EMPIRICAL STUDIES OF THE ARTS, Vol. 28(2) 173-191, 2010

VISUAL COMPLEXITY AND BEAUTY APPRECIATION: EXPLAINING THE DIVERGENCE OF RESULTS

MARCOS NADAL

ENRIC MUNAR

GISÈLE MARTY

CAMILO JOSÉ CELA-CONDE

University of the Balearic Islands

ABSTRACT

Although a number of studies have verified Daniel Berlyne's (1971) predicted maximum preference for intermediately complex stimuli, others have found that preference increased or decreased in relation to complexity. The objective of the present work was to assess whether differences in the kinds of stimuli used in prior studies or in the way complexity was defined could explain this divergence. In the first phase a set of 120 stimuli varying in complexity, abstraction, and artistry was assembled. In the second phase 94 participants were asked to rate the beauty of the stimuli. In the final phase the same participants rated 60 of the stimuli on seven complexity dimensions. We failed to detect any meaningful influence of complexity on beauty ratings for any of the kinds of stimuli. However, our results suggest that there are three different forms of complexity that contribute to people's perception of visual complexity: one related with the amount and variety of elements, another related with the way those elements are organized, and asymmetry. We suggest that each of these types of complexity influences beauty ratings in different ways, and that the unresolved relation between complexity and beauty appreciation is mainly due to differences in the conception, manipulation, and measurement of visual complexity.

doi: 10.2190/EM.28.2.d
http://baywood.com

The knowledge that order and complexity influence people's appreciation of beauty goes back at least to the ancient Greeks. However, it was not until Fechner's (1876) work that this issue was systematically studied, paving the way for Birkhoff's (1932) mathematical approach, which predicted that appreciation would increase with order and decrease with complexity. Although Birkhoff accompanied his formulation of the aesthetic measure with detailed definitions of order and complexity, together with examples for a large number of polygons, Eysenck (1941) reported considerably low correlations between predicted aesthetic measure and the beauty ratings actually awarded by people. In order to find a more satisfactory alternative, Eysenck (1941) studied people's responses to a broad number of geometrical figures. His results suggested that different features of those objects, related with complexity and order, could be used to predict the beauty ratings awarded by human participants. However, the relation between both factors was not the one predicted by Birkhoff. In a simplification of his original formula, Eysenck (1942) showed both order and complexity contribute positively to the appreciation of beauty.

It was not until Berlyne (1971) presented his influential framework that the study of complexity's influence on the appreciation of beauty was based on firm psychological and neurobiological grounds. In brief, Berlyne (1971) posited that the interaction of reward and aversion brain systems would lead people to prefer intermediate levels of complexity, which was defined according to such aspects as pattern regularity, amount of elements, their heterogeneity, or the irregularity of the forms (Berlyne, 1963, 1970, 1971; Berlyne, Ogilvie, & Parham, 1968).

A considerable amount of research has since been conducted to test this hypothesis, employing diverse visual stimuli. Some studies have been carried out with simple materials, such as geometric shapes (Aitken, 1974; Katz, 2002; Vitz, 1966), or artificially generated images (Heath, Smith, & Lim, 2000; Ichikawa, 1985; Markovic & Gvozdenovic, 2001; Stamps, 2002), varying along a specific, clearly defined, and objectively measurable complexity dimension. Other studies have used artworks, though often including only a single class, such as abstract paintings (Krupinski & Locher, 1988; Nicki & Moss, 1975; Osborne & Farley, 1970), cubist artworks (Nicki, Lee, & Moss, 1981), figurative images (Messinger, 1998), or portraits (Saklofske, 1975). These studies were not able to carry out such a straightforward measurement of complexity. A serious limitation of both sets of studies is that in most instances the number of participants was quite low (in some experiments as few as eight per condition) or the number of stimuli was reduced (only five in some instances). Additionally, the large majority of them used correlational methodology, which precluded the proposal of causal explanations of complexity's influence on aesthetic appreciation.

The review of these prior studies reveals a considerable divergence in results. Some of them did actually find the expected inverted U-shaped distribution of preference and beauty scores as a function of complexity. Conversely, others found that preference and appreciation increased with complexity and, yet others, found that ratings decreased as the stimuli's complexity grew. We believe that two reasons could explain this disparity: differences in the kinds of materials used in prior studies, and the disparate ways in which they conceived, measured, and manipulated visual complexity.

The objective of the present study was, therefore, to clarify the influence of visual complexity on beauty appreciation, and examine the role of stimuli features such as degree of abstraction and artistry. Additionally, we wished to explore the notion of visual complexity. Specifically, we addressed the following questions: (i) Do people consistently rely on the same features to perform complexity judgments of diverse visual stimuli? (ii) Is visual complexity reducible to a single measure or is it multidimensional in nature?, and (iii) Do all forms of complexity influence beauty ratings in the same way?

METHOD

Participants

All participants were students enrolled in their fourth or fifth year of psychology, philosophy, or history studies, who voluntarily took part in the experiments. We included no one who had received formal art training or had studied art history. Participants were divided into two groups. The first one took part in the first phase of the experiment, aimed at reaching an adequate set of stimuli to be used in subsequent phases. The second group participated in phases two and three, which were designed to explore the relation among different kinds of visual stimuli, different levels and forms of complexity and aesthetic appreciation. Group 1 consisted of 240 participants: 112 men (46.7%) and 128 women (53.3%). Their ages ranged from 18 to 44 years, with a mean of 22.03 and a standard deviation of 3.75. Group 2 included 94 participants: 38 men (40.4%) and 56 women (59.9%). Ages ranged from 18 to 46 years, with a mean of 22.41 and a standard deviation of 4.1.

