

A STUDY OF ESTHETIC JUDGMENTS

BY KATE GORDON

Introduction.—The most fundamental question of esthetics, and the most tantalizing, is the question of a standard. Are beautiful things really beautiful, or do we only think them so, because they give us a certain kind of pleasurable feeling, feeling which we have been taught to call ‘disinterested,’ ‘immediate,’ ‘universal,’ etc.? Is beauty constant? That is, are there any common external characteristics which beautiful objects must have in order to arouse such feelings? If there are none, then all men are equal in the value of their esthetic judgments, because then we have no objective facts with which to compare or verify their judgments. But if there are such external characteristics, they are the basis of a standard of beauty. The very existence of schools of art, and of rules of composition and of criticism seem to imply a belief in a standard. If the standard is there, how do we approach it to know it? These reflections led me to try to find out how well different persons would agree in their estimates of the relative beauty of the objects in a concrete series of color designs.

The Material Used.—For this purpose fifty colored plates of oriental rugs were used. All were printed on cards 4 by 6 inches in dimensions. The pictures were excellent reproductions of fine rugs. They appeared to afford a genuine esthetic experience to most of those persons who took part in the experiment. The rugs appeared to have a fairly wide range of esthetic values, though it is probable that the difference between ‘best’ and ‘worst’ represents a different emotional range for the different subjects.

Manner of Presentation.—The pictures were divided into two series in the following way. Three persons arranged the fifty pictures in an order of merit and their judgments were averaged. The pictures which stood first, third, fifth, etc.

were then called 'Series One,' and those which stood second, fourth, sixth, etc., were called 'Series Two,' thus making two series of 25 members each. This was done to see whether results obtained with Series One could be repeated with Series Two. The pictures of a given series were laid in chance order on a gray background, and to the person who was to judge them was handed a card of typewritten instructions as follows:

Here are some pictures of Oriental Rugs. Please arrange them in the order of their beauty. At the extreme left put the one which seems to you to be the most beautiful, next to it the one which is second in beauty, etc., and at the extreme right the one which is least attractive. Do not try to think of rules or theories of art. Judge them by the pleasure or the displeasure they give to you. Take all the time you want, and make any rearrangements you wish.

Each person arranged the two series separately, and the order of choice was recorded by me.

The Subjects.—The number of persons who gave judgments was 207. All, except one negro, were Caucasians. About one half were undergraduate students in the Carnegie Institute of Technology. Among the others were a few graduate students, and members of the faculty and staff of the same institution. There were a few students of the University of Southern California. The rest were men and women with no academic connection. The latter were probably above the average in artistic training. Out of the whole group there were 20 who may be looked upon as 'experts.' I have counted as experts those whose professional duties demand some exercise of judgment in the matter of color. There were among the 20 some 11 persons engaged in art instruction as professors or teachers or supervisors, 3 rug dealers of more than local reputation, 3 painters, and 3 persons in professional connection with painting departments of museums.

Wide Diversities of Individual Judgments.—The first striking result is the great variation in opinion as to the relative merits of the rugs. Every rug, without exception, was rated very high by some persons and very low by others, that is, it was put within 3 places of the top and the bottom of the series respectively. Nearly every rug was put quite at

the top and at the bottom by different judges. The following distributions of choices will illustrate the variations.

TABLE I
DISTRIBUTIONS OF CHOICES ON RUGS

Place Given	Rug 1. No. of Persons	Rug 5. No. of Persons	Rug 8	Rug 17	Rug 22	Rug 25
1.....	24	6	29	2	0	1
2.....	25	6	12	4	2	3
3.....	24	8	12	4	2	1
4.....	21	16	11	3	3	1
5.....	11	18	6	6	4	3
6.....	20	15	4	5	2	2
7.....	14	12	6	6	7	2
8.....	9	10	3	3	11	5
9.....	4	12	8	12	1	1
10.....	6	8	9	10	5	3
11.....	7	8	4	14	7	6
12.....	2	7	7	9	7	8
13.....	6	11	1	10	8	2
14.....	3	6	5	9	12	5
15.....	5	7	7	15	7	5
16.....	4	10	7	13	13	3
17.....	4	3	6	15	17	12
18.....	3	7	3	10	18	6
19.....	1	9	4	13	12	11
20.....	3	5	5	14	12	7
21.....	1	6	8	9	14	12
22.....	2	5	8	4	10	16
23.....	1	3	14	8	9	14
24.....	0	2	13	2	10	29
25.....	0	1	9	1	7	43

In this table the first column at the left gives the places assigned in an order of merit to the rugs. The remaining six columns show the number of persons who assigned a given rug to such places. For example, Rug 1, which had the highest average position in this series, was given first place by 24 persons, second place by 25 persons, third place by 24 persons, etc., whereas Rug 25, which had the lowest average position, was assigned first place by 1 person and last place by 43 persons. Rug 8 is illustrative of a small group which showed bi-modal distributions. There were a few cases in which the distributions were similar to those of pure chance.

The distributions of choices in Series Two were closely similar to those of Series One in the several types. In spite of the wide spread of these placements there is after all a preponderance of choices in favor of certain rugs.

Several of the subjects felt that their judgments were likely to vary largely from day to day. It was possible to secure a second trial of the two series from 38 persons, the time interval between the trials being not less than 3 weeks. In every case the subjects said that they had by that time forgotten practically all of their previous judgments.

Stability of Individual Preferences.—The measure of a person's agreement with his own previous preferences was obtained by correlating the rank order of the rugs in the two trials, by the formula

$$r = 1 - \frac{6\sum D^2}{n(n^2 - 1)}.$$

For Series One the coefficients ranged from $-.23$ to $+.94$, with a mean of $+.71$ and standard deviation 18.5. For Series Two the range was $-.17$ to $+.96$, mean $+.72$, S.D. 15.7. It is clear from these figures that in some cases a person's second arrangement of the pictures had no significant correlation with his first. But these cases are exceptional, and the greater number showed a fair degree of self consistency.

