British Journal of Psychology (2004), 95, 41–56

www.bps.org.uk

Individual and group modelling of aesthetic judgment strategies

Thomas Jacobsen*
Institut für Allgemeine Psychologie, Universität Leipzig, Germany

Individual differences in aesthetic judgments were investigated by comparing quantitative group and individual performance models of the judgment processes. Aesthetic judgments of beauty over novel, formal, graphic patterns were collected from 34 non-artist college students using a two-step ranking-rating procedure. Their judgment processes were individually modelled using Judgment Analysis. The participants showed noted individual differences. Certain features of the stimulus material, which were considered to contribute to the picture's beauty by one participant, were used in an opposing fashion by another. A group model was derived based on the average ratings of the patterns' beauty. It was concluded that the group model was not an adequate representation of the present data, whereas the data revealed systematic judgment processes at the individual subject level.

In ancient days, the Romans said, 'De gustibus non est disputandum'. There is no accounting for tastes. Fechner (1876) stated, in his seminal monograph, 'Vorschule der Aesthetik': 'Es ist eine alte Rede, dass sich ueber den Geschmack nicht streiten laesst; indes streitet man doch darueber, ja ueber nichts mehr als den Geschmack; es muss sich also doch darueber streiten lassen'.

> 'It is an old saying that there is no accounting for tastes, nevertheless people argue about it, about nothing more than taste; it must thus be possible to argue about taste'. (translation by the author; see also Mandler, 1982, p. 34 and Berlyne, 1971, p. 29).

This saying is still frequently used today, and versions of it exist in a large number of languages and countries. Hence, it seems worthwhile to ask under which empirical conditions it holds.

As there appears to be argument about taste, one may also expect to find a considerable body of empirical work using an idiographic approach to the study of aesthetic judgment. In fact, there are few such studies as most of the research effort in experimental psychological aesthetics has indeed been devoted to finding universal

*Correspondence should be addressed to Thomas Jacobsen, Institut für Allgemeine Psychologie, Universität Leipzig, Seeburgstraße 14–20, 04103 Leipzig, Germany (e-mail: jacobsen@uni-leipzig.de).

laws, to the nomothetic approach (Berlyne, 1971, 1974a, 1974b; Cupchik, 1992; Konecni, 1979; Martindale, 1988; see also Fechner, 1876). In this vein, Berlyne revived the interest in experimental aesthetics in the early 1970s with his biopsychological theory of aesthetic appreciation. He initiated the 'new experimental aesthetics' (Berlyne, 1974a, 1974b). More recently, his theory was challenged by Martindale's (1988; Martindale, Moore, & Borkum, 1990) cognitive theory of aesthetic processing.

Little is known about interindividual differences of aesthetic judgments within groups, as most of the work in this field was concerned with group differences. The typical study contrasted the performance of groups of experienced judges with the performance of groups of naive or inexperienced judges (e.g. Eysenck, 1972; Eysenck & Hawker, 1994; Locher, Smith, & Smith, 2001; Neperud & Marschalek, 1988; Nodine, Locher, & Krupinski, 1993; Winston & Cupchik, 1992; but see also Eysenck, 1970). There is also a considerable amount of literature based on personality structure research (e.g. Eysenck, 1972; Eysenck & Hawker, 1994).

This paper is concerned with individual differences in aesthetic judgments of beauty, judgments of novel, formal, graphic patterns in particular. Little research has focused on such individual differences of aesthetic judgments within homogeneous groups of individuals; that is, groups whose members were comparable with respect to specific training, cultural background, level of general education and social status. McManus, Jones, and Cottrell (1981), for instance, reported individual differences in aesthetic colour preferences. Whitfield (1984), also, found large individual differences represented by a low agreement within homogeneous groups in a colour-appropriateness ranking task. The author "anticipated that systematic differences would be more readily detected in within-group agreement than mean differences of groups" (Whitfield, 1984, p. 184). "This difference, however, is in internal consistency only and not in the mean ranks assigned to the colors for each style" (Whitfield, 1984, p. 185). Also, in the psychological research on the aesthetics of the golden section (see Green, 1995, for a review) no broad consensus as to whether or not the ratio of the golden section bears a preferred aesthetic property has been reached. However, temporally stable but idiosyncratic preferences have been reported (e.g. McManus, 1980). Likewise, McManus and Kitson (1995) found no group-level preference for compositional geometry in synthetic stimuli. In addition, details of results reported in the literature, such as huge error variances or null effects (e.g. Kubey & Barnett, 1989), for example, might also indicate individual differences. This, of course, may hold for certain types of stimuli, such as those mentioned in this paragraph, whereas it does not hold for others.

The present study was designed to investigate individual differences in aesthetic judgments of novel graphic patterns by comparing group and individual performance models which attempt to quantify the judgment process. In other words, the goal is to illuminate individual judgment processes that lead to individual differences within homogeneous groups by using the idiographic approach of individual case modelling, as well as group-level modelling. Thus, this method allowed the identification of agreement and disagreement between judges equally well.

For capturing individual judgment strategies, the framework of Social Judgement Theory (Hammond, Stewart, Brehmer, & Steinmann, 1975) was adopted, which defines judgment as a process that involves the integration of information from a set of cues into a judgment about some distal state of affairs. The theory provides a foundation for Judgment Analysis in a Brunswikian approach (e.g. Brunswik, 1952), and individual as well as group models are derived (e.g. Cooksey, 1996; Stewart, 1988). In accordance with Stewart (1988, p. 41), Judgment Analysis is defined as "using statistical methods to derive algebraic models of the judgment process". This is a paramorphic way of modelling, which is an input–output mapping of the judgment process under the premise that the material to be evaluated is perceptually salient. From this perspective, the perceptual process itself does not need to be modelled, and only input to and output from the judgment process is considered. Here, Judgment Analysis was used to assess the participants' ways of making aesthetic judgments. Stimulus features (cues) used by a participant for the aesthetic judgment process are included in the individual model, and some others are not. The relative judgmental importance is also established and reflected by the magnitude of the respective constituents of the model.