Materials

Visual Stimuli

We collected over 1,500 digitalized images, which were either abstract or representational, and either artistic or not artistic. The distinction between abstract and representational stimuli referred to the absence and presence of explicit content, respectively. Artistic stimuli were reproductions of catalogued pieces created by renowned artists and exhibited in museums. Following Heinrichs and Cupchik's (1985) recommendation, we included images belonging to diverse styles and schools, such as realism, cubism, impressionism, and so on. As a guide for our initial selection we used the compendium *Movements in Modern Art* from the *Tate Gallery, London* (Cottington, 1998; Gooding, 2001; Malpas, 1997;

Thomson, 1998) and added reproductions of XVIII and XIX century American and European artworks. Non-artistic stimuli included postcards, photographs of landscapes, artifacts, urban scenes, and so on, taken from the series of books *Boring Postcards* (Parr, 1999, 2000), photographs taken by ourselves, as well as digital images from the series of CDs *Master Clips Premium Image Collection* (ISMI, San Rafael, CA), used in industrial design, book illustrating, etc.

Stimuli Selection and Modification

The initial set of images was subjected to a series of modifications in order to eliminate the influence of potentially confounding variables. Only relatively unknown pieces were selected to avoid the impact of familiarity, as recommended by Eysenck (1940). In order to avoid the influence of ecological variables we eliminated those stimuli that contained clear views of human figures and human faces, as well as those stimuli portraying scenes that could elicit strong emotional responses. The undesired influence of psychophysical variables was controlled by adjusting all stimuli to the same resolution of 150 ppi, setting a common size of 9 by 12 cm, standardizing the color spectrum, and adjusting luminance to between 370 and 390 lx. Stimuli that could not be reasonably modified to comply with this homogenization were discarded. Finally, the signature was removed from all signed pictures. This process of stimuli selection and modification was carried out to leave us with 800 images, 200 of which were abstract artistic, 200 were abstract non-artistic, 200 were representational artistic, and 200 were representational non-artistic.

Hardware and Software

All stimuli were presented to participants and their responses registered by means of a specifically designed software running on Compaq EVO300 Pentium IV / 1,7 GHz computers with Windows 2000 SP4. After participants had introduced demographic information and understood the instructions, they went on to the actual experimental protocol. All stimuli were presented within a grey frame. In the upper segment of the frame there was a brief reminder of the task they were asked to perform. In the lower segment of the frame there was a reminder of the response scale they were required to use. Based on previous studies that showed no effect of viewing time on aesthetic appreciation (McWhinnie, 1993; Smith, Bousquet, Chang, & Smith, 2006), we decided not to impose a time limit on participants' responses. Hence, each stimulus was presented until participants responded. They did so by pressing a key between 1 and 5 in the phases 1 and 2, and a key between 1 and 9 in phase 3. If participants pressed any other key there was no response from the program. If the response was within the appropriate values it was fed-back on to the screen for 1.5 s, after which the 2-s masking screen appeared again. This same pattern was implemented for the 100, 120, and 60 stimuli in the first, second, and third phases, respectively (see in the Procedure section). The computer program registered all the demographic information given by the participants and each of their responses to the stimuli in each phase.

Procedure

In order to achieve the objectives outlined above, the present work was structured into three phases. The first phase was designed to create a set of stimuli suitable to carry out both subsequent phases. The second phase consisted in the exploration of the relation between complexity and beauty appreciation by means of a beauty rating task using the set of stimuli created in the first phase. Finally, the aim of the third phase was to address the three issues we specified with regards to the nature of visual complexity.

Phase 1: Creating an Adequate Stimulus Set

Our aim was to obtain a set of 120 stimuli equally divided into three complexity levels: low, intermediate, and high. Each of these complexity levels would include 10 abstract artistic (AA), 10 abstract non-artistic (AN), 10 representational artistic (RA), and 10 representational non-artistic (RN) stimuli.

The 240 participants recruited for phase 1 were divided into eight different groups of 30 individuals, attempting to balance each of them in relation to sex. The set of 800 stimuli was divided into eight groups of 100, using a stratified randomized method, such that there were 25 stimuli of each kind (AA, AN, RA, RN) in each group. They were presented to all participants in the same random sequence. Participants were asked to rate the complexity of each of the 100 images on a 1 to 5 Likert scale (very simple–very complex). At this stage complexity was not defined to participants. Participants were only instructed to focus on their general impression of the visual complexity of each stimulus, not on the complexity involved in producing it.

Two statistics were calculated for each stimulus: the average rating awarded by the 30 participants, considered as the complexity score, and the standard deviation, considered as the measure of participants' agreement on that score. The selection of stimuli for each complexity level was based on both statistics, and was carried out separately for each of the stimuli types (AA, AN, RA, RN) according to the following procedure. The 200 stimuli in each of the four kinds of images were ordered according to their complexity score. To select stimuli for the low complexity level, the experimenter began at the bottom of the list of stimuli (those with the lowest complexity score). If the standard deviation for the first stimulus was below .80, it was selected. If it was .81 or above, the stimulus was discarded and the operation was repeated with the stimulus immediately above in complexity score. This process was carried out for each of the four stimuli types until 10 stimuli of each particular type had been selected. In order to select stimuli for the high complexity level the same procedure was followed, except that it began at the top of the list and moved down the complexity scores. Again, this finished when 10 stimuli of each kind had been selected. In order to select stimuli for the intermediate level of complexity, the median of the complexity scores was calculated for each stimulus kind. The experimenter started the selection at that point, using the same agreement criterion as mentioned above, only that he alternatively moved up and down the list to select or discard the stimuli. When the 10 images of each kind had been selected, the process was ended.

