the pleasantness judgment. The number of colors factor had a very weak effect. In the color relations factor, the same tone combinations of light and vivid tone and the similar hues combinations were judged pleasant, whereas the same tone combinations of dull and dark tone, the contrast and the random combinations were judged unpleasant.

Finally, for the interestingness judgment, the coefficient of determination and partial correlation coefficients for both color factors were high. These findings confirm that the interestingness judgment is explained by both factors of the number of colors and the color relations. In the number of colors factor, the two-color combinations were judged extremely uninteresting. On the other hand, there was no clear tendency between the combinations other than for two-colors and overall they were judged interesting. When differences between each of the number of colors were tested for significance using a Tukey’s multiple

<table>
<thead>
<tr>
<th>Rating scale</th>
<th>$R^2$</th>
<th>Color relations</th>
<th>Number of colors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complexity</td>
<td>.94</td>
<td>.83***</td>
<td>.96***</td>
</tr>
<tr>
<td>Pleasantness</td>
<td>.81</td>
<td>.90***</td>
<td>.21*</td>
</tr>
<tr>
<td>Interestingness</td>
<td>.82</td>
<td>.84***</td>
<td>.83***</td>
</tr>
</tbody>
</table>

* $p < .05$, *** $p < .001$. 

Table 3. Multiple Correlation Coefficients and Partial Correlation Coefficients Obtained by the Quantification Theory Type 1 Analysis
comparison test, the number of colors from 4 to 12 did not differ significantly. In addition, there was a significant correlation between the number of colors and mean interesting ratings \((r = .67, p < .001)\); however, when the two-color combinations were excluded, there was no correlation \((r = .19, n.s.)\). Concerning the color relations factor, warm and saturated color combinations containing the color attributes such as Red or vivid tone were judged interesting, whereas cool and dull color combinations containing the color attributes such as Blue or dull and dark tone were judged uninteresting.

**Complexity, Pleasantness, and Interestingness**

**Complexity and Pleasantness**

The scatter diagram showed a relationship suggestive of an inverted-U function between mean ratings of complexity and pleasantness. The inverted-U shaped scatter diagram was also found between the first principal component scores of complexity and pleasantness with the multiple correlation coefficient of .57 (see Figure 2). This inverted-U relationship suggests that the same tone combinations of light tone and the similar hues combinations were judged pleasant, and the same tone combination of vivid tone, the contrast and the random combinations as unpleasant. In addition, within each of the color relations, this kind of relationship appeared only for the same tone combinations and the similar hues combinations \((R = .77\) and \(.89, \text{respectively})\).

**Complexity and Interestingness**

There was a significant positive correlation between the mean ratings of complexity and interestingness \((r = .67, p < .001)\). And there was a highly significant positive correlation between the first principal component scores of complexity and interestingness \((r = .83, p < .001)\); in addition, the multiple correlation coefficient of this linear regression was very high, \(R = .84\) (see Figure 3).

**Pleasantness and Interestingness**

There was a significant but low positive correlation between the mean ratings of pleasantness and interestingness \((r = .30, p < .01)\). In addition, there was a significant, but very low negative correlation between the first principal component scores of both judgments \((r = -.26, p < .01)\). Thus, the relation appears extremely weak.
DISCUSSION

Complexity Scale

The complexity judgment of the color combinations showed high stability, discriminability, and homogeneity, suggesting that the subjective complexity rating scale employed, has a potential as an effective explaining variable of aesthetic judgments. Furthermore, this high level of precision may indicate that ratings of complexity involve different features than those for pleasantness and interestingness.

A tendency for perceived complexity to increase with the number of colors in combination was detected. This corresponds with past findings involving complexity scales based on the number of shape elements. This tendency, however, was diminished logarithmically, which suggests that there is an upper limit to the contribution of the number of colors to the complexity of color
Concerning the color relations factor, the similar hues and part of the same tone combinations tended to be judged as simple, and conversely the contrast and the random combinations increased the complexity ratings. These findings, too, resemble outcomes from traditional research on complexity scales; however, though traditional research using visual shapes usually relies upon structurally limited or abstract stimuli, such color combination stimuli as we used in the present research are closer to visual arts.

