Convergence Principles: Information in the Answer Sets of Some Multiple-Choice Intelligence Tests

A. P. White and J. E. Zammarelli
University of Durham

It is hypothesized that some common multiple-choice intelligence tests exhibit the property that the correct answer and the distractors together form a set of elements that, considered apart from the question, contain information as to which member of the set is the correct answer. Three formalized principles (couched in terms of set theory) are suggested, which enable the correct answer to be deduced from the answer set. The application of these principles to two intelligence tests is demonstrated and an experiment that supports the hypothesis is reported.

Typically, a multiple-choice test will feature a series of items, each of which leaves either a blank space or an incomplete entity as part of the question. In the set of possible answers to an item there is, of course, the correct one. However, for purposes of discrimination between people, this set must also include incorrect answers that some people are likely to choose. These incorrect answers are sometimes referred to as distractors.

In this paper it will be the contention that the set of answer options may contain information that "converges" on the correct answer. A set of theoretical principles designed to take advantage of this fact will be presented. In addition, an experiment is reported which was designed to see if untrained subjects can use this information to choose the correct answer when presented with only the answer options to test items.

Theory

A set of principles can be suggested for deriving the correct answer from the answer set (A), which work well on some multiple-choice items of a spatial or diagrammatic nature. The essence of the principles is that the answer set can be said to converge on a subset of itself, which will be termed the *reduced answer set* (R). This subset contains among its elements the correct answer. The principles of convergence are as follows:

1. *Frequency principle.* If all the dimensions of variation among the elements of the answer set (A) are considered, then a subset of A may be defined such that its elements are those members of A that possess the most frequent value or values in the answer set on each of the dimensions of variation. This subset is termed the *frequency subset* (F).

2. *Magnitude principle.* If all the dimensions of *ordered* variation among the elements of the answer set (A) are considered, then a subset of A may be defined such that its elements are those members of A that possess the central value or values on all such dimensions exhibiting more than two levels

among the levels of A. This subset is called the *magnitude subset* (M).

3. *Oddity-resemblance principle*. If there is present in the answer set (A) an element that is distinctly different from the rest and that is not a member of F, and if this element bears a definite resemblance to any other elements of A, then these latter elements constitute the *oddity-resemblance subset* (O).

The manner in which these principles are best applied should, perhaps, differ slightly from test to test. However, the following scheme is proposed as a possible approach on the grounds that it yields high levels of performance with two different tests, as described later in this paper. The procedure is as follows:

1. Apply the frequency principle to yield the frequency subset (F).
2. If applicable, apply the magnitude principle to yield the magnitude subset (M).
3. If applicable, apply the oddity-resemblance principle to yield the oddity-resemblance subset (O).
4. There are 7 possible ways in which the 3 subsets F, M and O may be combined to yield R. Listed in order of priority they are
   a. $R = F \cap M \cap O$,
   b. $R = F \cap O$,
   c. $R = M \cap O$,
   d. $R = F \cap M$,
   e. $R = O$,
   f. $R = F$, and
   g. $R = M$.

The rule is to select that solution set for R that occurs highest on the list, with the constraints that it must be both defined and nonempty.

5. If R (as defined above) contains one element only, this is clearly the choice for the answer. If R is defined as empty, then the algorithm has failed to provide any advantage and an answer must be selected at random from A. The third possibility—that R contains more than one element—is dealt with by random selection of an element from R to provide an answer. However, R will generally contain fewer elements than the original answer set and when it does, guessing from R will be clearly more likely to yield the correct answer than guessing from A.

Although the principles and their scheme of application sound complicated, it is rarely necessary to deal with them in their full complexity, even with answer sets containing as many as eight items. Frequently, only one principle of convergence is necessary and the authors have never needed to use more than two in solving an item. To illustrate the application of the principles of convergence, an example of the derivation of each type of subset from hypothetical answer sets is given.