This procedure was followed with the objective of maximizing the difference between the three complexity levels and to minimize the difference in complexity within levels. Choosing images whose complexity score had a small standard deviation was aimed at including stimuli for which people tended to agree on their degree of complexity.

Phase 2: Beauty Appreciation Test

The objective of this phase was to explore the influence of complexity, degree of abstraction, and artistry on beauty appreciation. The 94 participants described above were asked to rate the beauty of the 120 stimuli selected in the previous phase on a 1 to 5 (very ugly–very beautiful) Likert scale. Stimuli were randomized and presented in the same order to all participants, who were sat at different computers in the same room.

The average rating awarded by participants was calculated for each stimulus. This was the dependent variable used in our analysis of the influence of the three independent variables: Complexity (low, intermediate, high), Abstraction (abstract and representational), and Artistry (artistic and non-artistic).

Phase 3: The Nature of Visual Complexity

This phase was conceived to explore the influence of different features of visual stimuli on judgments of complexity, the relations between these features, and their relation with the appreciation of beauty. In order to do so we selected 60 of the 120 stimuli used in Phase 2. This subset was constituted by including five stimuli from each of the four kinds in each level of complexity from the 120 stimuli set. The median value of complexity was calculated for each of the 12 subgroups. We included the stimuli corresponding to the median value and the two adjacent stimuli on both sides.

Based on our review of the literature on visual complexity and its influence on the appreciation of beauty, we selected seven features we believed could relate to different aspects of visual complexity. These dimensions were: Unintelligibility of the elements (difficulty to identify the elements in the image), Disorganization (difficulty to organize the elements into a coherent scene, Amount of elements, Variety of elements, Asymmetry, Variety of colors, and Three-dimensional appearance.

The same group of 94 participants who served as subjects in phase 2 took part in this third and last phase. In this case they were asked to rate each stimulus on a 1 to 9 Likert scale for each of the seven aforementioned scales. All the stimuli in the subset were rated on each dimension separately. Stimuli were presented in a different random order for each dimension. Before rating the stimuli on each particular dimension, participants received written and verbal instructions and a brief definition. After all participants had finished the task, their ratings were collected. The average rating awarded on each dimension were calculated for each stimulus.

RESULTS

Phase 1: Creating an Adequate Stimulus Set

The descriptive statistics of the set of 120 images, selected from the initial 800 according to the criteria mentioned in the procedure section, are shown in Table 1 for each complexity level by abstraction and artistry.

A series of Kruskal-Wallis tests were carried out to make sure that the set was well suited to use in subsequent phases. Results showed that there were differences between complexity scores of stimuli included in the high, intermediate, and low complexity levels for each of the four stimuli kinds, for all abstract stimuli, all representational stimuli, all artistic stimuli, and all non-artistic stimuli, as well as for whole set taken together. All differences were highly significant, with the lowest statistic corresponding to the representational non-artistic subset ($\chi^2 = 25.88$, $p < .001$) and the highest to the whole set ($\chi^2 = 71.02$, $p < .001$). Additional Mann-Whitney pairwise comparisons revealed that scores of stimuli

Table 1. Descriptive Statistics for the Set of 120 Visual Stimuli to be Used in Phases 2 and 3

	Artistry							
	Artistic				Non-artistic			
Complexity level	Abstraction	*n*	*M*	*SD*	Abstraction	*n*	*M*	*SD*
High	Abs	10	4.66	.15	Abs	10	2.98	.44
	Rep	10	4.37	.16	Rep	10	3.75	.19
Intermediate	Abs	10	3.47	.06	Abs	10	2.09	.04
	Rep	10	3.49	.03	Rep	10	2.52	.15
Low	Abs	10	2.48	.20	Abs	10	1.19	.03
	Rep	10	2.68	.12	Rep	10	1.40	.08

Note: Abstraction refers to abstract (Abs) and representational (Rep) images.

included in the high complexity level were significantly greater than those in the other two levels, and that stimuli included in the intermediate level had been rated as more complex than those included in the low level. This is true for each of the stimuli categories and the whole set of stimuli taken together. In this case, the lowest statistic corresponded to the comparison between low and intermediate complexity levels of abstract images ($Z = 2.71$, $p < .006$) and the highest to the comparison between low and high complexity scores for the whole set of stimuli ($Z = 7.34$, $p < .001$). Hence, the objective of the first phase of the investigation, the creation of a set of diverse visual stimuli grouped in three distinct levels of complexity, had been accomplished.

Phase 2: Beauty Appreciation Test

Table 2 shows the descriptive statistics for the beauty ratings to stimuli varying in complexity, abstraction, and artistry.

Given that homogeneity of variances could not be assumed, and that some of the distributions of beauty ratings could not be considered to approach normality, we used non-parametric techniques to test our hypotheses regarding the influence of complexity on beauty ratings awarded to visual stimuli varying in abstraction and artistry.