New stimulus material was systematically constructed for this study to accommodate a number of factors. Note that a study that controls external influences on the process of aesthetic judgment is more likely to find agreement between judges, because more possible sources of variation in the judgments have been ruled out by the experimental control. Eysenck (1988) suggested that better controlled studies in experimental aesthetics are required, if finding agreement between judges is the goal. In this vein, arrangements of line drawings of squares were constructed. These stimuli were simple in comparison with a large number of works of art, albeit adequate for this research (cf. e.g. work on the golden section using line segments or line drawing; e.g. Green, 1995). While being composed of familiar basic elements (e.g. Berlyne, 1970), the set of patterns was novel for all participants (Berlyne, 1970, 1974b; Cupchik, 1992), and thus did not allow a ready assimilation into existing cognitive schemata (Mandler, 1982) or allow for the recollection of prior aesthetic judgments stored in memory (cf. Kant, 1790, reflecting versus determining judgments). In terms of complexity (Berlyne, 1970, 1974b; Eysenck & Hawker, 1994; Nicki & Moss, 1975), the stimulus set as a whole, as compared with other aesthetic material like representational graphics or paintings, was fairly homogeneous. The individual items of the present set differed in the number of elements contained and, as a function of the latter, differed in complexity within the set. This was also formal, i.e. semantically shallow (Cupchik, 1992, p. 85; Martindale, 1988), and no stimulus was asymmetric (e.g. Eisenman, 1967; Jacobsen & Höfel, 2001). Financial interest, social status and cultural predication in general (Mandler, 1982) were unlikely factors influencing preference judgments over the present stimuli. As Hammond *et al*. (1975) put it, "all perceptual confusion" has to be eliminated to keep the processing "wholly judgmental in nature". Therefore, the stimuli were designed to have highly salient features so that measurable physical properties of the material, such as the size of the stimuli's basic elements, were available to the subject's cognitive system and used for the aesthetic judgments, to keep the task fully judgmental and hence prevent any confounding of perceptual and judgmental processes.

The study focused on two issues. First, the extent to which the present non-artist judges agree in their aesthetic judgment of beauty of novel, formal, graphic patterns was examined. Second, individual judgment strategies were examined. Paramorphic quantitative models determined which cues were used and in which manner these cues were employed by the individual judges.

Method

Participants

Nineteen female and 15 male students of the Freie Universität Berlin participated in the study. The average age was 26 years (ranging from 19 to 39). None of the participants was an art major or had otherwise received professional training in the arts. Eighteen

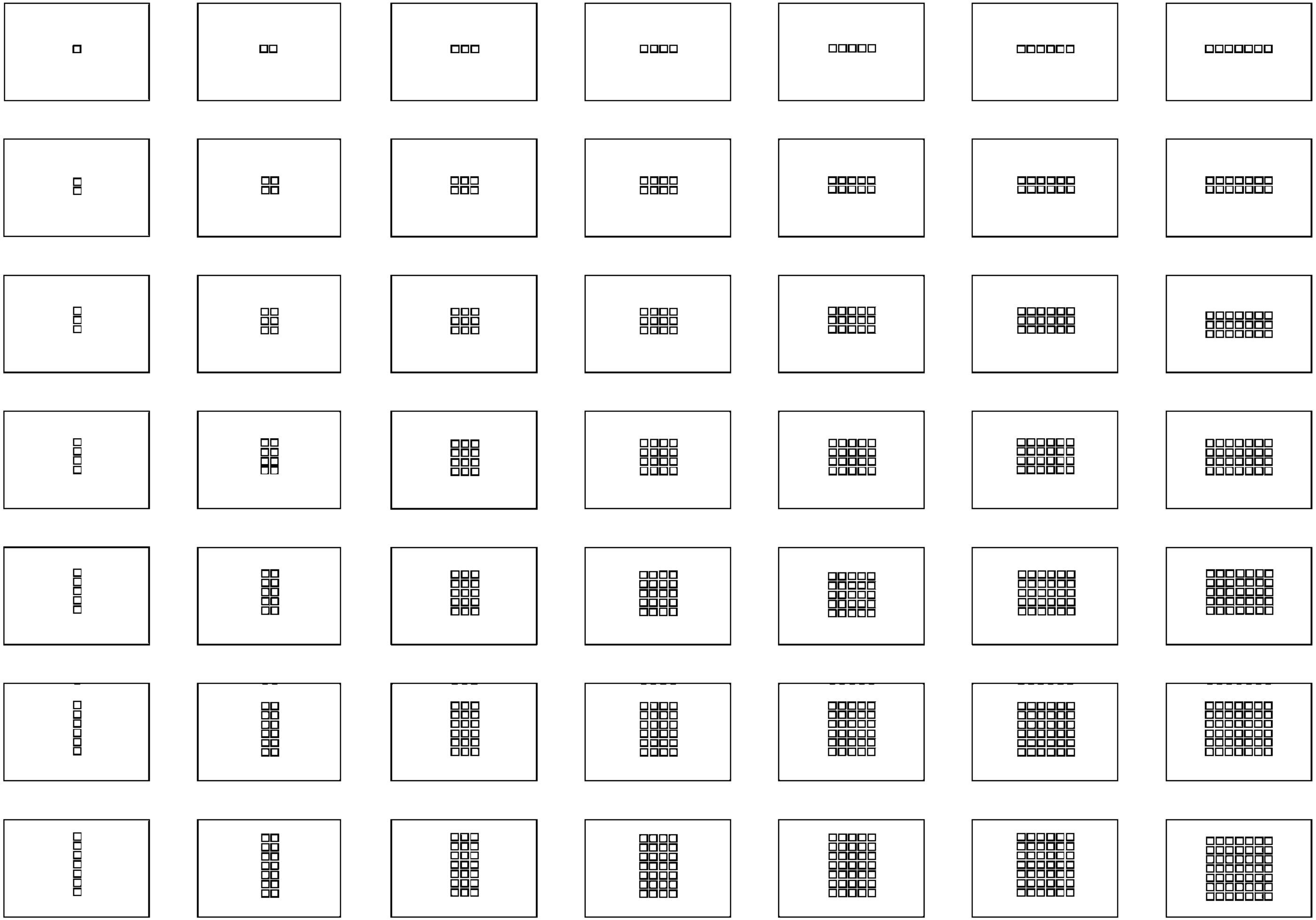

Figure 1. Complete set of stimulus material in miniature scale. Patterns are arranged in columns, starting with Picture 1 in the upper left-hand corner.