Examples

As an example of the derivation of the F subset, in Figure 1 there are two dimensions of variation in the answer set: number of dots and orientation of the row of dots. Three elements in A possess four dots and only two possess three dots. Therefore, all elements of F will have four

Figure 1
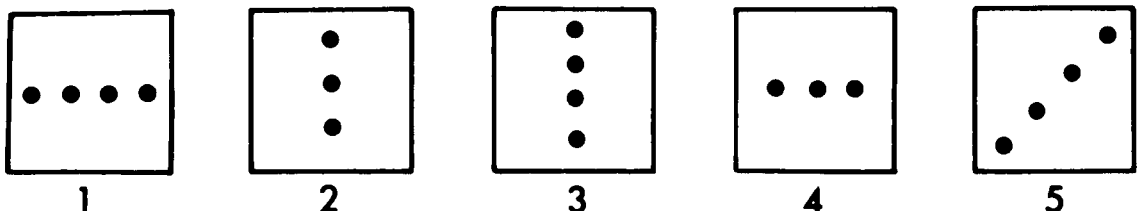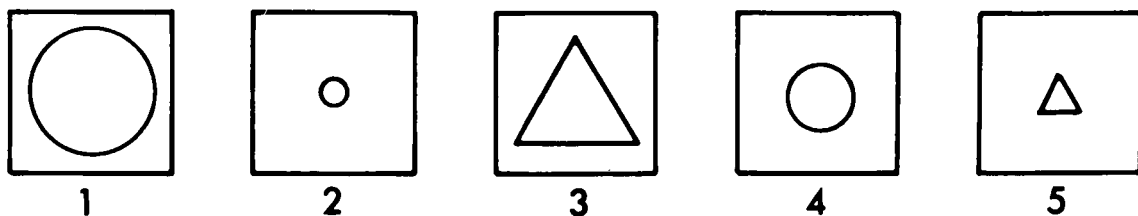Hypothetical Answer Set Illustrating Derivation of F

Figure 2
Hypothetical Answer Set Illustrating Derivation of M

dots. Two elements in A have vertical arrangements of dots and two have horizontal arrangements. Only one element has a diagonal arrangement. Therefore, all elements of F will have either a vertical or a horizontal arrangement of dots. Bearing both these constraints in mind, it can be seen that F consists of Elements 1 and 3.

As a demonstration of the application of the magnitude principle, there are two dimensions of variation in the answer set displayed in Figure 2: size and shape. Only size can be said to be an *ordered* dimension and it possesses three values, which will be designated small, medium, and large. Therefore, M will consist of all elements containing medium-sized figures; M thus contains only Element 4.

Finally, Figure 3 illustrates the derivation of the O subset. Here, Element 1 is clearly an oddity. No other elements have the small black circle incorporated in them. The only other element in the answer set bearing a distinct resemblance to Element 1 is Element 3. (Only Elements 1 and 3 contain an upright cross). Thus O contains Element 3 only.

Sample Applications

To demonstrate how these principles might be applied in practice, scales have been selected from two widely used intelligence tests—Test 3 of Scale 2 in Cattell's Culture Fair Test of *g* (Cattell & Cattell, 1973) and Set I of Raven's Advanced Progressive Matrices (Raven, 1958, 1965). For copyright reasons, the answer sets are not reproduced here, but a detailed analysis of the 12 items on each scale follows. There are 5 items in each answer set in the Cattell test and 8 items in each answer set in the Advanced Progressive Matrices. Three further points should be noted. Firstly, the dimensions identified in the answer set are indicated by numbering in parentheses the values chosen on those dimensions in the definition of F for each line. Secondly, the reduced answer set (R) for each item should be considered as being equal to the last subset listed in each case. Thirdly, an indication is given for each item as to whether the reduced answer set (R) obtained contains the correct answer.