Our results show that our three independent variables had significant main effects on participants' beauty ratings, revealing that they preferred high complexity stimuli over low complexity stimuli ($Z = 2.95$, $p < .0083$), representational over abstract stimuli ($Z = 5.12$, $p < .001$), and artistic over non-artistic stimuli ($Z = 5.40$, $p < .001$). However, these results must be viewed in light of the triple

Table 2. Descriptive Statistics for Participants' Beauty Scores Awarded to Each Kind of Stimulus

	Artistry							
	Artistic				Non-artistic			
Complexity level	Abstraction	*n*	*M*	*SD*	Abstraction	*n*	*M*	*SD*
High	Abs	10	2.94	.44	Abs	10	2.15	.40
	Rep	10	3.53	.56	Rep	10	3.55	.58
Intermediate	Abs	10	2.77	.45	Abs	10	2.10	.47
	Rep	10	3.48	.54	Rep	10	2.74	.74
Low	Abs	10	2.67	.48	Abs	10	1.84	.23
	Rep	10	3.40	.42	Rep	10	2.11	.53

Note: Abstraction refers to abstract (Abs) and representational (Rep) images. Artistry refers to artistic and non-artistic images.

interactions we identified, specifically the effects of complexity within each abstraction by artistry level. Interaction analysis revealed that complexity had significant effects on participants' beauty ratings only of representational non-artistic images ($\chi^2 = 13.88$, $p < .001$). Scores were higher for high complexity representational non-artistic stimuli than for low complexity images of the same kind ($Z = 3.40$, $p < .001$).

Phase 3: The Nature of Visual Complexity

The first issue we explored in relation to complexity was the possibility that participants relied on different features when judging the complexity of different kinds of visual stimuli. In order to do so we performed a series of discriminant analyses, which allow determining the best combination of independent variables (complexity dimensions) to predict the level of complexity (low, intermediate, high) of diverse kinds of stimuli (AA, AN, RA, RN). Table 3 shows the results of these analyses.

It is interesting to note that for most stimuli kinds only one or two complexity dimensions were required to accurately predict the complexity level of the stimuli. Furthermore, remarkably high levels of explained variance, correct classification, and agreement between predicted and actual complexity level were achieved. From the results presented in Table 3, it seems that dimension 3 (Amount of elements) is the best overall predictor for general complexity ratings. Conversely, dimensions 2, 4, and 6 (Disorganization, Variety of elements, and Variety of colors), were of little relevance to predict complexity level.

Table 3. Results for the Discriminant Analyses Carried Out on Participants' Ratings of Each Kind of Stimulus

Stimuli	PD	EV (%)	CC (%)	Kappa
Abstract artistic	1 & 3	98.9	100	1
Abstract non-artistic	3	100	80	.70
Representational artistic	3, 5, & 7	98.3	100	1
Representational non-artistic	3	100	93.3	.90

Note: PD: predictive dimensions; EV: explained variance; CC: correct classification. Kappa: agreement between actual and predicted complexity level based on predictive dimensions. All kappa values are significant at $p < .001$.

With regards to kind of stimuli, results showed that dimension 3 (Amount of elements) was a sufficiently adequate predictor of complexity ratings of both abstract and representational non-artistic stimuli. The prediction of complexity level of artistic images required including other aspects. In addition to dimension 3, discriminant analysis revealed the importance of dimension 1 (Unintelligibility of the elements) when rating the complexity of abstract artistic stimuli, and of dimensions 5 (Asymmetry) and 7 (Three-dimensional appearance) when rating the complexity of representational artistic stimuli.

The second issue we explored in relation to the concept of complexity was whether there was any relation among the seven dimensions of complexity we had considered. This was done by means of factor analysis. Sedimentation tests recommended the extraction of three factors for participants' ratings. Although the third eigenvalue was below 1 (.970), we decided to maintain the recommendation of the sedimentation test, given that dimension 5 (Asymmetry) only loaded satisfactorily on this third factor. We believe this decision is justified by the historical relevance of symmetry in the study of the relation between complexity and the appreciation of beauty, and the fact that eigenvalues of the additional four factors were well below 1. Table 4 shows the initial eigenvalues for the seven components, as well as the percentage of variance explained by each of the three extracted factors, together with rotated extracted values.

The first extracted factor explained close to 48% of the variance, the second factor explained about 31%, and the third factor explained approximately 14.5%. Overall, the three factors accounted for over 93% of the variance in participants' ratings. Table 5 shows the rotated component matrix.

Table 4. Results for the Factor Analysis Carried Out on Participants' Scores on the Seven Complexity Dimensions, Including Initial Eigenvalues and Percentage of Explained Variance before Rotation, and Rotated Extracted Factors

	Initial eigenvalues			Rotation sums of squared loadings		
Factor	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	3.360	47.995	47.995	3.347	47.809	47.809
2	2.230	31.855	79.850	2.202	31.464	79.273
3	.973	13.898	93.748	1.013	14.475	93.748
4	.265	3.789	97.537			
5	.106	1.512	99.050			
6	.044	.631	99.680			
7	.022	.320	100			

Table 5. Rotated Component Matrix

	Factor		
Dimension	1	2	3
1	.089	.981	–.075
2	–.036	.987	–.048
3	.947	.230	–.093
4	.965	.115	.027
5	–.044	–.082	.994
6	.942	–.055	.003
7	.788	–.437	–.085

After the rotation it became very clear that the first factor received loadings mainly from dimensions 3 (Amount of elements), 4 (Variety of elements), 6 (Variety of colors), and to a lesser degree, 7 (Three-dimensional appearance), while saturations from other dimensions were negligible. Dimensions 1 (Unintelligibility of the elements) and 2 (Disorganization) showed high loadings on the second factor, whereas dimension 3 had a low positive loading and dimension 7 a moderate negative loading on this factor. Finally, only dimension 5 (Asymmetry) loaded heavily on the third factor.

