of toys, heredity, inadequacy of nutrition and early childhood practises of abrupt weaning and neglect following weaning.

Numerous statements on personality differences between Africans and Europeans can be found throughout the literature (cf. Carothers, 1953). Biesheuvel (1959), for example, has suggested that Africans have a somewhat lower constitutional tempo and activity level than Europeans; and that they show, in terms of Heymans & Wierma's (cf.Biesheuvel & Pitt, 1955) dimension of primary-secondary function, more tendency to primary function than Europeans. Assuming that some basic personality differences do exist these could also affect the manifestation of curiosity in the two cultural groups.

Nutrition is another factor which may play an important part in the development of curiosity. That malnutrition can have serious effects on growth, health and energy and hence is bound to affect human efficiency is obvious (Cravioto, 1968). It has been shown that chron c protein deficiency affects growth, learning ability and temperamental stability in the rat (Cowley & Griesel, 1964, 1965) and that kwashiorkor a protein deficiency disease common in African children, retards the development of brain rhythms in early childhood (Nelson, 1959). Diets of both tribal and urban African children are often far below what is necessary for optimum nutrition.

5. Summary and Conclusions

Some theoretical assumptions and methodological difficulties underlying cross-cultural research are outlined in this chapter. It is assumed, in this study that attributes of

behaviour are a product of both heredity and environmental influences, and it was shown that from this point of view there can be no such thing as a culture-free test since the abilities of man are themselves not culture free. Some difficulties in the concept of the culture-fair test are also discussed.

In this study, it is held that in principle a comparison in performance between different cultures, which goes beyond a listing of differences in repertoire, can be made. However rigorous requirements as to the type of data that is acceptable should be laid down. Discussion of some of the requirements suggested by various authors reveals that while these are so strict as to be almost unattainable, they still leave a number of loopholes and unanswered questions. Nevertheless, the requirements for comparability put forward in this chapter can be regarded as of considerable value in providing guidelines for the analysis and interpretation of cross-cultural data.

Some reasons why a cross-cultural study of curiosity is regarded as important (implications for education, job selection; to test the universality of the construct), and some factors which may differentially effect the development of curiosity in Africans are mentioned.

CHAPTER III

THE CONSTRUCTION AND ADMINISTRATION OF THE CURIOSITY TESTS

1. The Aims of the Study

Two main aims can be specified in the construction of the curiosity tests.

(i) To determine whether the collative stimulus properties, complexity, novelty, incongruity and uncertainty regarded as the prime determinants of curiosity by Berlyne (1960) and used individually in a variety of studies as measures of curiosity, do in fact tap the same construct, viz. curiosity.

(ii) To study curiosity be means of its determinants in a group of African and European subjects and to discover whether the determinants of curiosity operate in a similar manner in a Western and African group.

2. The Subjects

The subjects were 97 male, South African undergraduate psychology students, i.e. 50 Africans from the University of Fort Hare and 47 students from the University of the Witwatersrand. The mean age of the European sample was 20,1 years, that of the Africans, 21,7 years. English was the medium of test instruction for all subjects. Of the African subjects 40 were Xhosa, 5 Sotho, 3 Zulu, and 2 Fingo. All were fluent in English which was the

teaching language at the University. The home language of 44 of the European Subjects was English, 2 Italian and 1 Afrikaans. Since the curiosity tests involved little or no verbalization and the test instructions were extremely simple, it is assumed that linguistic differences had no effect on test scores.

All subjects were volunteers who were recruited during one of their psychology classes. In both groups the percentage of subjects who volunteered to participate in the experiment was over 90% of those approached. It is thus felt that there was no biased sampling effect in favour of either group. Subjects were paid a nominal fee of R 2,50 on completion of the testing program which lasted approximately 6 hours.

Subjects with a history of illness likely to affect vision or psychomotor skills were excluded. All subjects were tested for colour-blindness with the test of Ishihara (1958). No African subjects were found to be colour-blind but 4 Europeans were red-green colour-blind. In spite of reservations about the use of colour-blind subjects they were included in the sample because of a shortage of European subjects. Since no significant difference in responses on the novelty test, the only test involving colour vision, between these and other European subjects was found, it was felt that the colour-blindness of these subjects did not greatly affect test results.

University students, compared with the rest of the general African and European population are, of course, a

highly selected group in terms of intelligence and motivation. This is probably even more the case for the African than for the European sample. However, it was decided to use students as subjects for the following reasons:

(i) Psychology students form a fairly homogeneous group compared to the general population, thus making for greater comparability of the two samples.

(ii) The familiarity of the African university student with the European culture made it easier to overcome specific cultural effects on test performance.

3. Conditions of Testing

An attempt was made to construct tests for the 4 principal determinants of curiosity, viz. complexity, novelty, incongruity and uncertainty. A test measuring conflict reaction time explicitly, was also included in the battery in order to test whether there was a relationship between curiosity and a direct measure of conflict. A Curiosity Questionnaire and a Self-Rating Scale, in which subjects were asked to rate themselves on adjectives pertaining to curiosity, were constructed to give an indication of the face validity of the above-mentioned tests. Some intelligence measures were also included to provide a means of testing the discriminant validity of the measures (cf. Campbell & Fiske, 1959) and their functional equivalence in the two samples (cf. Poortinga, 1971,p.79). The African sample, in addition, filled in a biographical questionnaire based on Grant's (1969) Urban-Rural Scale.