participants reported no interest at all in the arts or visual composition. The others reported little (11) to strong interest (5). The participants' self-reported interest in the arts and visual composition did not correlate with their age, gender, time on task or any of the dependent measures. Participants received course credit for their participation.

Material

A set of 49 pictures was designed in accordance with the construction criteria given above. A familiar and common basic element was chosen for the pictures, the line drawing of a square. The length of the edges of the square was 15 mm. Each picture consisted of an arrangement of one to 49 of these basic elements. The squares were placed in a grid of seven rows and seven columns. The grid had gaps of 5 mm between one row or column and the next. There were seven different pictures for each number of columns, using one, two, three, four, five, six or seven rows. Using this procedure, each picture showed a rectangle or a square consisting of at least one and up to 49 squares. Every arrangement of those basic square elements was positioned in the center of a DIN A4 sheet of cardstock. The pictures were consistently numbered in one corner on the back of the sheet. Figure 1 shows the material set in small scale.

The following perceptual cues obtained in the material were C 1 number of square elements in the arrangement, C 2 number of columns in the arrangement, C 3 number of rows in the arrangement, C 4 vertical symmetry axis between columns, C 5 horizontal symmetry axis between columns, C 6 proportion of width by height of the arrangement (a larger value denotes a more horizontally dominant arrangement; e.g. Pattern 43), C 7 proportion of height by width of the arrangement (a larger value denotes a more vertically dominant arrangement; e.g. Pattern 7) and C 8 proportion square, in a binary coding, i.e. dissociating arrangements that again formed a square from those that did not. Table 1 shows the intercorrelations of these cues. Note that the material was designed in such a way that the number of rows and the number of columns were orthogonal, resulting in zero correlations of a number of cues (Table 1). The symmetry measures were unrelated to the number of elements, rows and columns, as well as to each other.[1] The 'square proportion' is independent of the number of rows or columns, and the symmetry cues of a picture.

Table 1. Stimulus design: cues realized in the material and intercorrelations of cues

	c1	c2	c3	c4	c5	c6	c7	c8
c1 number of elements	–							
c2 number of rows	**.67**	–						
c3 number of columns	**.67**	.00	–					
c4 vertical symmetry	.00	.00	.00	–				
c5 horizontal symmetry	.00	.00	.00	.00	–			
c6 proportion width by height	−.23	**.47**	**−.64**	.00	−.18	–		
c7 proportion height by width	−.23	**−.64**	**.47**	−.18	.00	**−.49**	–	
c8 proportion square	.14	.00	.00	.00	.00	−.14	−.14	–

Note. Bold type denotes cue–cue combinations that were excluded from entering the same model.

[1] *Stimuli that have two symmetry axes may be processed differently to those that have only one. This double-symmetry feature applied to a number of the present stimuli. It was thus included as an additional cue in the Judgment Analysis and did not contribute to any of the models.*

Furthermore, based on the simple, familiar basic elements from which the patterns were constructed, it was assumed that the goal of perceptual saliency of the cues was achieved. This was mandatory, because the measurable physical properties of the stimuli were used as the 'cues' in the single case models. Also, the stimulus material was designed such that stimuli differ only in respect to a relatively small and well-controlled number of features—which were cues for the subject during the experiment.

Procedure

Participants were tested individually. They were given a booklet containing instructions and asked to read them carefully. The 49 pictures were then presented to each subject a random order. They were instructed to look at the pictures in landscape format only, without turning them upside down. They were asked to study the pictures carefully, without any time pressure. Participants were instructed that they were otherwise free to browse through the pictures. They could change their order, put them on different piles or lay them out in front of themselves. This served to ensure that the participants studied the stimuli adequately to become acquainted with the 49 pictures.

Rank order

After participants had studied the pictures carefully, they were asked to rank them in descending order according to their beauty, their aesthetic property. This rank ordering procedure was chosen, because the more reliable method of paired dominance comparisons ($k = 49$; 1,176 comparisons) was impractical for this study. Note that using a much smaller number of pictures that would have made paired dominance comparisons possible was prohibited here because of the methodological requirements of Judgment Analysis. The rank-order instructions allowed the participants to drastically reduce (as compared with fully paired dominance comparisons) the number of comparisons they had to make. They could form rough stacks first and then make a more fine-grained ranking judgment by paired dominance comparisons within these stacks whenever they felt these were necessary.

Bortz (1984, p. 126) reported that investigators are more likely to obtain reliable data using a rating procedure if the participants are familiar with material, and sequence effects are thus prevented. As a consequence, the ranking procedure was intended to improve the quality of the subsequently gathered rating data, because it ensured that participants were familiar with the material before they started to rate the pictures.

Furthermore, the rank order could be used later to estimate the subject's consistency in judging the aesthetic property of pictures by computing the Kendall's tau rank correlation between the rank order and the ratings introduced below. Note that this study was not designed to make statements about the temporal stability of the participants' judgments (see Discussion).