The final issue we wished to clarify in relation to complexity was the possibility that different forms of complexity influenced beauty appreciation in diverse ways. Given that the present study was not designed expressly to test this possibility experimentally, our approach is merely exploratory in nature. We performed a curve fit test for each of the three complexity factors identified above and the beauty scores awarded by men and women. The factor scores corresponding to each stimulus were entered as independent variables and the beauty scores as dependent variables and calculated the fit to linear, quadratic, and cubic functions. If more than one function, or none, produced a significant fit, we chose the one with a lower significance level. If significance was equal, we chose the one with the greater R^2 value. The relation between factor 1 and beauty scores fit a cubic function ($F = 61.29$, $p < .001$). The relation between factor 2 and beauty scores fit a quadratic function ($F = 6.86$, $p < .002$). Finally, the best fit of the relation between factor 3 and scores was to a quadratic function, though non-significantly ($F = 1.79$, $p = .180$). These results are shown graphically in Figure 1. There are clear differences among the relations between the three complexity factors and beauty ratings. Ratings of beauty increased linearly with complexity factor 1, it had a descending U-shaped relation with complexity

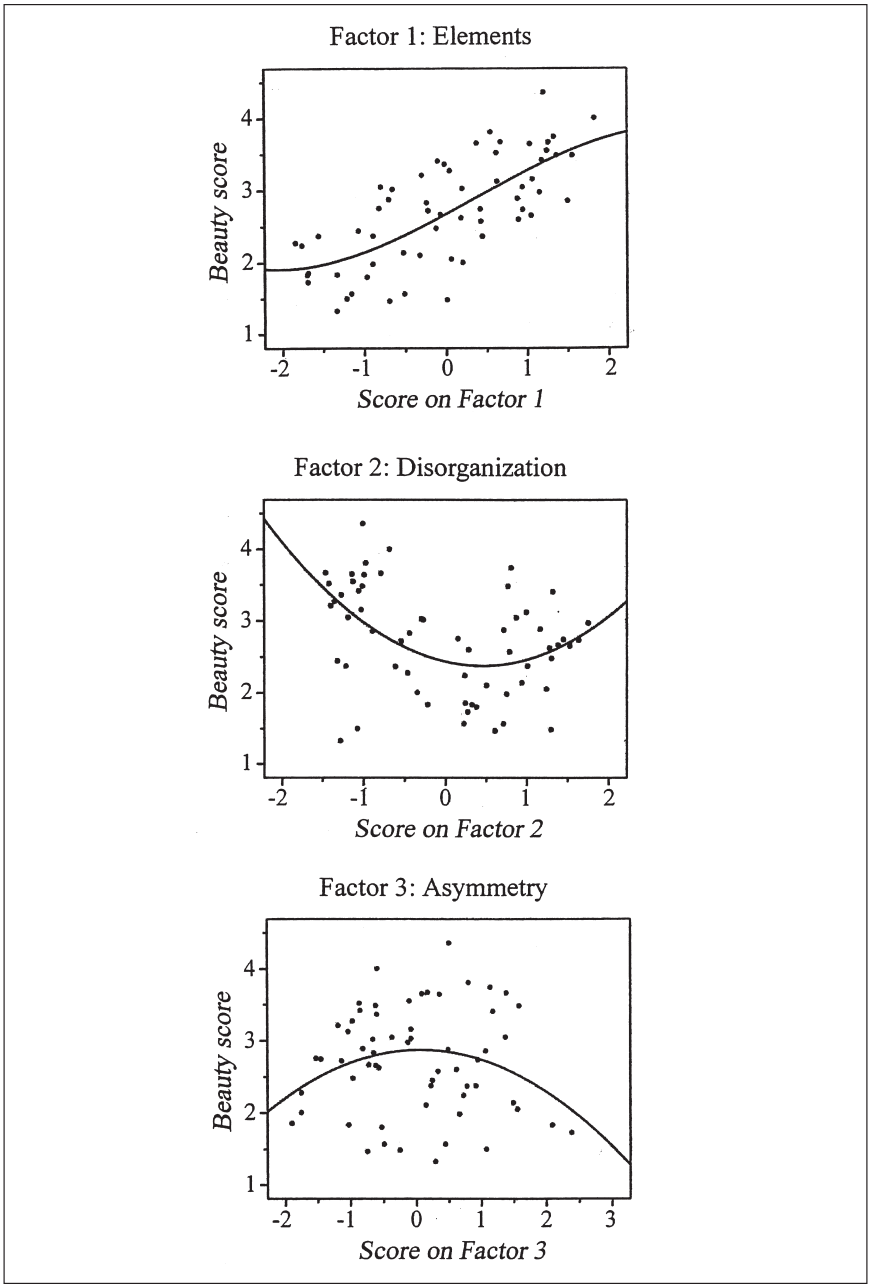

Figure 1. Results for the curve fit between the three complexity factors and beauty scores.

factor 2, and, finally, both variables showed an inverted U-shaped relation, whereby maximum beauty scores corresponded to intermediate levels of complexity factor 3.