Figure 3
Hypothetical Answer Set Illustrating Derivation of O
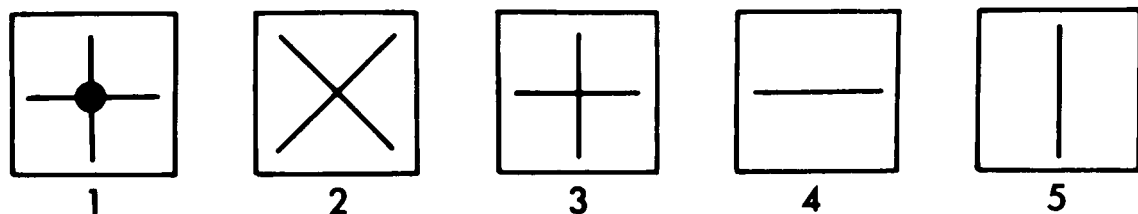
Cattell's Culture Fair Test of g

Item 1. Definition of F: 2 dots (1) in horizontal configuration (2); Elements in F: 1; Correct.

Item 2. Definition of F: 1 bar (1) at either horizontal or vertical angle (2); Elements in F: 3,4; Probability of guessing correctly = .5.

Item 3. Definition of F: Large (1) triangle—either white or striped (2); Elements in F: 2,3; M may be defined on color, regarding striped as intermediate between all-black and all-white; Elements in M: 2,4; Elements in F∩M: 2; Correct.

Item 4. Definition of F: One (1) short (2) line, sloping diagonally (3) from top left to bottom right (4); Elements in F: 5; Correct.

Item 5. Definition of F: Small (1) × (2) with thin (3) bars; Elements in F: 3; Correct.

Item 6. Definition of F: Black (1) square (2), standing on edge (3); Elements in F: 1; Correct.

Item 7. Definition of F: Large or medium-sized (1) black (2) circle; Elements in F: 2,3; M can be defined on size; Elements in M: 1,2; Elements in F∩M: 2; Correct.

Item 8. Definition of F: Both dots same color (1), sloping top left to bottom right (2); Elements in F: 2,4; Probability of guessing correctly = .5.

Item 9. Definition of F: Left hand (1) pointing up (2); Elements in F: 5; Correct.

Item 10. Definition of F: Arrow points to a top corner (1); three elements are rotationally equivalent (2); diagonal slopes top left to bottom right (3); Elements in F: 1; Correct.

Item 11. Definition of F: Three (1) long (2) lines, widely spaced (3), sloping top left to bottom right (4); Elements in F: 2; Correct.

Item 12. Definition of F: Narrow (1), all-white (2) cross with thin bars (3); Elements in F: 2; Correct.

Summary. Applying the principles in the manner outlined above, the correct answer can be converged on in 10 out of 12 items with a 50% chance of guessing correctly on the remaining 2 items. Thus, an expected score of 11 out of 12 could be claimed. If only the frequency principle is used, the correct answer can be identified for 8 items with a 50% chance of success on the remaining 4, resulting in an expected score of 10 out of 12.

Raven's Advanced Progressive Matrices

In the analysis of this test, there are a number of cases where a dimension of variation takes the form of a particular feature that may be either present or absent. Such instances are indicated with an asterisk.

Item 1. Definition of F: 1 vertical bar (1); 3 rows of dots (2); Elements in F: 8; Correct.

Item 2. Definition of F: 1 vertical bar (1); 1 horizontal bar (2); bars colored white (3); contrasting center (4)*; Elements in F: 4; Correct.

Item 3. Definition of F: Lines widen left to right (1) and top to bottom (2); wide spacing on both horizontal (3) and vertical (4) lines; black center (5)*; Elements in F: 5; M can be defined on line spacing on both horizontal and vertical dimensions; Elements in M: 4; Elements in F∩M: none; Therefore, R defined as F; Correct.

Item 4. Definition of F: Background with three (1) horizontal and three (2) vertical dotted lines, with equal spacing on both dimensions (4); figure drawn in heavy lines on background (5)*; Elements in F: 1; Correct.

Item 5. Definition of F: Background half white and half horizontal lines (1); complete (2) diamond, shaded with vertical lines (3); Elements in F: 4,7; Incorrect.

Item 6. Definition of F: Three (1) horizontal bars, with one (2) block; Elements in F: 5; Correct.

Item 7. Definition of F: Background either dots or horizontal lines (1) with a single (2) superimposed figure; Elements in F: 2,3,4,5,7; Incorrect.