Each subject was tested individually on 2 consecutive days with equivalent versions of the tests devised to measure Conflict Reaction Time, Uncertainty, Complexity and Novelty, in an attempt to eliminate non-experimental effects due to differential familiarity with the testing situation and apparatus in the two cultural samples. (cf. Pcortinga 1971 p.8) Tests were administered in the order in which they are mentioned above. The test for Incongruity was administered on the second day of testing only, immediately after the test of Uncertainty. In the 3rd testing session, paper and pencil tests, consisting of the Curiosity Question are, the Self-Rating Scale, the Biographical Questionnaire and the Intelligence Measures were administered to the subjects in small groups. Each of the 3 testing sessions lasted about 2 hours.

4. The Conflict Reaction Time Test

Apparatus A photograph of the apparatus is shown in Figure 1. The apparatus consists essentially of an eyepiece jutting out of a black box which contains a round opaque glass window, 10 cm in diameter and approximately 25 cm from the eyepiece, which could be illuminated by means of a light source situated behind the box. This light could be varied in intensity. There were two buttons for terminating the light stimulus, one on the left and one on the right.

Stimulus Material Subjects were tested under 4 conditions on each of the two individual testing sessions.

(a) Simple Reaction Time, in which the subject was required to react as quickly as possible to a light of constant brightness intensity. (Luminance value = 547.3 lamberts).

(b) Conflict Reaction Time, Level 1, in which the subject was required to discriminate between 2 lights of different brightness intensity. (Luminance values = 171,8 and 855,9 lamberts respectively).

(c) Conflict Reaction Time, Level 2, in which the subject was required to discriminate between 2 lights whose brightness intensities differed less from each other

Figure 1. Apparatus for the Conflict Reaction Time Test.

than the previous two. (Luminance values : 171,8 and 547,3 lamberts)

(d) Conflict Reaction Time. Level 3 in which the subject was required to discriminate between 2 lights whose brightness intensities differed even less from each other than the previous two. (Luminance values : 171,8 and 299,1 lamberts)

The Testing Procedure. (a) Simple Reaction Time: On entering the testing situation the subject was given the following instructions:

> "Place your eyes in front of the eye-piece... you will see a light coming on. It is your task to press the button on the right as soon as you see the light coming on. Please be as quick as you can."

A few practise trials were given and then testing was begun. The subject was given 20 trials with his right hand and then told to press the button on his left with his left hand for a further 20 trials. Reaction Time (RT) was recorded in hundredth's of a second.

(b) Conflict Reaction Time Level 1. The subject was then required to discriminate between the two lights of different brightness intensities. He was instructed as follows:-

> Place your eyes in front of the eye-piece I am now going to show you 2 lights of different intensities....This is the light of lower intensity. When you see it you should press the button on your left .... This is the light of higher intensity, when you see it you should press the button on your right. Please be as quick as you can. If you should make a mistake and press the wrong button, you will hear a loud buzzer".

After a few practise trials, the subject was given 40 test

trials. RTs and errors were recorded.

(c) Level 2. The subject was instructed that he would be required to discriminate between two lights which differed less from each other in intensity than the previous two. He was shown both lights and then given 40 test trials.

(d) Level 3. This was followed by 40 trials with two lights which differed even less from each other in intensity than the lights shown in level 2. RTs and errors were recorded.

5. The Uncertainty Test

The Stimulus Material. Series of situations in which hands manipulated common objects (e.g. a screwdriver, scissors) were photographed, firstly in focus and then in varying degrees of "out of focus" or "blurredness". It was found necessary to vary the degree of blurredness with the size and structure of the object being photographed. Of these pictures, 40 photographs, 20 in focus and 20 showing in the opinion of the author, an intermediate degree of blurredness, were selected as the stimulus material.

The Testing Procedure On each of the two testing sessions 20 slides (10 in focus and 10 blurred) were projected onto the screen, one at a time, in a random order. On the first testing day the test slides were preceeded by a picture of the experimenter and 2 demonstration slides (1 in focus and 1 showing an intermediate degree of blurredness).

Instructions were as follows:

"This is a picture of the experimenter; as you can see he is pouring water into a glass. The pictures which follow show only the

experimenter's hands; they are doing something different in each picture. Some of the pictures are clear and some are not." (The next two demonstration slides were shown) "You may look at each picture for as long or as short as you wish. When you have finished looking press the button in your right hand; the picture will disappear and a new one will appear in a couple of seconds. Are these instructions clear to you?"

Instructions on the second day of testing were essentially similar except that no demonstration slides were shown.

The time (in hundredth's of a second) that the subject looked at each picture was recorded by means of a Heath Digital Timing Module (EU 801) and timer attached to a shuttor in front of the projector. The interstimulus interval of 5 seconds was also controlled by the same apparatus.

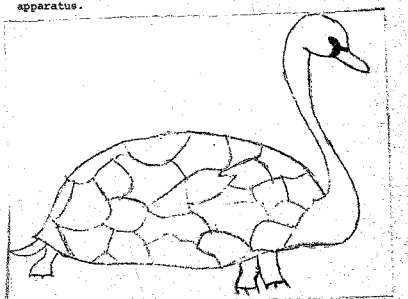

Figure 2. Incongruity Test: An example of the stimulus material.

6. The Incongruity Test

The Stimulus Material A series of 12 simple pictures of animals were drawn. These are referred to as the "congruous pictures". Six further incongruous pictures were drawn of mismatching combinations of heads and bodies of the congruous animals. An example is shown in Figure 2. These 18 drawings were then photographed.

The Testing Procedure The 18 slides, consisting of the 12 congruous drawings presented in a random order followed by the 6 incongruous pictures also in random order were projected onto the screen, one at a time. The subject was instructed that he would be shown some simple drawings of animals and that he could look at each picture for as long or as short as he wished. When he had finished looking he should press the button in his right hand in order to proceed to the next drawing. The time that the subject looked at each picture and the interstimulus interval were controlled by the same apparatus as in the uncertainty test.