DISCUSSION

The Influence of Complexity on Beauty Appreciation

Our results showed that complexity had a negligible effect on beauty ratings of most of our stimuli kinds. In fact, it turned out to influence only the beauty ratings of representational non-artistic images. Specifically, participants awarded higher beauty ratings to high complexity images of this kind than to the low complexity ones. This might seem to suggest that the effects of complexity on beauty ratings are modulated by the kind of stimuli, which would support one of our initial explanations for the divergence of results among earlier studies. However, an alternative explanation for the effect of complexity observed for this kind of stimuli arises from the careful examination of the data and the stimuli themselves. Participants rated highly complex representational non-artistic stimuli as beautiful as artistic representational stimuli belonging to any complexity level. Conversely, their beauty ratings of low complexity representational non-artistic stimuli were similar to those awarded to most abstract non-artistic stimuli, independently of their complexity. The comparison of low complexity and high complexity representational non-artistic stimuli reveals an unforeseen yet clear difference between both groups of stimuli. Representational non-artistic stimuli that were included in the level of low complexity are simple or schematic drawings or photographs of individual objects, such as a car, a biker, bananas, a pencil, and so on. Conversely, representational non-artistic stimuli assigned to the high complexity level are, for the most part, paintings or photographs of natural sceneries, such as landscapes or seascapes. Hence, it seems that the beauty ratings awarded by our participants without artistic training reflect a tendency to consider the art-looking postcards or illustrations as artistic and to reject simple depictions of individual objects. This suggests that our category of non-artistic stimuli could probably be subdivided into a category of what Lindauer (1990) and Winston and Cupchik (1992) might consider cheap or popular art and a category of what we could call icons or objects.

The Concept of Visual Complexity

Since Berlyne and colleagues' (1968) initial distinction among complexity forms, such as the amount of elements, their heterogeneity, the irregularity of their shapes, the irregularity of their disposition, the degree with which the different elements are perceived as a unit, asymmetry, or incongruence, there has not been much work aimed at determining whether these features impact perceived complexity in the same way and to the same extent. In addition, there has been little research on the relation among the complexity features themselves. Finally, there has been no attempt at determining whether different complexity features affect beauty appreciation in the same way or to the same extent. Most research in empirical aesthetics and visual perception has regarded complexity as a one-dimensional concept. Although many studies have dealt with the relation between complexity and the appreciation of beauty, their conception of complexity has not always emphasized the same aspect. Whereas some studies have conceived complexity as the amount of elements in a stimulus—lines, angles, turns, and so on—others have regarded it as the degree of asymmetry, or the degree of incongruity. This obviously creates problems when comparing their results, which may differ precisely because complexity refers to different aspects in different cases.

The results of our analyses revealed that the complexity level of each kind of visual stimuli could be predicted very reliably from solely one or two dimensions. Importantly, not all aspects of an image are taken into account by participants to the same extent when asked to rate its visual complexity. Overall, and in agreement with Berlyne and colleagues' (1968) results, participants' ratings of visual complexity seem to be driven mostly by the amount of elements. Conversely, disorganization, variety of elements, and the variety of their colors seem to have little influence on people's impression of visual complexity. This is in agreement with Hall's (1969) results, which suggested that the variety of colors did not represent an important factor when rating the complexity of linear stimuli. An interesting finding is that the artistic status of the stimuli influenced visual complexity ratings. Whereas the amount of elements seems to be the most relevant feature when people express their impression of the complexity of non-artistic stimuli, additional aspects appear to be taken into account when rating the complexity of artistic stimuli. These include unintelligibility of the elements in the case of abstract artistic images and asymmetry and three-dimensional appearance in the case of representational artistic images. These exploratory results suggest hypotheses that require future experimental testing under rigorously controlled conditions and with stimuli specifically manipulated to this end.

Regarding the relation among the seven complexity dimensions, our results indicated the existence of two factors that explained most of the variance, and a third one which we included to account for asymmetry. The first factor received high loadings from the following dimensions: Amount of elements, Element heterogeneity, Variety of colours, and Three-dimensional appearance. The second factor received high loadings from Unintelligibility of the elements and Disorganization. And the third factor, as we just mentioned, received high loadings only from Asymmetry. These three factors could be referred to as *elements*—which has to do with the amount and variety of the elements, *organization*—related with how the elements are grouped to form identifiable objects and how these are organized into a coherent scene, and *asymmetry*.

Generally speaking, these results are in line with prior studies. Berlyne and colleagues' (1968) factor analysis indicated the existence of two main factors, one related with the amount of elements and another which was a composite of several dimensions, and named it unity versus articulation into easily recognizable parts. It stands out that these two factors are very similar, or even equivalent, to our *elements* and *organization* factors. However, whereas our *elements* factor accounted for just under half of the variance in complexity ratings, Berlyne and colleagues (1968) found that it accounted for between 70% and 90% of the variance. This difference in the relevance of the amount of elements on complexity ratings might be due to the fact that the stimuli used by Berlyne and colleagues were simple line drawings in which the constituting elements were much more salient that in most of the stimuli used in the present study.

Other studies have also found participants' impression of complexity to depend on two kinds of features. Nicki and Moss (1975) suggested that there might be two kinds of complexity factors, a "perceptual" one related with the number and variety of elements, and a "cognitive" one related with the amount of associations or cognitive tags elicited by stimuli. Chipman (1977) distinguished between a qualitative component of complexity judgments, determined largely by the amount of elements, and a structural component, related with symmetry, the repetition of motifs and other organizational processes. Chipman noted that the first factor, related to the amount of elements, seems to set an upper threshold of perceived complexity and the second one can act to reduce this impression, a suggestion that was later experimentally corroborated by Ichikawa (1985).

