Table 2. Reliability coefficients for original and popular responses computed by test-retest and odd-even methods.

Response	Combined ($N = 56$)	First test ($N = 28$)	Second test ($N = 28$)
Test-retest reliability estimates			
Popular	.80		
Original	.86		
Odd-even reliability coefficients			
Popular	.82	.85*	.81
Original	.91	.91	.87

* N = 46 for this sample.

by these three methods of computation for describing the relationship between numbers of popular and of original responses are given in Table 1. The difference between the coefficients obtained by the correction methods is not significant; both coefficients are significantly different from the uncorrected *r* ($p < .05$ and $p < .01$).

The raw data were obtained by administering the Kent-Rosanoff word association test to 28 volunteer college students on two days 15 days apart. The respondents did not realize that they would take the test a second time until they were presented with the task 15 days after first taking it. The instructions for the two presentations were identical—namely, to give the first single word that comes to mind after reading the stimulus word.

The results from two methods of estimating the reliability of popular-response and original-response scores are presented in Table 2. By either method of computation, the original-response scores are the more reliable.

In summary, the scoring of popular responses in the word-association test provides a fair estimate of the number of original responses, accounting for from 55 to 62 percent of the variance rather than for the 77 percent estimated from an uncorrected correlation of popular with original responses. The number of original responses is significantly more stable than the number of popular responses; this is not surprising in view of the fact that more than one response can be scored as original for any given stimulus word but only one response can be scored as popular (*5*).

FREDERICK R. FOSMIRE
H. EDWARD TRYK
Department of Psychology, University of Oregon, Eugene

References and Notes

1. G. H. Kent and A. J. Rosanoff, *Am. J. Insanity* **67**, 37 (1910).
2. D. W. MacKinnon, *Am. Psychologist* **17**, 491 (1962).
3. R. S. Woodworth and H. Schlosberg, *Experimental Psychology* (Holt, New York, rev. ed., 1954). Similar statements are found in R. S. Woodworth, *Experimental Psychology* (Holt, New York, 1938).
4. *Popular response,* in our terminology, signifies the most common associate of any stimulus word; *original response* signifies an associate not included in the 1954 University of Minnesota norms for 1008 college students [see W. A. Russell and J. J. Jenkins, "The Complete Minnesota Norms for Responses to 100 Words from the Kent-Rosanoff Word Association Test" (Univ. of Minnesota Press, Minneapolis, 1954)].
5. This research was supported in part by the National Institute of Mental Health [grants M-3841(A) and M-4978)]. Ralph Hollingworth assisted with the computations.

8 October 1962

Personality Test Interpretation by Digital Computer

Abstract. *In this study a set of decision rules was devised for interpreting profile patterns of the Minnesota Multiphasic Personality Inventory (MMPI) of maladjusted and adjusted college students. The procedure used was that of computer programming of the "maladjusted" versus "adjusted" decisions of an expert test interpreter. The interpreter's decision-making processes were tape-recorded while he was thinking aloud during the sorting of the profiles of 126 college students. The programmed decision rules, which were based on the interpreter's protocol and which were improved upon by a process of trial-and-error statistical checking, yielded a greater hit percentage than the decisions of the original interpreter. In its final form, the set of objective configural inventory rules identified correctly large numbers of maladjusted college students in two cross-validation samples.*

Within the past few years research on human thinking has been facilitated by the introduction of the electronic digital computer as a research tool in the behavioral sciences and by the demonstration that this tool is much more than just a machine which performs rapid arithmetical operations. Among those who have contributed most to this development have been Allen Newell and Herbert A. Simon of Carnegie Institute of Technology and J. C. Shaw of the Rand Corporation (*1*). These workers have provided considerable and impressive evidence that the digital computer, when appropriately programmed, can carry out complex patterns of processes.

We now report how a computer has recently been applied as a tool to aid in interpretations of personality tests. The computer was used to approximate the rules expressed in the tape-recorded verbalizations of an expert test interpreter.