Item 8. Definition of F: Three (1) concentric figures, all of same (2) shape; the figures are diamonds (3); Elements in F: 3; Correct.

Item 9. Definition of F: Three (1) straight (2) dotted (3) lines, oriented diagonally (4); Elements in F: 7; Correct.

Item 10. Definition of F: Two quarters of large square shaded black (1). Elements in F: 1,2,6; O can be defined using 5 as the odd element. Element 8 is most similar to it; Elements in O: 8; Elements in F∩O: none; Therefore, R defined as O; Correct.

Item 11. Definition of F: Background of intersecting arcs (1); upright cross (2); no dots (3)*; Elements in F: 7; Correct.

Item 12. Definition of F: 3 lines meeting at center (1)*; large circle present (2)*; no small black circle at center (3)*; Elements in F: 8; Incorrect.

Summary.   The full set of principles gives the correct answer on 9 out of the 12 items. If only the frequency principle is used, the correct answer can be identified on 8 items. It is instructive to examine Item 10 before leaving the Raven Matrices. Here, if F alone is used, an incorrect answer is obtained. However, O gives the correct answer. Generally, the authors have found the oddity-resemblance principle to be the most powerful of the three; and so, when it is used, it has precedence over the frequency principle in the priority ordering of subsets defining R, shown earlier in the rules of application.

Method

Thirty-five school pupils (aged between 14 and 17) were administered the "AH4" intelligence test (Heim, 1968). This test of general intelligence is in two parts, each of 65 items. The first part consists of various types of verbal and numerical items and the second part uses items of a spatial or diagrammatic nature. Further details are given in the manual for the test (Heim, 1970).

The pupils were then tested individually using just the answers from the Culture Fair Test of *g* (Cattell & Cattell, 1973), as follows. Each subject was presented with a sheet containing five examples of answer sets (i.e., correct answers accompanied by their distractors). The examples chosen were the answers only (no questions) from the first five items of Test 1 of Scale 2 of Cattell's test. The correct answer in each answer set was indicated on the example sheet. The nature of the contents of the example sheet was explained to the subjects and they were invited to try to develop ways of deriving the correct answers from their respective answer sets. Subjects were then presented with the answer sets (alone) for all 12 items of Test 3 of Scale 2 of the Cattell test and asked to do their best to derive the correct answer to each item. (In later parts of this paper, this task is referred to as the Cattell Answer Test). Unlimited time was allowed for those parts of the experiment dealing with the Cattell test. The experimenter gave no hints concerning principles of inference which might be used, nor were the subjects given any indication as to what the questions were. The test was scored in the conventional fashion and split-half reliability coefficients were computed using odd- versus even-numbered items.

Results

The mean number of items scored correct in the Cattell answer test was 4.1. The actual frequency distribution and the frequency distribution expected by chance (calculated using the binomial theorem) are shown in Table 1. It is clear that substantially fewer subjects scored closer to the expected value (2.4 items correct) than expected by chance and also that there were more extreme scores (either none correct or else more than four). When the data were grouped into three categories (the middle one being 2 or 3 items correct) and tested for conformity to the expected frequencies, a χ^2 value of 57.7 was obtained ($df = 2$; $p < .001$).

Low positive values were obtained for all the Pearson product-moment correlations computed between the Cattell Answer Test and the AH4. The total AH4 score yielded a correlation coefficient of .31. Parts I and II of the AH4 taken separately both gave correlations of .29.

Table 1
Frequency Distribution of Number of Items
Correct on Cattell Answer Test

Number of Items Correct	Number of Subjects	
	Actual	Expected
0	8	2.41
1	5	7.22
2	2	9.22
3	2	8.27
4	1	4.75
5	4	1.86
6	2	0.54
7	3	0.17
8	4	0.02
9	2	0.00
10	2	0.00

The split-half correlation coefficient for the Cattell Answer Test was .82, giving a value of .90 with the Spearman-Brown formula. The Guttman split-half reliability was .89.