Rating

The participants' second task was to rate the degree of the pictures' aesthetic property on a 7-point rating scale. Participants were again instructed to make their judgments according to the perceived beauty of the pictures. They were given a relative instruction (Russell & Gray, 1994), which meant that participants were asked to make their judgments of beauty exclusively taking into consideration the set of 49 pictures presented and not in relation to any other (aesthetic) objects, e.g. works of art, they knew.[2] Finally, participants were asked to report how they judged the degree of the aesthetic property of the pictures.

Results

To accommodate the two main questions asked in this study, how the individual judges process their aesthetic judgments over the stimuli presented and to what extent participants agree in these judgments, the analysis procedure pursued had two qualitatively distinct levels. On the first level, the Judgment Analysis (Cooksey, 1996; Stewart, 1988), individual case models of the participants' ways of judging the aesthetic value of the pictures presented were derived. On the second level, a group model was derived, and a group analysis determined the degree of agreement between the participants' judgments. Group level results on the number of elements in a pattern, the square proportion, symmetry, and horizontal versus vertical dominance in the picture are described. Subsequently, a Monte Carlo study with the purpose of securing the results against artefacts due to chance is reported.

Verbal reports

Two independent judges evaluated the verbal reports. They agreed completely that all participants gave useable, task-appropriate rationales for their aesthetic judgments. No participant reported that they could not perform the task. Only three participants reported that they did not find the stimuli very appealing as such, and that they could only perform the task by judging the stimuli one relative to the others (see Procedure, relative instruction). For 22 participants, verbal reports of judgment strategies showed convergence with the Judgment Analysis models. This convergence is shown in bold in Table 2. Examples of verbal reports are given below.

Judgment Analysis

Judgment Analysis is defined as using statistical methods to derive algebraic, paramorphic models of the judgment process (see above, Stewart, 1988). In accordance with this definition, multiple-regression equations using the cues' values for the pictures as predictors and the participants' ratings as the criterion were computed for every single subject using a constraint stepwise method. The cue explaining most of the variance was entered into the model first. Cues explaining incremental variance were entered into the model, if there was no substantial cue–cue intercorrelation ($r < .25$; see Table 1) and $p < .005$ to avoid accumulation of alpha errors due to the number of tests performed per participant.

A substantive model was obtained for every participant ($N = 34$). The median R was

[2] *The 7-point scale was preferred over a preference scale since the pictures were quite simple, and participants could not be expected to agree to a point on a rating scale labelled 'like very much indeed'. For the same reason, the relative instruction (Russell & Gray, 1994) was given. Moreover, a preference instruction might induce judgments other than of purely aesthetic nature, because participants are likely to fill in a notion of 'prefer for what' or 'in which context' themselves. In general, an individual might prefer an otherwise less beautiful item over a very beautiful one for economical, practicability or other reasons (see e.g. Boselie, 1991, p. 66, for a discussion of this preference problem). These were factors that could not be fully controlled by means of stimulus design. Also, this study had to rely on verbal communication of the concept of aesthetics via the word 'beauty' (see also Jacobsen & Schröger, unpublished), because of the lack of an external definition of aesthetic judgment and an independent external criterion. This was a rather general problem for studies in the experimental aesthetics and cannot be considered especially problematic for this study in particular.*

.71 (min .29, max .92). Table 2 shows the multiple regression models containing the beta weights. The beta weights obtained in the multiple-regression analysis reflect the participant's relative cue use. Converging with their verbal reports, Participants 03, 08, 12 and 17 adopted a strategy of judging patterns containing more elements less beautiful. For example, Participant 17 reported regarding their cue use: "simplicity: arrangements with less elements I liked more", whereas, Participants 05, 09, 18 and 21 adopted the opposite strategy, and patterns containing more elements were deemed more beautiful. This was also confirmed by their verbal reports. Participant 09, for instance, reported that they "did not like arrangements with just a few or just one square". For these six judges, the number of elements was the only significant cue. Strategies making use of two or more cues were also adopted. For example, Participant 04 judged larger square arrangements more beautiful than other patterns. Participant 29, for instance, reported that they used "1) harmony of the arrangement, i.e. the square is best; 2) ratio of height by width; 3) ratio of size of arrangement and the paper (constant); 4) height and width of the arrangement on the paper". This verbal report also converged with the regression model. Moreover, the amount of linearly explained variance in the participants' ratings varied considerably (see Table 2).

One estimate of the participants' consistency was Kendall's tau b correlation between the participants rankings and ratings. The mean tau b was .73 (*SD* tau b = .19, min = .22, max .=.93) The correlation of tau b and multiple R was substantial with $r = .40$. This indicates that, to a certain degree, the more consistent participants produced the more systematic ratings. It was certainly not the case that the more consistent participants were less linearly predictable.

One of the prerequisites of the multiple regression is the normal distribution of predictor and criterion values. However, the method of multiple regression is known to be very robust, even if this prerequisite has been violated. Kolmogoroff–Smirnoff (K–S) values for the fit of the distribution of ratings to the normal distribution were computed for every single participant. Only three K–S values missed significance (Lillefors statistic), indicating that most of the participants did not produce normally distributed ratings. It can now be argued, that the non-normal distribution of the ratings might increase the multiple correlation coefficient artificially and thereby produce artificially high degrees of linearly explained variance in the criterion. For this reason, the correlation between the participants multiple R and the K–S value was computed. The non-significant $r = -.06$ ($N = 34$; $p = .73$) suggests that there is no positive correlation between K–S value and multiple R. In fact, there was no correlation at all.[3]

A low multiple correlation between the cues and the participant's ratings could have been due to the participant's way of configurally combining the cues for their own judgments, i.e. using a rather holistic judgment criterion (e.g. Participants 11, 23, 24 and 33) or to poor, unreliable performance of the tasks resulting in noisy data. Hence, the possible alternative explanation for the inter-individual differences in R is the accuracy with which the participants performed the task of ranking and rating the pictures according to their degree of aesthetic value to them. The assumption that an unreliable judge in this experiment would spend less time than a better judge (possibly to leave the

