

Heritability of Factors in Creative Thinking and Esthetic Judgment

Frank Barron

Although the heritability of general intelligence has been the subject of much study and can be considered to be well established, the heritability of specific factors in intellectual functioning has not as yet been the object of such intense scientific inquiry.

The recent work of Nichols (1965), utilizing a large sample of twins from the National Merit Scholarship testing program and employing a battery of tests measuring achievement in five different academic areas, is an excellent pioneering study. Nichols used a composite score derived from the five subtests, presuming this to be a measure of general ability, and he found that MZ twins were much more alike than DZ twins with respect to the composite score; about 70% of the variance in the composite could be attributed to heredity. The novel result of his investigation is his discovery that the specific subject areas also have significant heritability components. When the influence of the composite general ability upon subtests was removed statistically, the residual subtest scores also showed considerable heritability.

A limitation of Nichols' findings is that the subtests do not offer any potential theoretical link to basic factors in mental functioning. Such measures, however, may be provided by the fundamental factor analytic explorations of Guilford and his associates during the past twenty years. Since our own interest has been in the area of creative thinking and esthetic judgment, we have restricted ourselves in the present investigation to the use of tests in those areas whose factorial classification has been studied previously.

Tests for the following factors were employed: (1) Adaptive Flexibility; (2) Expressional Fluency; (3) Ideational Fluency; (4) Originality; (5) Esthetic Judgment for Visual Displays.

A word about this latter factor-test is in order. The factor was identified in a study done jointly by the present author and the Guilford group (Barron et al, 1957). However, the authors were of two minds about its interpretation. The highest loading of the factor was on the Barron-Welsh Art Scale, a set of geometrical figures and drawings whose preference-rank for professional artists had been established independently and whose scoring expresses the degree of agreement of the respondent with professional artistic judgment. The standardization would thus justify the factor name given above. A complication stems from the fact that virtually all the figures ranked high by artists are of greater complexity than those they rank low, and it is possible to interpret the measure as basi-

Tab. I

Variable	Italian sample						American sample					
	r_{MZ}	r_{DZ}	r^2_{MZ}	r^2_{DZ}	h^2	HR	r_{MZ}	r_{DZ}	r^2_{MZ}	r^2_{DZ}	h^2	HR
Gottschaldt Figures	0.86	0.35	0.74	0.12	0.78	1.186	0.49	0.13	0.24	0.02	0.41	1.469
Expressional Fluency	0.92	0.70	0.85	0.49	0.75	0.485	0.28	0.76	0.08	0.58	-4.33	7.429
Consequences, fluency score	0.38	0.71	0.15	0.51	-1.14	-1.708	0.45	0.10	0.20	0.01	0.39	1.556
Consequences, originality score	0.72	0.75	0.52	0.57	-0.13	-0.086	0.42	0.49	0.18	0.24	-0.14	-0.333
Symbol Equivalence, fluency	0.42	0.73	0.17	0.53	-1.17	-1.506	0.29	0.36	0.08	0.13	-0.11	-0.483
Symbol Equivalence, originality	0.15	0.58	0.02	0.34	-1.04	-5.634	0.22	0.35	0.05	0.12	-0.20	-1.182
Unusual Uses, fluency	0.72	0.59	0.52	0.35	0.31	0.348	0.42	0.45	0.18	0.20	-0.06	-0.143
Unusual Uses, originality	0.59	0.47	0.35	0.22	0.23	0.411	0.38	0.18	0.14	0.03	0.24	1.053
Barron-Welsh Art Scale	0.58	0.07	0.34	0.01	0.55	1.750	0.66	-0.02	0.44	0.00	0.67	2.061

cally a preference for complexity, as the present author has done (Barron, 1953). The factor might therefore be understood simply as “preference for complex visual displays”.

As we shall see, this may be an important point for the interpretation of the findings to be reported. Our measure for the Adaptive Flexibility factor is the Crutchfield revision of the Gottschaldt Figures, in which the task is to discover in a complex figure a simpler embedded one. This too was taken as the test of choice, because of its high loading on Adaptive Flexibility in the earlier factor analysis (Barron et al, 1957).

The present investigation employed, in addition to the Art Scale and the Gottschaldt Figures, the following tests from the Guilford battery: Expressional Fluency, Unusual Uses, Consequences, and Symbol Equivalence. These latter three tests are scored for Ideational Fluency (low quality score, or a simple count of the number of acceptable but not highly original or clever responses) and for Originality (number of responses scored 4 or 5 on a 5-point scale of originality).