Agreement of Individuals with Group Judgment.—When 101 persons had performed the task of arranging the pictures their various estimates were combined in an average. For example a certain rug would be given first place by one person, third place by another, twelfth place by another, etc. These numbers were added, and the rug which had the smallest resulting number was assigned first place in the composite order of merit. The rug which had the largest resulting number was called last or twenty-fifth in the composite order. Comparisons were made between this composite order and the individual arrangements of each member of the group, and also of another 106 persons who were not members of the first group. The above quoted correlation formula was used. Counting the two series this made 414 coefficients which were computed between individuals and the group order. The range of these is very great,—from $-.12$ to $+.80$ for Series One, and from $-.32$ to $+.84$ for Series Two. The respective means are $+.42$, S.D. 19.6, and $+.41$, S.D. 22.5.

The individual experts had lower correlations with the group than did the non-experts. Their coefficients for Series One ranged from $-.10$ to $+.60$, mean $+.30$, and for Series Two from $-.32$ to $+.71$, mean $+.31$.

A few inter-correlations were computed between pairs of experts selected at random, 12 such coefficients for each series. These ranged on the first series from $-.26$ to $+.65$, with a mean of $+.20$, and on the second series from $-.29$ to $+.61$, with a mean of $+.19$. In other words, the experts agreed very little with the group, and hardly at all among themselves.

The question here suggests itself whether agreement with the group is a constant quality or characteristic in the individual. That is to ask whether the persons who agree most closely with the group in judging one series of rugs will be the ones to agree most closely with the group in judging the other series. This was answered as follows. The correlations of individuals with the group for Series One were plotted against the correlations of the same individuals with the same group on Series Two. [See Table 2.]

Since the correlation coefficients were exact numbers the Pearson r was used, and the result was $+.47$, p.e. $.036$.

Agreement of Group with Group.—Five groups of subjects, 20 in each group, were compared. Groups I., II., III., and IV. were mutually exclusive groups selected at random. Group V. comprised the 20 experts mentioned above. The averaged order of merit for the rugs for the 20 persons of Group I. was correlated with the averaged order for the 20 persons of Group II. and so on. In this way 10 group inter-correlations were secured for each series. The mean inter-correlations of the 4 non-expert groups is $+.76$ (range $+.61$ — $+.86$) for Series One and $+.84$ (range $+.79$ — $+.94$) for Series Two. The mean correlations of the expert group with the other groups is $+.55$ and $+.53$ respectively for the two series. Taking the five groups together and combining the series the mean intercorrelation is $+.70$.

Four mutually exclusive groups of 50 persons each were compared, the averaged order of rugs for each group being

TABLE 2

Correlation Coefficients for Series One

	-20	-10	0	10	20	30	40	50	60	70	80	90
90							1			3		
80							1	1	7	4	3	
70					3	3	4	4	4	3	2	
60				2	3	4	7	11	9	2		
50			1	2	5	11	8	8	2	1	1	
40			4	1	1	5	11	7	8	1		
30												
20	1			2	1	6	5	3	1	4		
10												
0												
-10			1		2	1		1				
-20												
-30			2									
-40												

Scatter Diagram of Correlation Coefficients. Agreement with Series One Plotted Against Agreement with Series Two. $n = 207, r = +.47, p.e. = \pm .036.$

correlated with the averaged order for the other groups. The intercorrelations between the groups were, for Series One $+ .81$, $+ .82$, $+ .83$, $+ .85$, $+ .93$, and $+ .96$, and for Series Two $+ .84$, $+ .87$, $+ .87$, $+ .90$, $+ .90$, and $+ .92$. The mean is $+ .87$ for one series and $+ .88$ for the other.

A comparison was made between the average arrangement of the first 30 men tested, and the first 30 women. The coefficients were $+ .86$ for Series One and $+ .86$ for Series Two.

Finally the group arrangement for the first 100 persons was compared with the group arrangement of the second 100 persons. The result is a coefficient of $+ .93$ for Series One, and of $+ .95$ for Series Two. It appears then that as the numbers of judges included in the groups are increased the higher becomes the correlation between groups. In the case of the present data, groups of twenty agree to the extent of $+ .70$, groups of fifty by $+ .87$, groups of one hundred by $+ .94$.

SUMMARY OF CONCLUSIONS

1. Every observation made with Series One was confirmed by the results of Series Two.

2. There was very wide diversity of individual judgments on the rugs, and this was even more marked in the case of experts than of others.

3. The majority of persons who judged the pictures twice were fairly consistent with themselves, mean coefficients were $+ .71$ and $+ .72$.

4. The correlation of individuals with groups tended to be positive, though not high, mean coefficients being $+ .41$ and $+ .42$.

5. Those who tended to reflect group judgment in the one series tended also to do it in the other series. Pearson $r = + .47$.

6. No large negative correlations were obtained.

7. Between group and group the measure of agreement is large and it increases with the size of the groups.

8. The experts who, as individuals, had mean correlations of $+ .19$ and $+ .20$ with a large group, had, as a group, mean correlations of $+ .53$, $+ .55$ with other groups.

9. It may be held that although the large groups agree on a certain order of merit for these rugs it may yet be a "wrong" order. I shall be willing to believe this when it can be shown that experts themselves, judging under experimental conditions, can achieve so high a degree of stability as the large groups here show.

UNIVERSITY OF CALIFORNIA, SOUTHERN BRANCH,
LOS ANGELES