7. The Complexity Tests

It was decided to divide this test in two parts, the first part consisting of non-representational items, having presumably a low cultural load, the second of representational items having a higher cultural load.

A. The Non-Representational Complexity Test

The Stimulus Material. A computer program was

written for the Hewlett-Packard Desk Calculator (9100B) and Plotter (9125A) by means of which line segments (½ cm in length and differing randomly in slope and direction) generated in one 45 degree segment of a circle could be automatically plotted in that segment and in the other 7 segments of the circle fo give bilaterally symmetrical patterns. Complexity of the patterns was defined in terms of the number of line segments contained in the 1st segment of the circle. (Complexity Level 1 contained 2 lines in the first segment of the circle, Level 2 - 3 lines, level 3 - 4 lines, level 4 - 6 lines, Level 5 - 8 lines, Level 6 - 12 lines, Level 7 - 16 lines and Level 8 - 24 lines in the first sigment of the circle. In terms of information theory these numbers of lines correspond respectively to approximately 1,1½,2,2½,3, 3½,4 and 4½,bits. By randomly varying the slope and position of the lines a large number of different patterns could be generated at each level of complexity.

A large number of different random patterns at each level of complexity were generated in this way. Patterns were then paired together in such a way that for each of the two testing days 2 different patterns at each level of complexity were combined twice with patterns on each of the three adjacent levels of complexity. No particular pattern was used in more than one combination. Thus the following combinations were made twice for each of the two testing days.

Level 1	with	Level 2	
"	"	"	3
"	"	"	4
" 2	"	"	3
"	"	"	4
"	"	"	5
" 3	"	"	4
"	"	"	5
"	"	"	6
" 4	"	"	5
"	"	"	6
"	"	"	7
" 5	"	"	6
"	"	"	7
"	"	"	8
" 6	"	"	7
"	"	"	8
" 7	"	"	8

These pairs of patterns were then photographed, left and right positions of the more complex patterr of a pair on the slides being alternated randomly. An example of the stimulus material is shown in Figure 3. These patterns were then also photographed singly.

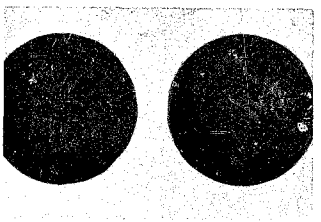

Figure 3. Non-Representational Complexity Test: An example of the stimulus material.

The Testing Procedure On each of the 2 testing sessions 36 pairs of patterns were projected, one pair at a time onto the screen for 3 seconds. The subject was asked to indicate by pressing the button on his right or his left whether he wished to see the left or right pattern again for a further 3 seconds. His choice was recorded.

B. The Representational Complexity Test

The Stimulus Material Series of slides of different forms of art (Bushman art, African art, Graphic art, Islamic and Oriental art, Abstract Modern art and Child art) were photographed. Two different sets of ratings of the complexity of the pictures were then made. The first

rating was used in the selection of items for the test, the second to facilitate the analysis of the results and provide some means of determining an overal complexity preference level for the African and European samples (cf.Chapter IV 1(b)).

The 1st Rating Four African and 4 European male judges (Staff members of the National Institute for Personnel Research) rated the pictures within each series separately for complexity on an 8-point scale (1 referring to the lowest level of complexity, 8 the highest level of complexity in the series). The procedure was as follows:

The judges were first shown each picture in a particular series for 3 seconds with a two second interval between each presentation. After an entire series had been viewed in this way, each slide was then again projected for a further 3 seconds followed by a 5 second interval during which each judge was required to independently write down his rating of the picture's complexity on an 8-point scale with respect to the other pictures in the series. The same procedure was carried out for each of the 6 series of different forms of art. The ratings of the European judges on the pictures correlated highly with each other (ranging from 0,7 to 0,95) for all the series. The ratings of the African judges, however, yielded low correlations both with each other and the European judges. It was therefore decided to exclude their ratings in the selection of the stimulus material and to use only

the ratings of the European judges.

By pairing slides within the same series but differing in complexity level, 35 pairs of test slides were obtained for each of the two testing days.

The 2nd Rating Four different African and European judges, again staff members of the National Institute for Personnel Research, then re-rated the 70 slides selected to be used on the 1st testing day with respect to each other and the 70 slides selected to be used on the 2nd testing session with respect to each other on 2 consecutive days on an 8-point scale of complexity.

On each day the judges were first shown each slide for a period of 3 seconds, followed by a 2 second interval. Each slide was then shown for a further 3 seconds followed by a 5 second interval during which each judge was required to independently write down his rating of the picture's complexity. The ratings of the European judges, again correlated highly with each other (0.76 to 0.84) but 2 of the African judges showed very low correlations with all the other judges; their ratings were therefore excluded. The correlations of the 2 remaining African judges ranged from 0,5 to 0,63. The complexity rating for each slide with respect to all the other slides shown on the same day of testing was thus based on the mean ratings of the 4 European judges and the 2 Africans.

The Testing Procedure On each of the 2 testing sessions, 35 pairs of pictures were projected one pair at a time onto the screen, for a period of 3 seconds. The subject

was required to indicate after each presentation whether he wished to see the picture on the right or on the left again by pressing the button on his right or his left. His choice was recorded.