Two samples of adolescent twins were employed: one group of Italian twins (whose cooperation was secured through the good offices of the Mendel Institute in Rome and the Institute for Medical Genetics in Florence), and a second group of American twins, studied at the Institute of Personality Assessment and Research, University of California, Berkeley. In the Italian sample, diagnoses were arrived at from medical data, including blood tests, and it is safe to assume they are highly accurate. The American sample was classified as to zygosity on the basis of their responses to the questionnaire developed by Nichols (1965). While the method can be expected to result in some misclassifications, Nichols reported 93% accuracy against the blood grouping criterion. Error in diagnosis should lead to a higher intraclass correlation for DZ twins and a lower one for MZ twins for inherited traits, so that heritability will be underestimated, rather than overestimated, on the average, when errors of classification do occur.

The Italian sample consisted of 59 pairs of like-sexed twins: 30 MZ (15 ♂♂ and 15 ♀♀) and 29 DZ (14 ♂♂ and 15 ♀♀). One ♀♀ DZ pair had to be eliminated later, because of incomplete test results. The average age of the sample was 17.04 years; all subjects were secondary school students or graduates. The American sample consisted of 57 pairs of like-sexed twins: 29 MZ (14 ♂♂ and 15 ♀♀) and 28 DZ (13 ♂♂ and 15 ♀♀). The average age of the sample was 17.42 years, and all were either recent high school graduates or seniors.

Results will be reported first for the Italian sample and then for the American one, as this was the actual order of work; and a preliminary partial report on the Italian sample has already been made (Barron, 1969).

Adaptive Flexibility. As we have indicated, the form of the Gottschaldt Figures employed as a measure of this factor is the Crutchfield revision (MacKinnon et al, 1956). The test is well adapted for group administration and is closely timed. The task is to discover in a complex figure a specific simpler figure embedded in it. Fifteen problems are presented.

Scores, in the Italian sample, ranged from 0 to 15, with a mean of 7.16. The intraclass correlation for the MZ group is 0.86; for the DZ group, 0.35. Both correlations are significantly different from zero, and they are significantly different from one another as well ($t = 4.79$; $P < 0.01$).

The Holzinger heritability coefficient h^2 is 0.79 (correlations not corrected for attenuation).

In the American sample, the mean score is 7.81, not significantly different from the Italian mean. The MZ intraclass correlation is 0.49; the DZ one, 0.13, giving $h^2 = 0.41$. The two correlations are significantly different from one another, as in the Italian sample, and the indications are for a significant degree of heritability in the intellectual function measured by this test.

Expressional Fluency. This is a test from the Guilford battery, in which the subject is given the initial letters of four words in sequence and asked to complete the words in such a manner as to form a sentence. His score is the number of complete sentences he can form in a given time.

In the Italian sample, the MZ correlation is 0.92, the DZ correlation 0.70, yielding $h^2 = 0.75$. In the American sample, however, the finding does not hold. The MZ correlation is -0.28 , the DZ 0.76, yielding $h^2 = -4.33$.

Ideational Fluency. Here again, the findings are inconsistent and heritability is not indicated. Low quality scores for Consequences, Symbol Equivalence, and Unusual Uses for the two samples show the following results:

	Italian sample			American sample		
	MZ	DZ	h^2	MZ	DZ	h^2
Consequences	0.38	0.71	-1.14	0.45	0.10	0.39
Symbol Equivalence	0.42	0.73	-1.17	0.29	0.36	-0.11
Unusual Uses	0.72	0.59	0.31	0.42	0.45	-0.06

Originality. The results for the “high quality” scoring of these tests are as follows:

	Italian sample			American sample		
	MZ	DZ	h^2	MZ	DZ	h^2
Consequences	0.72	0.75	—0.13	0.42	0.49	—0.14
Symbol Equivalence	0.15	0.58	—1.04	0.22	0.35	—0.20
Unusual Uses	0.59	0.47	0.23	0.38	0.18	0.24

At this point, it should be noted that these tests of expressive verbal behavior, independent of verbal comprehension and reasoning, are not scored objectively, as the other tests in the battery are, but rather by subjective ratings. Their reliabilities are correspondingly lower, revealing interrater agreement in the range of 0.40 — 0.70. Inconsistencies in the findings may thus be in part a function of scoring error.