[3] *To investigate this point further, a small Monte Carlo study was conducted. One single ranking of the pictures used in this study was projected into six different sets of ratings. These differed in the degree of their fit to the normal distribution, but all maintained the same initial ranking. These simulated ratings were entered into a Judgment Analysis as described above. The deviation from the normal distribution did not increase the amount of explained variance. On the contrary, the greater the deviation from the normal distribution, the smaller was the multiple R. Using this procedure, it could be shown that participants that produced ratings deviating from the normal distribution did not cause artificially high multiple Rs by doing so.*

Table 2. Multiple-regression-models

Participant	Number of elements	Number of columns	Number of rows	Vertical symmetry	Horizontal symmetry	Proportion width by height	Proportion height by width	Proportion square	R
P01						**−.45**			.45
P02							**−.77**		.77
P03	**−.84**								.84
P04	.40							.53	.70
P05	**.84**								.84
P06	.54					**−.34**			.71
P07		−.45							.45
P08	**−.44**								.44
P09	**.79**								.79
P10			−.53						.53
P11					−.29				.29
P12	**−.60**								.60
P13		**.54**						.54	.76
P14			**.40**						.40
P15		.63							.63
P16			**.40**						.40
P17	**−.87**								.87
P18	.73								.73
P19		−.32	−.64					.30	.90
P20	.78						−.25		.87
P21	**.92**								.92
P22	.72							**.33**	.83
P23			−.30						.30
P24							−.41		.41
P25								**.55**	.55
P26	.60							**.40**	.76
P27		**.72**						**.46**	.86
P28	**.66**							**.44**	.84
P29	**.62**						**−.24**	**.31**	.82
P30							**−.45**	**.34**	.60
P31	**.32**						−.23	**.67**	.86
P32	.47								.47
P33	.44								.44
P34								**.42**	.42
Group	.44						−.36	.43	.82

Participants are shown in the rows, and cues are shown in the columns. Cells contain the beta weights of the multiple-regression models. The last column contains the multiple R value. Cues converging with participants' verbal reports of judgment strategy are shown in bold.

experimental situation earlier) seems sensible. If a lower multiple R is due to poor performance, and a poor performer finishes the tasks faster than a good performer, the correlation between multiple R and time on task would be expected to be positive. The opposite was, indeed, the case. The correlation between time on task and R was $-.39$ ($N = 34$; $p = .031$). This substantial negative correlation indicated that the faster participants were not the less linearly predictable ones, as one might expect. On the contrary, the faster participants were the more systematic ones. They tended to employ rules and thus were more linearly predictable and faster. The verbal reports confirmed this view (Table 2). Furthermore, there was a correlation of $r = -.28$ ($N = 34$; $p = .056$)

between degree of holism of a participant's judgment strategy, as derived from the verbal report by ratings of two independent judges, and the models' R value. This also suggests that there was a trend that systematic rule use resulted in stronger linear models than holistic processing. Participant 23, for example, gave a holistic rationale for his/her judgments: "I have decided completely based on my subjective impression". Participants that spend more time on performing their aesthetic judgments hence tended to perform the task in a more holistic, less linearly accountable way. Taken together, this analysis argues against a poor-performance account.

Group analysis

The median ranking data, their ranges, the average rating data, their standard deviations and ranges of an item-based analysis are shown in Table 3. The Judgment Analysis group model based on the average ratings is shown in the last row of Table 2. This multiple-regression equation was derived using the method for the individual case models described above.

The time the participants spent on the task, viewing, rank ordering, and rating the patterns, was recorded. The average time on each task was 54 min, *SD* 17, min 30, and max 90. The correlations between the sets of judgments made by two participants reflect the overall agreement between these two participants. First, the Spearman and Kendell's tau rank correlations of all combinatorily possible pairs of participants' rankings ($N = 34$, $k = 561$) were computed, mean Kendell's tau = .11 (median tau = .11). The standard deviation of the distribution was .32 (min = − .78, max = .79; mean Spearman correlation .14, median .15), and the standard deviation of the distribution was .43, min = − .92, max = .94). Second, the Pearson correlations of all combinatorily possible pairs of participants' ratings ($N = 34$, $k = 561$) were computed, mean Pearson correlation $r = .13$ (median $r = .13$; standard deviation of the distribution .38, min $r = -.90$, max $r = .95$). The results showed that there were participants who maximally agreed (correlation of .95) and participants who maximally disagreed for the rankings and for the ratings. Overall, the agreement is close to null.

The reason for this immense variation becomes obvious from an inspection of the single case models in Table 2. The participants as a group used seven out of eight cues, all of which were employed as the primary cue by at least one judge; that is, there was no vast agreement among the participants as to which features of the stimuli were critical for the aesthetic evaluation of the pictures. Furthermore, if participants used the same cue for their judgment, they still may have used it in the opposite direction. The cue 'number of squares' (c1), for example, was used by 16 participants as the primary indicator of beauty. For 75% of them, a picture was more beautiful when it contained more elements. Twenty-five percent used this same cue in the reverse fashion (Participants 03, 08, 12, and 17). For the cues number of columns and number of rows, this contrast was even stronger. About half of the individual judges are not adequately represented by the group model, counting individuals whose primary cues did not enter the group model or were opposite the average usage.

Apart from these marked individual differences, there were also group-level results that showed agreement between participants. The majority of judges who used the number of elements cue preferred larger arrangements over smaller ones. If participants used the square cue, it always had a positive correlation, indicating that a square arrangement of square was preferred. However, extreme proportions of vertically or horizontally dominant pattern were rejected by all participants who made use of these cues.