8. The Novelty Test

The Stimulus Material. Short series of 5 to 15 slides, depicting the same type of object (e.g. a series of waterlilies) were photographed in colour. Each series also included a picture of a different but related object (e.g. in the middle of a series of lilies a protea would be inserted). For each of the two testing sessions 4 different series of slides (flowers, landscapes, 1 species of bird and 1 species of animal), each containing a novel but related item, were prepared.

The Testing Procedure On each of the 2 testing days, the slides were projected one at a time onto the screen. The subject was instructed to look at each picture for as long or as short as he wished and to press the button in his right hand when he wished to proceed to the next slide. There was an interstimulus interval of 5 seconds between one picture and the next. The time that the subject looked at each picture was recorded and the interstimulus interval was controlled by the same apparatus as in the Uncertainty test.

9. The Curiosity Questionnaire

The Construction of the Questionnaire Staff members of the National Institute for Personnel Research were asked to submit possible items for a curiosity questionnaire.

Following this a preliminary version of the questionnaire, which also included a number of suitable items adapted from the Sensation Seeking Scale (Zuckerman, Kolin, Price & Zoob, 1964) was drawn up.

The preliminary version consisted of 66 pairs of statements, one statement of a pair supposedly indicative of high curiosity, its opposite of low curiosity. An attempt was made to make both alternatives equally desirable by manipulating the item wording. One statement of a pair was presented on the left-hand side of the page, the other on the right, the 2 statements being separated by 4 dashes. The subjects were required to mark the left-most dash if they agreed strongly with the statement on the left and the right-most dash if they agreed strongly with the statement on the right. The two intermediate dashes indicated a weaker preference for the left or the right-hand statements.

This preliminary version of the questionnaire was administered to 227 undergraduate African students (155 males and 72 females) from the University College of the North and 310 students (33 males and 277 females) from the Johannesburg College of Education. An item analysis with iteration procedure using Gulliksen's index (Gulliksen, 1950, p.363) (criterion r=0,15,step=0,03) was carried out for the African Male, African female, European male, European female and the Combined groups. With 2 iterations KR 20 values for the 4 groups were 0,83, 0,82, 0,85, 0,82 respectively. It was also found that roughly the same items were discarded in all groups with this procedure.

These items were left out in the final version of the questionnaire which then contained 48 items. A copy of the final version of the questionnaire can be found in Appendix 1.

The Testing Procedure The questionnaire was administered to small groups of approximately 10 subjects. Subjects were told to read the instructions carefully and then proceed to fill in the questionnaire. There was no time limit.

10. The Self Rating Scale

A 29-item self rating scale on which subjects were required to rate their conception of "Myself" on a 7-point scale was also constructed. The end points of the scale were anchored by means of contrasting adjectives pertaining mainly to curiosity and activity. A copy of the self-rating scale can be found in Appendix 1. The testing procedure was the same as that for the Curiosity Questionnaire.

11. The Intelligence Tests

The Intelligence tests were included in the test battery, primarily to provide some means of testing the discriminant validity of the curiosity measures and to investigate the functional equivalence of the curiosity tests in the two samples.

Adopting a multifactor approach to the measurement of intelligence, paper and pencil tests, measuring four of Thurstone's six primary mental abilities (viz. numerical ability (N), verbal meaning (V), spatial ability (S) and reasoning ability (R)) were selected. Of these tests, the

test for spatial ability was considered especially important. A number of studies have commented upon the inferior performance of Africans on tests involving spatial ability (Grant & Schepers, 1969; Hudson, 1967; Rollings, 1961). Ability to perceive spatial relations could have an effect on performance on all the curiosity tests since they all involved visual perception. The following tests were administered to the subjects in small groups in the third testing session.

A. The Mental Alertness Test (A/1/1) from the Intermediate Battery (National Institute for Personnel Research, 1964). The 65 items in the test measure verbal and arithmetical reasoning ability and are intended for subjects with more than 12 years of education. (Wilcocks, 1969).

B. The Blox Test (National Institute for Personnel Research, 1961) This test contains 45 items consisting of line drawings of a set of cubes. The correct answer to each question is found by comparing the particular set of cubes with 5 other sets and picking out from the other 5 sets, the set which represents the same arrangement of blocks but viewed from a different angle. The test supposedly measures spatial ability (cf. Verster,1973).

C. The Letterset Test (Part 1.)(Educational Testing Services, 1962). Each of the 15 items consists of 5 groups of letters with 4 letters in each group. Four of the groups of letters are alike in some way. The correct answer is found by finding the rule that makes the 4 groups alike and drawing an X through the group

of letters that is different. The test supposedly measures inductive reasoning. (French et al., 1963; Verster, 1973). The subjects were first given 7 minutes to do this test with a pencil of one colour and were then told to carry on with a pencil of a different colour for a further 7 minutes. Two scores were thus obtained for each subject - after 7 minutes (speed test) and 14 minutes (power test). This was done because a number of investigations in the cross-cultural field (cf. Poortinga, 1971) have found that Africans are unduly penalized in speed tests.

D. The Locations Test (Part 1) (Educational Testing Services, 1962). Each of the 14 problems in this test consists of 5 rows of small dashes, separated into groups of dashes by blank spaces. In each of the first 4 rows one dash is replaced by an "x". In the 5th row 5 of the dashes are replaced by numbers. In each problem there is a rule guiding the placement of the "x" in each of the first 4 rows. The correct answer is obtained by finding the rule and then applying this rule to the 5th row and marking a number which appears where an "x" belongs. This test also measures inductive reasoning. (French et al., 1963).

As in the previous test, subjects were given 6 minutes to do the test with a pencil of one colour and then another 6 minutes to complete the test with a pencil of a different colour.