Esthetic Judgment of Visual Displays. In the Italian sample, the MZ intraclass correlation is 0.58; the DZ one, 0.07. The MZ correlation is significantly different from zero, but the DZ one is not. They are significantly different from one another ($t = 3.05$, $P < 0.01$), and a high heritability component is indicated by the h^2 value of 0.55.

This finding is strongly confirmed in the American sample. The MZ intraclass correlation is 0.66, the DZ one, —0.02, a highly significant difference. The corresponding h^2 value is 0.67.

Leaving aside for the moment the hypothesized factorial structure for which some of these tests are presumably measures, the most obvious feature of the results is that the tests calling for expressive verbal behavior show, for the most part, substantial similarity of all twin pairs, MZ and DZ alike. The tests calling for visual discrimination of complex forms, as well as reasoning and judgment in the visual sphere, reveal significant differences between MZ and DZ twins, with the intraclass correlations being much higher in the MZ samples.

Since expressive verbal behavior is much more likely to be influenced both by imitation and propinquity than are purely visual discrimination and reasoning — which are largely silent and unobserved because of their infrequent practical consequences and a lack of social need for their expression — it does make sense that simply being part of the environment called twinship would produce high correlations in the first set of behaviors, while leaving the second unaffected by the social environment. Thus, the findings create a strong argument for the existence of a significant hereditary component in visual judgment. The possible significance of this for esthetics needs clarification, however, since the Gottschaldt Figures, on the face of it, call for discrimination and reasoning but not esthetic judgment; while the Barron-Welsh Art Scale, although known to be relevant to esthetic judgment and creativity, has not been studied from the point of view of the respondent's capacity to discriminate form and to reason about patterns, but only from the point

of view of the correlates of his *preferences*. The question arises: can one exhibit good taste in preferences without having the capacity to discriminate and reason about the stimulus material in which the problem of choice is presented? Could these considerations play a part in the analysis of creative mathematical reasoning, which traditionally has employed esthetic criteria in judging of the fitness of a solution, or in the degree of promise of steps towards a solution? Certainly, further study of these effects and their implications is needed, both in new studies with the twin method and with expanded test measures in the figural sphere, as well as with individuals of noteworthy ability in the arts and in mathematics.

Acknowledgments. We are especially indebted to Drs. Luigi Gedda and Paolo Parisi, of the Mendel Institute in Rome, and to Dr. H. B. Young, of the Harvard Florence Project. Other members of the Mendel Institute staff, including especially Dr. Lucio Braconi, also contributed time and effort to the research in a spirit of generosity that we found most memorable. The work in Florence also owes much to Dr. A. Noferi, who interviewed the subjects and administered the tests. The Institute for Medical Genetics in Florence and the Institute for Applied Psychology there also contributed their facilities and technical help.

The research in Berkeley was accomplished with the valuable help of Drs. Arnold Mordkoff and Ann-Marie Allerand Bloch, the interviewing and test scoring assistance of Barbara Bell, Lorraine Granit, Jerrell Kraus, and Dr. Isabella Conti, and the statistical advice and labors of Dr. Wallace Hall and Susan Hopkins.

The research was initiated with the financial help and support of the Richardson Foundation, and parts of it were made possible by a grant from the US Office of Education for basic research in esthetics. The latter aspect of the work is continuing with valuable support from the Arts and Humanities Division of that agency.

References

- BARRON F. (1953). Complexity-simplicity as a personality dimension. *Abnorm. Psychol.*, **48**: 163-172.
- GUILFORD J. P., CHRISTENSEN P. R., BERGER R. M., KETTNER N. W. (1957). Interrelations of various measures of creative traits. Technical Memorandum AF 18 (600)-8. Institute of Personality Assessment and Research, Berkeley, California.
- MACKINNON D. W., BARRON F., BLOCK J., CRUTCHFIELD R. S., GOUGH H. G. (1956). An assessment study of Air Force Officers. Technical Report WADC-TR- 58-91, ASTIA Document No. AD 151040.
- NICHOLS R. C. (1965). The National Merit Twin Study. In S. G. Vandenberg: *Methods and Goals in Human Behavior Genetics*. Academic Press, New York.

Prof. F. BARRON, Institute of Personality Assessment and Research, University of California, 2240 Piedmont Avenue, Berkeley, Calif. 94720, USA.