The distribution of subjects scoring correct on each question in the Cattell Answer Test is shown in Table 2. This distribution is significantly non-uniform ($\chi^2 = 19.75$; $p < .05$). The table shows that only three items in the Cattell Answer Test were answered correctly by fewer than 12 of the 35 subjects. These were Items 2, 8, and 10.

Discussion

The results suggest that some subjects can utilize at least some of the information present in the answer sets of certain multiple-choice in-telligence tests. The Theory section above shows that the Cattell Answer Test contains two items (Numbers 2 and 8) that yield reduced answer sets with more than one element. It is worth noting that these items were among the three most difficult items in the test. This is further evidence that not only did the more successful subjects apply selection principles to the answer sets but also that those principles were probably similar to those derived by the authors.

It might be argued that the ability to do the Cattell Answer Test may well involve much the same sort of fundamental cognitive process as is required in most conventional intelligence tests. However, the low correlations obtained in this study between performance on the AH4 and performance on the Answer Test suggest that this is

Table 2
Frequency Distribution of Number of Subjects Scoring
Correct on Each Question in the Cattell Answer Test

Question Number	1	2	3	4	5	6	7	8	9	10	11	12
Number of Subjects	17	5	12	16	18	14	12	6	12	4	12	14

not the case. Indeed, it might tentatively be deduced from this evidence that quite different cognitive processes are involved in solving the problems described here, as opposed to those which appear in more conventional tests of intelligence.

Of course, a low correlation would be expected between an established test such as the AH4 and a variable that has a large component of random error. If the variance of the scores on the Cattell Answer Test were due to little except random noise, then the hypothesis of this study would not have been supported. However, there are two reasons this is not so. First, it is difficult to see why the distribution of scores obtained (see Table 1) should differ so markedly from that expected by chance if the scores were due largely to random error. Second, the split-half reliability coefficients for the Cattell Answer Test were as high as would be expected for a conventional psychometric test, suggesting that the Cattell Answer Test is a reliable measure. This, in turn, suggests that the proportion of variance due to random error must be small. Thus it can be inferred that the Cattell Answer Test is a reliable measure of an ability that has little in common with intelligence as conventionally measured.

If it is indeed true that different cognitive processes are required for performing the two types of task, then we might question the validity of a test that is clearly intended to be undertaken using one type of cognitive process when it can also be tackled using another of a quite different nature. This problem seems even more serious when multiple-choice achievement tests are considered. Imagine a test of, say, chemical knowledge in which a candidate could obtain a nearly perfect score without any chemical knowledge at all, merely by applying convergence principles to the answer sets.

An obvious way of overcoming this difficulty is to draw up sets of distractors for items in such a way that the answer sets contain no information as to which of their elements are the correct answers. A method for doing this is described by Guttman and Schlesinger (1967). Perhaps all the multiple-choice intelligence and achievement tests should be examined to see if they are susceptible to solution by the use of convergence principles. If this were done, how many of the commonly used multiple-choice tests would need to be provided with new sets of distractors and re-standardized?

References

Cattell, R. B., & Cattell, A. K. S. *Handbook for the Individual or Group Culture Fair Intelligence Test [a Measure of "g"]: Scale 2, Forms A and B.* Champaign, IL: Institute for Personality and Ability Testing, 1973.

Guttman, L., & Schlesinger, I. M. Systematic construction of distractors for ability and achievement test items. *Educational and Psychological Measurement*, 1967, 27, 569–580.

Heim, A. W. *AH4 Group Test of General Intelligence.* Slough, England: NFER Publishing Company, Ltd., 1968.

Heim, A. W. *AH4 Group Test of General Intelligence; Manual* (rev. ed.). Windsor, England: NFER Publishing Company, Ltd., 1970.

Raven, J. C. *Advanced Progressive Matrices; Set I.* London, England: H. K. Lewis & Co., Ltd., 1958.

Raven, J. C. *Advanced Progressive Matrices (Sets I and II). Plan and use of the scale.* London, England: H. K. Lewis & Co., Ltd., 1965.

Author's Address

Send requests for reprints or further information to A. P. White, Computer Centre, Elms Road, University of Birmingham, P.O. Box 363, Birmingham B15 2TT, England.