All scores on the intelligence tests were corrected for

guessing according to the formula : Score = C-I/(n-1), where C is the number of correct answers, I is the number of incorrect answers and n is the number of response alternatives for each item.

12. The Biographical Questionnaire

The African subjects were given a modified version Grant's Urban- Rural Scale (1969). This scale provides a measure of urbanization which has been shown by Grant (1969, Ch. VII) to influence performance on a number of cognitive tests. Some questions in the original scale were excluded and others modified to make the scale more suitable for students by Poortinga (1971, p.72). Questions in the modified version of the schedule related to residential history (rural v. urban), residential history of parents, education of parents, profession of parents, adherence to tribal life and customs, and outlook on life (egocentric v. community oriented). Each of the six groups of questions was scored separately from 0 (less urbanized) to 1 (more urbanized) and the total score was obtained by adding together scores on the six groups of questions for each subject. A copy of the Biographical Questionnaire can be found in Appendix 1.

CHAPTER IV

RESULTS AND DISCUSSION OF THE INDIVIDUAL TESTS.

1. The Complexity Tests

(a) The Non-Representational Complexity Test

The number of times each subject chose the more complex of 2 stimuli, on each of the 2 testing days was calculated and taken as his individual score on that particular day. Test reliabilities for the two samples were calculated by using the Kuder-Richardson Formula 20 (Lord & Novick,1968,p.91) and yielded values of 0,93 and 0,82 for the African and European samples respectively.

Table 1 shows the means, standard deviations and split half reliabilities of the "number of more complex choices" made by the 2 samples on the two testing days. The results of an analysis of variance (nested design, c.f. Brownlee, 1965,p.538) carried out on the data are presented in Table 2.

Table 1

Non-Representational Complexity Test: Means, S.D.s, and Reliabilities of Scores.

	Mean	S.D	Reliability
Africans Day 1	19,660	6,153	0,76
Day 2	19,480	8,667	0,88
Europeans Day 1	23,766	5,061	0,75
Day 2	24,213	5,056	0,71

Table 2

Non-Representational Complexity Test: Analysis of Variance with Culture (2x), Subjects, within culture (47x) and Day of testing (2x) as factors

Source of Variation	Denominator	Mean Square	df num.	df denom.	F
Culture, C	S(C)	846,814	1	92	6,122**
Subjects within culture, S(C)		138,320	92		
Day, D	D x S(C)	2,346	1	92	,010
C x D	D x S(C)	10,771	1	92	,458
D x S(C)		23,508	92		

In this and all subsequent Tables * → $p < .05$; ** → $p < .01$.

No significant differences between the first and second day of testing were found in either sample, but the results do show that the European sample preferred the more complex patterns on significantly more occasions than did the African group.

The Item Equivalence of the test for the two cultural samples: Originally, it was proposed to use a paired comparison method (incomplete design. cf. Bock and Jones, 1968, Ch.6) to obtain the preferred level of complexity in each sample. However, one difficulty with this method is that the results for the 2 cultural groups cannot be compared graphically, since for each sample, scale and location (based on the standard deviation and mean for the sample, respectively) are set differently. It was therefore decided to plot the % of preferred choices made at each level of complexity for each group. While this method is not as sophisticated as the paired-comparisons method, the form of the curve that is obtained is almost identical to that obtained with the paired comparison method, and this second method has the advantage that the graphs of the 2 cultural samples are directly comparable. The results for this method for the 2 samples are presented graphically in Figure 4. It can be seen that the graph of the African sample drops sharply after Complexity Level 6, this apparently being their level of complexity preference, while the curve for the European sample shows no inversion and continues to rise; as it stands, therefore, the test does not seem to fulfill the requirements for

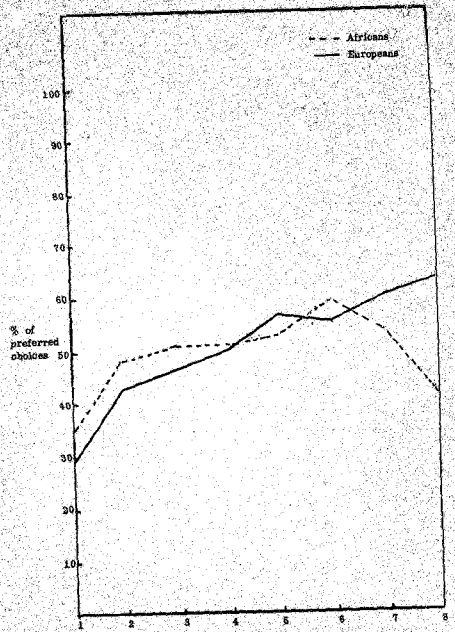

Figure4. Non-Representational Complexity Test: % of preferred choices made at each complexity level by the 2 cultural groups (results of Day 1 and Day 2 combined).

item equivalence set out in Chapter 2, namely that the pattern of item difficulties for the 2 samples should not be significantly different. (When results for the samples being compared are non-monotonic, as they are in this case, the requirements for item equivalence, stipulated by Poortinga, namely that the order of item difficulty in the 2 samples being compared should not be significantly different, must clearly be modified. In terms of the theoretical requirements outlined in Chapter II, a test can still be considered item equivalent for 2 samples provided that the pattern of item difficulties in the 2 samples is the same.) The difference in the form of the graphs for the 2 samples may be due to the fact that the preferred level of complexity for the European groups as a whole lay outside the range of the test. To check on this idea, the percentage of preferred choices made at each level of complexity by the 10 highest and 10 lowest scorers in each cultural sample were compared with each other. The results are shown in Figure 5. The graph for the 10 lowest scoring Europeans is rather irregular, but it can be seen in both the African and European samples that the peak of the graph has shifted to a lower level of complexity. Inspection of the graph for the 10 highest scorers in each cultural sample shows that the graph for the African sample now slopes down only at Complexity Level 8. It would thus appear that the range of complexity covered by the testing material was too short but that given a greater range of complexity, the pattern of item difficulties in the 2 samples