**Effectiveness and Limits of Complexity Scale in Color Combinations**

To estimate how well this scale explains pleasantness and interestingness, the nature of the relationships between complexity and aesthetic judgments were examined. In the first principal component score, the relationship between pleasantness and complexity follows an inverted-U function, though this relation was weak in the case of mean ratings. That is, a color combination that possesses

![Figure 3. The relationship between the first principal component scores for complexity and interestingness obtained by the Principal Component Analysis (N = 96).](image-url)
moderate complexity tended to be judged most pleasant. The inverted-U relation between complexity and hedonic tone has been reported in past research (e.g., Hekkert & van Wieringen, 1990; Russell & Milne, 1997), but it has been confined chiefly to studies involving abstract stimuli. The present results support these earlier findings because the stimuli employed in this research can be considered as abstract. With respect to the inverted-U pattern in the color combinations, however, some color relations were pleasant, and others were unpleasant. In addition, this relationship was only found in color combinations that contained color similarities, that is, those which involve the same tone or similar hues. This suggests that the inverted-U relationship between complexity and pleasantness in color combinations is a limited one which depends on specific stimulus attributes.

On the other hand, a strong linear relationship was found between the first principal component score of complexity and interestingness as well as mean ratings. That is, the more complex a color combination is, the more it is interesting. This result corresponds to those found in past research. Therefore, one can conclude that the complexity scale would be particularly effective as an explanatory variable for the judgment of interestingness.

Thus, the above-mentioned relations, whether linear or inverted-U, suggest the effectiveness of the complexity scale in explaining aesthetic judgments of color combinations. In addition, the complexity scale can explain the heterogeneity of pleasantness and interestingness.

**Pleasantness and Interestingness**

In the past, most researchers investigating pleasantness of color combinations focused on color harmony. Their goal was to identify universal principles of harmony judgments common to humankind (e.g., Eysenck, 1941). Unlike these approaches, the homogeneity of the pleasantness judgment was not especially high in the present research. However, it was found that the characteristics of color attributes and the color relations in color combinations were used to judge pleasantness and the number of colors was hardly involved. This confirms the relevance of color harmony theories based on color attributes. In addition, the present results that bright or orderly color combinations (e.g., the same tone combinations of light tone and the similar hues combinations) were judged pleasant, whereas dull or disorderly color combinations (e.g., the same tone combinations of dull and dark-tone, the contrast and the random combinations) where unpleasant corresponds nicely with findings reported by color harmony research (e.g., Nayatani et al, 1969; Ou & Luo, 2006).

As briefly touched on the introduction, if the judgment of pleasantness can be associated with the concept of unity, then the judgment of interestingness can be associated with variety. In this research, pleasantness was defined by parts of the color relations (corresponding to unity), and interestingness was chiefly defined by complexity (corresponding to variety). These results reflect
the main structural differences between pleasantness and interestingness. Aside from complexity, however, both the color relations and the number of colors also defined interestingness. In the former, the bright color combinations composed by warm or vivid colors were more interesting, suggesting a relation to the affective meanings of color. In the latter, judgments of two-color combinations showed a different tendency from those of other numbers of colors as described in result section and thus, it seems that the criteria for two-color and more than four-color combinations are different. The high partial correlation coefficient for the number of colors factor for the interestingness judgments may be due to this heterogeneity of two-color combinations. This suggests the limitations to the role of color harmony using two-color combinations as stimuli because participants may adopt different criteria between two-color and multiple-color combinations.

Finally, the present research confirms that the complexity scale has a potential as an effective explanatory variable for aesthetic judgments. As already indicated, however, homogeneity of pleasantness and interestingness judgments were not especially high and eigenvalues of both judgments suggested that two or more judgment criteria were inherent in the mean ratings. Though not mentioned in the results, linear relationships of both a positive and a negative type between complexity and pleasantness were found as the second and the third principal component scores, respectively. In addition, an inverted-U relationship with complexity was found as the second principal component score for interestingness. The former finding corresponds to those of Avital and Cupchik (1998) and Walker (1971); the latter to Walker’s results. Taken together, these findings suggest that the linear or the inverted-U criterion may be inherent in both aesthetic judgments. In the future, we hope to replicate these suggestive finding using other variations of stimulus attributes and different participants as well as different procedures and analysis method.

REFERENCES


Direct reprint requests to:

Ako Tsutsui
Graduate School of Humanities and Sociology
The University of Tokyo
Department of Psychology 7.3.1 Hongo Bunkyo-ku
Tokyo 113-0033, Japan
e-mail: ako_221@l.u-tokyo.ac.jp