Table 3. Item-based results. median rank, range of rankings, mean ratings, standard deviation and range of ratings

Picture	Median	Range	Mean	*SD*	Range
1	32.0	48.0	3.6	2.4	6.0
2	43.5	48.0	2.4	1.8	5.0
3	37.5	47.0	2.9	1.9	6.0
4	42.0	46.0	2.5	1.8	6.0
5	42.0	44.0	2.5	1.8	6.0
6	41.0	43.0	2.1	1.4	5.0
7	43.5	48.0	2.1	1.5	6.0
8	36.5	47.0	3.1	2.2	6.0
9	13.0	48.0	4.5	2.0	6.0
10	25.0	34.0	3.7	1.4	6.0
11	29.0	46.0	3.2	1.6	6.0
12	29.5	42.0	3.2	1.6	6.0
13	32.0	38.0	2.9	1.7	6.0
14	35.0	42.0	2.6	1.5	6.0
15	26.5	46.0	3.5	1.9	6.0
16	24.0	41.0	3.5	1.4	5.0
17	9.5	44.0	5.3	1.5	4.0
18	19.0	42.0	3.8	1.4	5.0
19	23.0	38.0	4.0	1.6	6.0
20	25.5	43.0	3.6	1.4	6.0
21	25.5	48.0	3.3	1.7	6.0
22	37.5	42.0	2.7	1.5	6.0
23	27.5	42.0	3.4	1.4	5.0
24	18.5	43.0	4.2	1.6	6.0
25	12.5	45.0	4.5	1.9	6.0
26	20.0	43.0	3.9	1.6	6.0
27	25.0	43.0	3.6	1.6	6.0
28	25.0	45.0	3.7	1.8	6.0
29	37.5	36.0	2.9	1.7	6.0
30	31.5	41.0	3.3	1.5	6.0
31	21.0	40.0	4.2	1.5	6.0
32	17.0	42.0	4.4	1.6	6.0
33	14.0	45.0	4.9	1.8	6.0
34	21.5	41.0	4.1	2.0	6.0
35	20.5	46.0	4.1	1.9	6.0
36	38.0	47.0	2.4	1.4	5.0
37	33.5	42.0	2.9	1.6	6.0
38	21.0	44.0	4.0	1.7	6.0
39	16.5	44.0	4.3	1.9	6.0
40	13.5	44.0	4.3	1.9	6.0
41	12.0	47.0	4.4	2.0	6.0
42	20.5	45.0	4.0	1.9	6.0
43	35.0	45.0	3.0	1.9	6.0
44	32.5	45.0	2.7	1.6	6.0
45	24.5	40.0	3.5	1.8	6.0
46	19.0	46.0	3.8	2.0	6.0
47	14.5	44.0	4.3	2.0	6.0
48	17.0	47.0	4.5	2.1	6.0
49	15.5	48.0	4.8	2.2	6.0

Number of elements

Most of the participants used the number of elements as the primary cue. If only the group tendency is considered, more elements tended to make the pattern more beautiful. This was also reflected in the group model in which number of elements was the strongest predictor.

Square

Seven of the 49 pictures of the stimulus set showed a special congruency between elements of the pictures and their arrangements. In these pictures, the arrangements of the squares formed a square again (cue C6). This cue was used in a positive manner by all judges who did use it. It was also the primary cue for 14% of the participants and the second strongest predictor in the group model.

Proportion

The ratio of width by height of the arrangements of squares in the pictures (cue 4) can be viewed as coding horizontal and vertical dominance of the pictures. The ratio is greater than zero for horizontally dominant pictures and lower than zero for vertically dominant pictures. The more the ratio deviates from zero, the more dominant a picture is in one direction. More extreme proportions were judged more negatively by the judges who used that cue. This held to a larger degree for vertically dominant patterns that were positioned orthogonally to the landscape format of the paper. Furthermore, all patterns in the stimulus set could be considered symmetrical. This fact seemed to be sufficient for the judges, as symmetry cues (c4 and c5) were used only once.

Monte Carlo Study

A Monte Carlo study was conducted to investigate whether the constellation of eight possible cues could lead to substantial models for random data, just because of their inherent redundancy and thus resulting potential to describe possible data sets. For this purpose, the rating values of the 34 participants were randomly assigned to the 49 pictures. This procedure was chosen to preserve the shapes of the distributions of the participants' ratings (see K–S analysis above). Thirty-four multiple-regression models were computed over these Monte Carlo subjects using the same method as for the participants. Eleven significant models resulted. The median multiple R of these Monte Carlo subjects was $R = .33$ (min .29, max .39). Using a binomial test, it was shown that 34 out of 34 models, as encountered with the actual data, would reach significance by chance with a probability of $p < .000001$. Selection of cues and positive or negative direction of cue use were not systematic in the Monte Carlo study.

Discussion

Aesthetic judgments of beauty of 49 novel, formal graphic patterns were collected from non-artist participants. In the framework of Social Judgment Theory (Hammond *et al.*, 1975), the data were subjected to individual Judgment Analyses deriving quantitative paramorphic models of the judgment processes. These models reflected the individual's strategy of aesthetic judgment, cue use, and degree of linear judgment capture, largely converging with participants' verbal reports of judgment strategies. A group model was constructed using the same method applied to group-average rating data. Filtering out individual judgment strategies, the group-level analysis showed that larger arrangements of the basic element, with a tendency towards forming a square itself, while not being vertically dominant, were preferred overall.