Thus, our results add further support to the idea that two or three processes contribute to the formation of subjective visual complexity. Probably the most important one is the determination of the number and variety of elements. The second one refers to how well the elements organize into a coherent scene. Although previous studies have subsumed asymmetry within organizational processes, the present results showed this was not an adequate solution for our data, and hence, we chose to include it as a separate factor. The temporal sequence of cognitive processes related with these factors remains to be elucidated, though based on Ichikawa's (1985) results, a plausible hypotheses is that the different features are processed in parallel, but that operations related with *elements* are faster than those related with *organization*, which are completed later.

The final part of this study was a tentative exploration of the possibility that the different factors of complexity are related in different ways to beauty ratings. This can only be regarded as a very tentative exploration because stimuli were not designed to vary independently on each of the complexity factors. Despite this limitation, our results suggest that complexity factors might influence beauty appreciation in very different ways. The *organization* factor seemed to have a U-shaped or descending relation to beauty. Specifically, stimuli receiving extremely low values on this factor were rated as more beautiful than those receiving intermediate and high scores. In contrast, the second factor, *elements*, had a positive relation with beauty: high-scoring images, those with a large amount and variety of elements, were those that had obtained the best beauty ratings. Finally, our results suggest that the last factor, *symmetry*, seems to have an inverted U-shaped relation to beauty: images rated as intermediately asymmetric were considered to be more beautiful than those rated as extremely asymmetric or extremely symmetric.

Our review of the literature supports the possibility that the diversity of relations between complexity and beauty that have been found in previous studies were due to their emphasis on different complexity factors. Based on our results, we would expect those studies that have varied complexity by manipulating the amount or diversity of elements to have found an increasing relation between complexity and beauty ratings. We would expect to find that studies manipulating complexity by means of organizational features obtained decreasing or U-like distributions between complexity and beauty ratings. Finally, we would expect prior studies that specified complexity along a symmetry-asymmetry dimension to have produced the expected inverted-U distribution of beauty ratings over complexity. We used 15 prior studies to test this retrospective prediction. We only selected those that utilized some sort of specific complexity measure, leaving out those that assessed complexity by means of a general complexity rating scale. Six studies had designed or employed stimuli which varied along the *elements* factor (Aitken, 1974; Day, 1967; Heath et al., 2000; Nicki, 1972; Nicki & Moss, 1975; Stamps, 2002), five had designed or used stimuli which varied along the organization factor (Krupinski & Locher, 1988; Neperud & Marschalek, 1988; Nicki et al., 1981; Nicki & Moss, 1975; Osborne & Farley, 1970), one had used stimuli varying only in asymmetry (Krupinski & Locher, 1988), three had used stimuli which varied in asymmetry as well as elements (Eisenman, 1967; Imamoglu, 2000; Munsinger & Kessen, 1964), and one had used stimuli varying in all three factors (Francès, 1976), which was discarded due to its combined use of measures related with the three factors. We also pooled the studies that had conceived complexity as asymmetry or the combination of asymmetry and number of elements into a single category. For each of the 15 studies, we summarized its main conclusion as supporting an increasing, inverted U-shaped, or decreasing/U-shaped, relation of awarded beauty scores and complexity. Table 6 shows the cross-tabulation of the main factors manipulated by these studies and the shape of the relation between complexity and beauty appreciation.

The results of the Chi-square test are highly significant ($\chi^2 = 15.33$, $p < .004$). Directional measures were calculated in order to assess the strength of the association. They revealed that there is a strong relation between the way in which previous studies have specified complexity and their resulting distribution of beauty scores as a function of complexity. Moreover, as the measures of association show, the kind of distribution can be predicted from the complexity factor manipulated by the experimenter with a considerable degree of accuracy

Table 6. Cross-tabulation of Prior Studies According to the Factor Representing the Complexity Feature They Manipulated and the Shape of the Relation between Complexity and Beauty Appreciation They Found

	Main complexity factor			
Main result	Elements	Organization	Symmetry	Total
Increasing	5	1	0	6
U-shaped	0	4	1	5
Inverted U	1	0	3	4
Total	6	5	4	15

($\lambda = .67$, $p < .011$; $\tau = .52$, $p > .006$; U = .54, $p < .001$). This shows that the choice of complexity factor among *elements*, *organization*, and *symmetry*, has had an impact on the shape of the resulting distribution of beauty over complexity. Just as our results had suggested, most studies manipulating the number or variety of elements found an increasing relation between complexity and beauty appreciation, most of those manipulating organizational features had found a U-like or descending relation, and most of those that had manipulated symmetry found an inverted-U distribution.

REFERENCES

Aitken, P. P. (1974). Judgments of pleasingness and interestingness as functions of visual complexity. *Journal of Experimental Psychology, 103*, 240-244.

Berlyne, D. E. (1963). Complexity and incongruity variables as determinants of exploratory choice and evaluative ratings. *Canadian Journal of Psychology, 17*, 274-290.

Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Perception & Psychophysics, 8*, 279-286.

Berlyne, D. E. (1971). *Aesthetics and psychobiology*. New York: Appleton-Century-Crofts.

Berlyne, D. E., Ogilvie, J. C., & Parham, L. C. C. (1968). The dimensionality of visual complexity, interestingness, and pleasingness. *Canadian Journal of Psychology, 22*, 376-387.