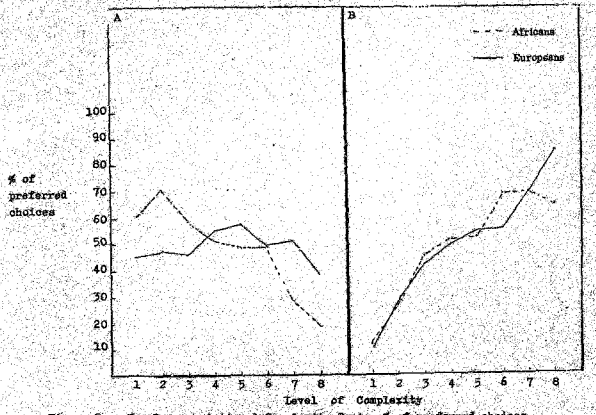

Figure 5. Non-Representational Complexity Test: % of preferred choices made at each level of complexity by: A. the 10 lowest scorers and B. the 10 highest scorers in each sample.

may have been similar.

The pattern of choices made by the subjects was clearly non-random in both groups and fits the pattern anticipated under Berlyne's theory and discussed in Chapter 1.

(b) The Representational Complexity Test

As in the previous test, the number of times the subject chose the more complex of a pair of pictures was taken as his individual score on each day of testing. The Reliability (Kuder-Richardson 20) of the test over the two testing days was 0,63 and 0,62 for the European and African samples, respectively. The mean, standard deviation of the scores for each cultural sample on the two testing days are shown in Table 3.

An analysis of Variance (nested design, Brownlee, 1965, p.528) carried out on the data showed no significant difference in performance between the two cultural samples, on either of the two testing days. This fact can possibly be explained by the low reliability of the tests.

Table 3

Representational Complexity Tests: Means, S.Ds and ~~Reliabilities~~ of scores.

	Mean	S.D
Africans Day 1	22,000	3,356
Day 2	22,840	3,981
Europeans Day 1	21,404	4,079
Day 2	23,106	3,546

Table 4

Non-Representational Complexity Test: Analysis of Variance with Culture (2x), Subjects, within culture (47x) and Day of testing (2x) as factors.

Source of Variation	Denominator	Mean Square	df num.	df denom.	F
Culture, C		4,473	1	92	0,121
Subjects within culture, S(C)	S(C)	36,902	92		
Day, D		58,644	1	92	3,464
C x D	D x S(C)	16,094	1	92	0,950
D x S(C)	D x S(C)	16,930	92		

Item Equivalence: Figure 6 shows the % of preferred choices made at each level of complexity by each of the cultural samples. The graphs for the 2 samples are highly similar, thus providing strong evidence for the item equivalence of the test.

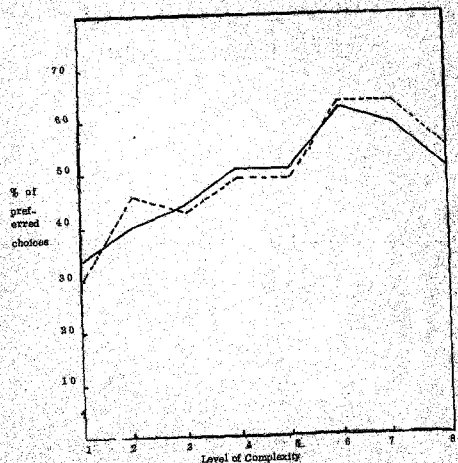

Figure 6. Representational Complexity Test: % of preferred choices made at each level of complexity.

Correlations between this test and the previous non-representational complexity test were 0,46 for the European group and 0,44 for the African group. Because of the low reliability of the representational complexity test in both samples, these correlation values are, of course, artificially lowered. Applying a correction for attenuation (cf.Lord & Novick, 1968, p.69) resulted in a correlation value of 0,59 for the African group and 0,63 for the European group, which points to the fact that the two complexity tests do to some extent tap the same construct in the two samples.

The fact that no difference in preference for complexity was observed on the representational complexity test makes it doubtful that differences in the mean scores obtained on non-representational complexity test can be interpreted solely in terms of cultural differences in "preference for complexity" in the 2 samples. It may well be that the non-representational testing material is more culture-bound than is usually supposed. In the construction of the battery of curiosity tests, it was originally thought that the non-representational complexity test would have a lower cultural content that the representational complexity test; it may well be that Europeans are more accustomed to dealing with abstract stimulus material and that the reverse is actually the case. A number of studies (Munsinger & Kessen, 1966; Vitz, 1966) have found that previous experience with the stimulus material, or similar stimulus material, results in a higher preference for complexity.

In both these tests, it is clear that the pattern of choices is non-random and provides confirming evidence for Berlyne's (1960) theory that subjects will in general look longer at the more complex patterns and for Dember and Earl's (1957) idea that subjects have an optimal or preferred level of complexity and will when given the choice prefer stimuli at this level of complexity to stimuli which belong either to a higher or lower complexity level.

2. The Uncertainty Test

Due to large inter-subject fluctuations in the total length of looking time at all the pictures on each day of testing, direct comparisons of looking time at individual pictures in the series between subjects were not possible: accordingly, the following transformation was applied to the data. The transformed looking time at any particular picture for each subject was calculated as a percentage of the subject's total looking time at all the pictures in the series on that day of testing. The transformed total looking time for each subject on each testing day was then equal to 100% and results could be compared over subjects.