Judgment Analysis

This study obtained noted individual differences in aesthetic judgments of beauty collected from a homogeneous group of non-artist college students. The judgments of a number of participants resulted in inter-individually contrary rankings and ratings, while being intraindividually consistent. This pattern of results was obtained despite the strict experimental control of the stimulus material. The set of stimuli was novel while consisting of familiar elements (e.g. Berlyne, 1970), formal (e.g. Martindale, 1988), symmetric (e.g. Eisenman, 1967) and not of social relevance (e.g. Mandler, 1982; see Introduction). As a result, the stimuli were rather simple and showed reduced feature dimensions compared with most works of art. Nonetheless, they were adequate for the aesthetic judgment task, as the verbal reports confirm (cf. e.g. Fechner, 1876). According to these methodological considerations (e.g. Eysenck, 1988; see above), this study was more likely to find agreement between judges than a study with a poorer experimental control and richer stimuli. However, this was not the case. Some participants used simple judgment rules, while others employed rather holistic criteria. It is unlikely that the degree of inter-individual variation decreases when more stimulus dimensions come into play (e.g. Eysenck, 1988). Thus, it is even more significant that these individual differences were obtained. Of course, studies using richer stimuli while maintaining strict experimental control can be conducted (e.g. Locher *et al.*, 2001; McManus & Kitson, 1995). Judgment Analysis can also be applied to such stimuli, if proper cues can be identified, cue–cue inter-correlations are controlled and models are appropriately restricted (e.g. Jacobsen & Höfel, 2001). For richer stimuli, the nature of the cues used for aesthetic judgments is likely to shift. Numerous, manifold cues will elicit different strategies. This deserves further empirical investigation.

By using Judgment Analysis (e.g. Cooksey, 1996), it is possible to model individual judgment strategies without asking for a verbal report. The beta weights of the regression equations derived from the participant's judgments reflect the relative use of the cues obtained in the stimulus material. These judgment strategies, identified by the models, largely converged with participants' verbal reports. That is, the paramorphic models show how participants derived their aesthetic judgments. As a consequence, a way of studying the accuracy of self-reports may become available here. Furthermore, in accordance with, Brehmer and Joyce (1988), for instance, the results also show that most individual judges make use of only a limited number of cues. Participants whose judgment strategies were less well accounted for by the model tended to spend more time on the task and employed more holistic judgment criteria.

It is also argued that the present results are not unlikely to be based on error variance, because completely contrary judgments were produced by the participants who also showed a high internal consistency at the same time. Opposite cue uses with higher correlations than in the Monte Carlo simulations were obtained. The Monte Carlo simulation suggests that a small number of the less systematic models may have been based on chance sampling. The fact that time on task tended to be longer for these participants argues against this view. The Monte Carlo also suggests that the majority of the stronger linear models were not based on chance sampling effects.[4]

Group analysis

There was a group-level tendency towards larger arrangements of elements, square proportion, against notably horizontally or, especially, vertically dominant patterns. However, about half of the individual judges are not adequately represented by the group model, counting individuals whose primary cues did not enter the group model or were opposite the average usage. This pattern indicates that the derivation of a group model was problematic, because of the noted individual differences. These, however, were effectively filtered out by the averaging procedure. As a consequence, the group model does represent the subgroup of the judges who agree that larger patterns are more beautiful than smaller patterns (c1, .44), more vertically dominant a patterns are less beautiful (c7, −.36), and square patterns are more beautiful than non-square patterns (c8, .43). Agreement in the sub-group was captured well by the group model, while covering only about half of the judges.

The present data converge with previous findings of individual differences in judgments of basic stimuli in aesthetics. As introduced above, there is no general agreement about the beauty of the golden section proportion (e.g. Green, 1995; McManus, 1980). Idiosyncratic preferences, however, were obtained (McManus, 1980). The author also reported a preference for the square proportion, which converges with the present results. Likewise, McManus and Kitson (1995) found no group-level preference for compositional geometry in synthetic stimuli. Furthermore, individual differences for colour preferences have been reported (McManus *et al.*, 1981; Whitfield, 1984). Also, no group-level preference was observed in the present data as to whether horizontally or vertically oriented arrangements were generally preferred. This finding converges with the results of Kubey and Barnett (1989), who did not find any preference for an orientation of horizontally or vertically presented rectangles. Eysenck (1988) suggested solving the problem of objectivity of aesthetic judgments by using average judgments. The averaging procedure, however, would not yield satisfactory results for some of the past, and especially the present, data, because it constitutes a strong filter that would camouflage the systematic, idiosyncratic differences in the aesthetic judgments.

Conclusion

Using novel, formal, graphic patterns, the present study was of limited scope. Aesthetic judgments are frequently performed on richer stimuli being influenced by a host of factors as evolutionary, cognitive, emotional, situational, temporal, social, cultural and historical factors. As a consequence, it is unlikely that a larger number of influential dimensions will result in less variation and more agreement. Thus, it also appears to be reasonable that some nomothetic studies may have camouflaged noted individual

[4] *Furthermore, participants might handle the ranking and the rating of the pictures as different aesthetic judgment tasks and not as an identical task, in respect to the aesthetic judgment that simply features two different answering formats. The ranking task is inherently relative within the material set, whereas the experimenter has to rely on the participants' compliance in this respect for the rating task. The relatively low consistency that a few participants showed, on the one hand, might indicate poor performance or it might, on the other hand, suggest that the two tasks, which were meant to require identical judgments just in different answering formats, were actually different tasks for some participants.*

differences by using data averaging. Individual case modelling can capture these differences. The present study also derived a group model based on data averages. This model, however, could sufficiently account for only half of the participants, whereas the individual models gave a much more precise account. Hence, there may be some debate about the mere nomothetic approach being justified, given such a data pattern. Thus, it is argued that the idiographic approach should be additionally adopted, if such an equivocal empirical situation is encountered. In that sense, there is (no) accounting for taste.

Acknowledgements

I thank Volker Bosch, Karl Christoph Klauer, Heinz-Martin Süß, Trevor Penney, Chris McManus, and three anonymous reviewers for helpful comments on an earlier version of the manuscript, Katharina Buchta and Michael Köhler for assistance with the analysis of verbal reports, and Maria Etzkorn.