Birkhoff, G. D. (1932). *Aesthetic measure*. Cambridge, MA.: Harvard University Press.

Chipman, S. F. (1977). Complexity and structure in visual patterns. *Journal of Experimental Psychology: General, 106*, 269-301.

Cottington, D. (1998). *Cubism*. London: Tate Gallery Publishing.

Day, H. (1967). Evaluations of subjective complexity, pleasingness and interestingness for a series of random polygons varying in complexity. *Perception & Psychophysics, 2*, 281-286.

Eisenman, R. (1967). Complexity-simplicity: I. Preference for symmetry and rejection of complexity. *Psychonomic Science, 8*, 169-170.
Eysenck, H. J. (1940). The 'general factor' in aesthetic judgements. *British Journal of Psychology, 31*, 94-102.
Eysenck, H. J. (1941). The empirical determination of an aesthetic formula. *Psychological Review, 48*, 83-92.
Eysenck, H. J. (1942). The experimental study of the 'Good Gestalt'—A new approach. *Psychological Review, 49*, 344-363.
Fechner, G. T. (1876). *Vorschule der Ästhetik*. Leipzig: Breitkopf und Härtel.
Francès, R. (1976). Comparative effects of six collative variables on interest and preference in adults of different educational levels. *Journal of Personality and Social Psychology, 33*, 62-79.
Gooding, M. (2001). *Abstract art*. London: Tate Gallery Publishing.
Hall, A. C. (1969). Measures of the complexity of random black and white and coloured stimuli. *Perceptual and Motor Skills, 29*, 773-774.
Heath, T., Smith, S. G., & Lim, B. (2000). Tall buildings and the urban skyline. The effect of visual complexity on preferences. *Environment and Behavior, 32,* 541-556.
Heinrichs, R. W., & Cupchik, G. C. (1985). Individual differences as predictors of preference in visual art. *Journal of Personality, 53*, 502-515.
Ichikawa, S. (1985). Quantitative and structural factors in the judgment of pattern complexity. *Perception & Psychophysics, 38*, 101-109.
Imamoglu, C. (2000). Complexity, liking and familiarity: Architecture and non-architecture Turkish students' assessments of traditional and modern house facades. *Journal of Environmental Psychology, 20*, 5-16.
Katz, B. F. (2002). What makes a polygon pleasing? *Empirical Studies of the Arts, 20*, 1-19.
Krupinski, E., & Locher, P. (1988). Skin conductance and aesthetic evaluative responses to non representational works of art varying in symmetry. *Bulletin of the Psychonomic Society, 26*, 355-358.
Lindauer, M. S. (1990). Reactions to cheap art. *Empirical Studies of the Arts, 8*, 95-110.
Malpas, J. (1997). *Realism*. London: Tate Gallery Publishing.
Markovic, S., & Gvozdenovic, V. (2001). Symmetry, complexity and perceptual economy: Effects of minimum and maximum simplicity conditions. *Visual Cognition, 8*, 305-327.
McWhinnie, H. J. (1993). Response time and aesthetic preference. *Perceptual and Motor Skills, 76*, 336-338.
Messinger, S. M. (1998). Pleasure and complexity: Berlyne revisited. *The Journal of Psychology, 132*, 558-560.
Munsinger, H., & Kessen, W. (1964). Uncertainty, structure, and preference. *Psychological Monographs: General and Applied, 78*, 1-24.
Neperud, R. W., & Marschalek, D. G. (1988). Informational and affect bases of aesthetic response. *Leonardo, 21*, 305-312.
Nicki, R. M. (1972). Arousal increment and degree of complexity as incentive. *British Journal of Psychology, 63*, 165-171.
Nicki, R. M., Lee, P. L., & Moss, V. (1981). Ambiguity, cubist works of art, and preference. *Acta Psychologica, 49*, 27-41.
Nicki, R. M., & Moss, V. (1975). Preference for non-representational art as a function of various measures of complexity. *Canadian Journal of Psychology, 29*, 237-249.

Osborne, J. W., & Farley, F. H. (1970). The relationship between aesthetic preference and visual complexity in abstract art. *Psychonomic Science, 19*, 69-70.
Parr, M. (1999). *Boring postcards*. London: Phaidon Press.
Parr, M. (2000). *Boring postcards USA*. London: Phaidon Press.
Saklofske, D. H. (1975). Visual aesthetic complexity, attractiveness and diverse exploration. *Perceptual and Motor Skills, 41*, 813-814.
Smith, L., Bousquet, S. G., Chang, G., & Smith, J. K. (2006). Effects of time and information on perception of art. *Empirical Studies of the Arts, 2*, 229-242.
Stamps, A. E., III. (2002). Entropy, visual diversity, and preference. *The Journal of General Psychology, 129*, 300-320.
Thomson, B. (1998). *Post-impressionism*. London: Tate Gallery Publishing.
Vitz, P. C. (1966). Preference for different amounts of visual complexity. *Behavioral Science, 11*, 105-114.
Winston, A. S., & Cupchik, G. C. (1992). The evaluation of high art and popular art by naive and experienced viewers. *Visual Arts Research, 18*, 1-14.

Direct reprint requests to:

Marcos Nadal
Department of Psychology
University of the Balearic Islands
Crta Valldemossa s/n, km 7,5
Palma de Mallorca 07122
Balearic Islands, Spain
e-mail: marcos.nadal@uib.es