The percentage of time spent looking at the 10 blurred pictures on each of the two testing days was taken as the subject's score. Means and standard deviations of scores are given in Table 5. Combining scores over the 2 testing days gave a reliability of 0,54 for the European sample and 0,80 for the African sample. The reliabilities were calculated by considering scores on the first and second day of testing for

each sample as two equivalent halves of the same test, correlating them and then correcting for the reduced length of the subdivisions by using the Spearman Brown formula (Lord & Novick, 1968,p.84) $r_{tt} = \frac{2r}{1+r}$ where r is the reliability obtained between the parts of the test. Table 5 also shows the results of t-tests carried out to establish whether looking time at the blurred pictures differed significantly from 50% (i.e. whether subjects looked longer at blurred than at clear pictures). Only in one case, the European sample on the first day of testing, did the subjects look significantly longer than 50% of the total looking time at the blurred pictures.

An analysis of variance (Table 6; nested design, cf. Brownlee, 1965, p. 538) showed no significant difference in score between the 2 cultural samples or between the 2 testing days.

Table 5

Uncertainty Test: Means, S.Ds of scores and the Results of t-tests carried out to test whether scores differed significantly from 50%.

		Mean	S.D	df	t
Africans	Day 1 (n=50)	51,060	11,160	49	0,67
	Day 2	48,320	11,449	49	- 1,65
Europeans	Day 1 (n=47)	53,851	8,325	46	2 981**
	Day 2	50,830	6,595	46	0,86

Table 6

Uncertainty Test: Analysis of Variance with Culture (2x), Subjects within Culture (47x) and Day of Testing (2x) as factors.

Source of Variation	Denominator	Mean Square	df num.	df denom.	F
Culture, C	S(C)	233,4	1	92	0,798
Subjects within culture, S(C)	-	293,8	92		
Day, D	D x S(C)	325,3	1	92	3,886
C x D	D x S(C)	9,1	1	92	0,0109
D x S(C)			92		

The average looking time (untransformed) at all the pictures in the test was compared over days for the two samples. The results are shown in Table 7. The degrees of freedom for the t-tests are corrected for inequality in variance between the 2 samples. F values for $p \leq 0,05$ are also given (cf. Kendal, 1946, p.114). No significant difference in the average length of looking time was found between the 2 samples on either day of testing.

Item Equivalence: As a test for item equivalence, the average transformed length of looking time at each item in the test was correlated in the 2 samples. A Pearson-product moment correlation of 0,59 was obtained between the African and European samples on the 2 days of testing. After correction for attenuation due to the low reliability of the test in the European sample (cf. Lord & Novick,1968 p.69) this value becomes 0,89, suggesting that the test does, to some extent, measure the same concept in the 2 samples. However, the reliabilities of the test for the 2 samples (0,80 for the African sample, and 0,56 for the European sample) differed significantly ($p \leq 0,05$) (cf. Edwards, 1953,p.131), casting some doubt on the item equivalence of the test for these 2 samples. "Culture" of the sample may well have been responsible for the higher reliability of scores in the African sample.

Our failure to find corroboration for Berlyne's hypothesis that ambiguous stimuli such as blurred pictures result in greater perceptual curiosity, and a consequently longer looking time, than pictures which are unambiguous, is perhaps

partly the result of unsuitable stimulus material.

A supplementar experiment, with 14 European and 12 African staff members of the National Institute for Personnel Research as subjects, was carried out as a check on the suitability of the stimulus material. The 40 pictures used in the main study were presented, one at a time, to these subjects for a period of 12 seconds. During this time, each subject was required to write down the content of the picture. The percentage of subjects, in each cultural group who correctly identified the picture was calculated for each of these 40 pictures and the results were correlated with the average looking time at these pictures for the African and European samples in the main study. The results are shown in Table 8.
(Spearman Rank Correlations were used since the data for the recognition test was not normally distributed, most of the non-blurred pictures being correctly identified.

Table 7

Uncertainty Test: t-values for the differences between average lengths of looking time(untransformed) between the two groups. P-values for $p \leq ,05$ are also given.

	df	t	P
Africans vs.Europeans Day 1	82,36	,42	2,30**
Day 2	95,00	,18	1,49
Africans Day 1 vs Africans Day 2	87,15	,30	1,94**
Europeans Day 1 vs Europeans Day 2	98,00	1,20	1,26

Table 8

Uncertainty Test: Spearman Rank Correlations between percentage recognition and average transformed looking time at items in the test.

		Looking Time		Recognition	
		Africans	Europeans	Africans	Europeans
Looking Time	Africans	1,00			
	Europeans	0,79	1,00		
Recognition	Africans	-0,84	-0,77	1,00	
	Europeans	-0,76	-0,81	0,86	1,00

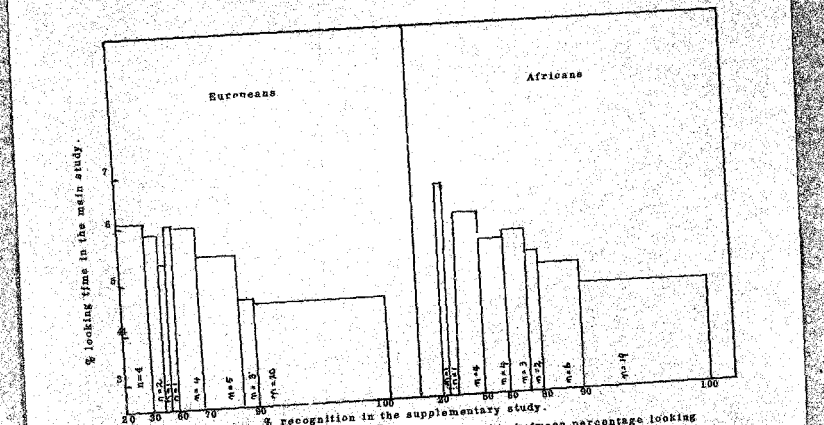

Figure 7. Uncertainty Test: Illustration of the relationship between percentage looking time in the main study and the degree of blurredness as measured by percentage recognition in the supplementary study. (n= no. of pictures at each level of recognition in the supplementary study).