References

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

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

Berlyne, D. E. (1974a). The new experimental aesthetics. In D. E. Berlyne (Ed.), *Studies in the new experimental aesthetics: Steps toward an objective psychology of aesthetic appreciation.* Washington, DC: Hemisphere.

Berlyne, D. E. (1974b). Novelty, complexity and interestingness. In D. E. Berlyne (Ed.), *Studies in the new experimental aesthetics: Steps toward an objective psychology of aesthetic appreciation.* Washington, DC: Hemisphere.

Bortz, J. (1984). *Lehrbuch der empirischen Forschung für Sozialwissenschaftler.* Berlin: Springer.

Boselie, F. (1991). Against prototypicality as a central concept in aesthetics. *Empirical Studies of the Arts*, *9*, 65–73.

Brehmer, B., & Joyce, C. R. B. (Eds.). (1988). *Human judgment: The SJT view.* Amsterdam: North-Holland.

Brunswik, E. (1952). *The conceptual framework of psychology.* Chicago: University of Chicago Press.

Cooksey, R. W. (1996). *Judgment analysis: Theory, methods, and applications.* San Diego, CA: Academic Press.

Cupchik, G. C. (1992). From perception to production: A multilevel analysis of the aesthetic process. In G. C. Cupchik & J. Laszlo (Eds.), *Emerging visions of the aesthetic process: Psychology, semiology, and philosophy.* Cambridge: Cambridge University Press.

Eisenman, R. (1967). Complexity–simplicity: I. Preference for symmetry and rejection of complexity. *Psychonomic Science*, *8*, 169–170.

Eysenck, H. J. (1970). Training in art as a factor in the determination of preference judgments for polygons. *British Journal of Psychology*, *61*, 65–81.

Eysenck, H. J. (1972). Personal preferences, aesthetic sensitivity and personality in trained and untrained subjects. *Journal of Personality*, *40*, 544–557.

Eysenck, H. J. (1988). Personality and scientific aesthetics. In F. H. Farley & R. W. Neperud (Eds.), *The foundations of aesthetics, art and art education.* New York: Praeger.

Eysenck, H. J., & Hawker, G. W. (1994). The taxonomy of visual aesthetic preferences: An empirical study. *Empirical Studies of the Arts*, *12*, 95–101

Fechner, G. T. (1876). *Vorschule der Aesthetik. Erster Teil.* Leipzig: Breitkopf & Haertel.

Green, C. D. (1995). All that glitters: A review of psychological research on the aesthetics of the golden section. *Perception*, *24*, 937–968

Hammond, K. R., Stewart, T. R., Brehmer, B., & Steinmann, D. O. (1975). Social judgment theory. In M. Kaplan & S. Schwartz (Eds.), *Human judgment and decision processes.* New York: Academic Press.

Jacobsen, T., & Höfel, L. (2001). Aesthetics electrified: An analysis of descriptive and symmetry and evaluative aesthetic judgment processes using event-related brain potentials. *Empirical Studies of the Arts*, *19*, 177–190.

Jacobsen, T., & Schröger, E. (unpublished). On describing the aesthetics of objects: A semantic field analysis.

Kant, I. (1790). *Critik der Urteilskraft.* Berlin: Lagarde und Friedrich.

Konecni, V. J. (1979). Determinants of aesthetic preference and effects of exposure to aesthetic stimuli: Social, emotional and cognitive factors. *Progress in Experimental Personality Research*, *9*, 149–197.

Kubey, R. W., & Barnett, A. L. (1989). Esthetic preference for rectangles of vertical and horizontal orientation. *Bulletin of the Psychonomic Society*, *27*, 239–240.

Locher, P. J., Smith, J. K., & Smith, L. F. (2001). The influence of presentation format and viewer training in the visual arts on the perception of pictorial and aesthetic qualities of paintings. *Perception*, *30*, 449–465.

Mandler, G. (1982). The structure of value: Accounting for taste. In M. S. Clark & S. T. Fiske (Eds.), *Affect and cognition, The Seventeenth Annual Carnegie Symposium on Cognition.* Hillsdale, NJ: Erlbaum.

Martindale, C. (1988). Aesthetics, psychobiology, and cognition. In F. H. Farley & R. W. Neperud (Eds.), *The foundations of aesthetics, art and art education.* New York: Praeger.

Martindale, C., Moore, K., & Borkum J. (1990). Aesthetic preference: Anomalous findings for Berlyne's psychobiological theory. *American Journal of Psychology*, *103*, 53–80.

McManus, I. C. (1980). The aesthetics of simple figures. *British Journal of Psychology*, *71*, 505–524.

McManus, I. C., Jones, A. L., & Cottrell, J. (1981). The aesthetics of colour. *Perception*, *10*, 651–666.

McManus, I. C., & Kitson, C. M. (1995). Compositional geometry in pictures. *Empirical Studies of the Arts*, *13*, 1995, 73–94.

Neperud, W. R., & Marschalek, D. G. (1988). Informational and affect bases of aesthetic response. *Leonardo*, *21*, 305–312.

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.

Nodine, C. F., Locher, P. J., & Krupinski, E. A. (1993). The role of formal art training on perception and aesthetic judgment of art compositions. *Leonardo*, *26*, 219–227.

Russell, P. A., & Gray, C. D. (1994). Ranking or rating? Some data and their implications for the measurement of evaluative response. *British Journal of Psychology*, *85*, 79–92.

Stewart, T. R. (1988). Judgment analysis: Procedures. In B. Brehmer & C. R. B. Joyce (Eds.), *Human judgment: The SJT view.* Amsterdam: North-Holland.

Whitfield, A. (1984). Individual differences in evaluation of architectural colour: Categorization effects. *Perceptual and Motor Skills*, *59*, 183–186.

Winston, A., & Cupchik, G. C. (1992). The evaluation of high art and popular art by naive and experienced viewers. *Visual Arts Research*, *18*, 1–14.

Received 13 May 2002; revised version received 24 February 2003