by all the subjects in our supplementary study). It can be seen from Table 8 and Figure 7 that there is a highly significant correlation between percentage recognition in the supplementary experiment and the length of looking time in the main experiment.

However, the average percentage of correct identification for blurred and clear slides in the supplementary experiment was 67,53% and 95,42% respectively for the African subjects in the supplementary experiment and 64,46% and 93,22% for the European subjects, whereas the percentage looking time at clear and blurred pictures differed by not more than 4% (cf.Table 5). Clearly, in spite of the high correlation between % recognition in the supplementary experiment and the length of looking time in the main experiment, other factors, for example, complexity at the stimulus material, or perhaps the fact that the stimulus material itself was not sufficiently interesting to produce significant changes in looking time may have been responsible for the lack of large significant differences between looking time at blurred and non-blurred pictures.

Discarding the looking times at all clear pictures which were correctly identified by fewer than 60% of the subjects in both groups in the supplementary study and all blurred pictures which were recognized by 100% of the subjects resulted in leaving out 2 pictures (1 clear, 1 blurred) on the first day of testing and 4 pictures (2 clear, 2 blurred) on the second day of testing. Scores were again taken as the percentage of time subjects spent looking at the blurred

pictures of a series.

Means, S.Ds, reliabilities (corrected for length according to the Spearman - Brown formula) and the results of t-tests carried out to establish whether subjects looked significantly longer than 50% of the [illegible] at the blurred pictures, are given in Table 9. It can be seen that the results for the European sample in the main study are now significantly different from 50% on both testing days, while those of the Africans now approach significance on the first day of testing. Unsuitability of the stimulus material may thus be partly responsible for the failure to find strong support for Berlyne's hypothesis.

3. The Incongruity Test

The means, and standard deviations of the average length of looking time at all the items in the test can be found in Table 10. Differences in the length of looking time between the two cultural samples were not significant. As in the previous test, the large intersubject fluctuations in the total length of looking time necessitated transformation of the data. The same transformation as in the Uncertainty test was applied (i.e. the length of looking time at any one picture was taken as a percentage of the subject's total looking time at all the pictures in the test.) The subject's score was taken as the percentage of time spent looking at all the incongruous pictures. Means, S.Ds and reliabilities (corrected for length according to the Spearman-Brown formula) of scores are given in table 11. No significant difference

Table 9

Uncertainty Test: Means, S.D.s and Reliabilities of time spent looking at the blurred pictures after unsatisfactory items had been deleted, and the results of t-tests carried out to test whether the scores differed significantly from 50%.

		Mean	S.D	Rel.	df.	t
Africans	Day 1	52,610	11,392	0,79	49	1,60
(n=50)	Day 2	50,595	12,706		49	0,33
Europeans	Day 1	55,539	8,743	0,54	46	4,30**
(n=47)	Day 2	53,432	7,616		46	3,06**

Table 10

Incongruity Test: Means and S.Ds of the average length of looking time at all the pictures in the test.

	Mean	S.D
Africans (n=50)	6,68 sec	6,05
Europeans (n=47)	7,49 sec	7,06

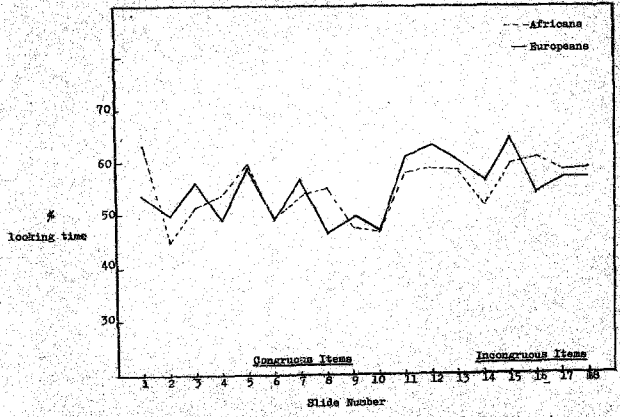

Figure 8. Incongruity Test: % of time spent looking at each picture in the series.

in score between the African and European samples was found.

The percentage of time spent looking at each picture in the series is shown in Figure 8. Mann-Whitney U-tests were carried out for both samples to test whether subjects looked significantly longer at the incongruous than at the congruous pictures in the test. Confirmation for this hypothesis

Table 11

Incongruity Test: Means, S.Ds and Reliabilities of scores

	Mean	S.D.	Reliability
Africans (n = 50)	35,16	10,94	0,74
Europeans (n = 47)	35,12	9,05	0,67

was found in the African sample ($p < 0,05$), but the results for the European group, although in the indicated direction are not significant.

Failure to confirm the findings of other workers in the field (Berlyne, 1960, 1963; Connolly & Harris,1971) is probably due to 2 reasons.

(i) The incongruous pictures all appeared together at the end of the series. Novelty of the test situation and items would presumably have had a greater effect the congruous items which appeared earlier in the test