One of the personality tests which has frequently been used in psychiatric settings to aid in diagnostic and prognostic decision-making is the Minnesota Multiphasic Personality Inventory. The inventory is conventionally scored on a profile sheet which contains four validity and ten clinical scales. In this study an experienced user of the inventory was instructed to discriminate between the test profiles for maladjusted college students ($N = 45$) and those for well-adjusted students ($N = 81$).

The profiles for 126 college students (72 males and 54 females) were used as a criterion sample upon which a set of decision rules was developed. Such profiles were obtained from students who belonged to subgroups as follows.

1) Adjusted and maladjusted counseling group. This group was comprised of 65 students who had voluntarily requested help from the Carnegie Institute of Technology Counseling Center. They were judged by two counselors, after the completion of several interviews, to have problems of either a vocational-academic or a personal-emotional nature. The students whose problems were vocational-academic were labeled "adjusted" ($N = 37$); those whose problems were personal-emotional were called "maladjusted" ($N = 28$).

2) Adjusted and maladjusted no-counseling group. There were 31 students in this group, all members of fraternities and sororities. They were classed as either "adjusted" or "maladjusted" on the basis of the way in which their fraternity brothers (or sorority sisters) perceived them. Each member of each fraternity and sorority on campus, under supervised conditions, nominated from a roster of names of his fraternity brothers or sorority sisters four individuals, two of whom he considered the least well adjusted and two the best adjusted. A student was retained in this group if 60 percent or more of his peers nominated him, or her, for one of the two categories. Finally, 31 of these students (the 17 least well adjusted and the 14 best adjusted) were given the personality inventory test. This type of sample was chosen in order to counterbalance the

effects of test set; that is, in principle it may be expected that at the time students seek aid at a counseling center they are probably experiencing the full force of their emotional difficulties. Consequently their scale scores may be higher than they would be under less threatening circumstances. The fraternity-sorority maladjusted sample, therefore, may be similar in personality makeup to the personal-emotional-problem counseling group, except that the former are not in immediate need of counseling.

3) Random normal group. Thirty inventory profiles (15 males and 15 females) were randomly selected from a group of Multiphasic profiles for 800 entering freshmen. None of these students had either been in for counseling or been in the sample of persons nominated by the fraternity-sorority groups. This group, the vocational-academic-problem sample, and the "most adjusted" fraternity-sorority sample constituted the "adjusted" criterion groups. In all, then, the 126 profiles comprised profiles from 81 adjusted and 45 maladjusted students.

The 126 profiles were then prepared on 4½- by 4½-inch cards and Q-sorted by ten highly reputed experts in the United States. The mean profiles for the adjusted and maladjusted students are presented in Table 1. The experts were instructed to sort the profiles into a 14-step forced normal distribution which ranged from least to most adjusted; the number in each of the piles (from least to most adjusted) was 2, 3, 4, 9, 12, 15, 18, 18, 15, 12, 9, 4, 3, and 2, respectively. The cutting line between "maladjusted" and "adjusted" profiles was arbitrarily drawn in the middle of the distribution (between the two piles of 18 profiles).

Of the ten experts, one was chosen for intensive study and for computer programming because he had achieved the highest hit percentage. His sort/resort reliability, with an interval of 1 day between sorts, yielded a correlation coefficient of .96 and thus reflected considerable stability. On his first sorting he classified 80 percent and 67 percent of the profiles into the valid positive and valid negative categories, respectively; his second sorting yielded hit rates of 76 and 64 percent for the same categories. He was then instructed to Q-sort the profiles for the various subgroups and was asked to "think aloud" into a tape recorder during the process of sorting. The information which was obtained during approximately 30 hours of tape-recorded protocol was carefully edited, compiled, and then programmed into computer language so that an electronic digital computing machine could make decisions about profiles similar to the decisions made by the expert. In other words, the computer was given the sorter's information and procedure for processing that information (for example, the sorter's heuristics). Portions of the sorter's protocol and the corresponding decision rules are presented in Table 2.

Table 1. Mean T-score values for MMPI scales of the criterion group.

MMPI scale	Male		Female	
	Adjusted ($N=48$)	Maladjusted ($N=24$)	Adjusted ($N=33$)	Maladjusted ($N=21$)
?	50.0	51.0	50.0	50.9
L	48.1	47.6	48.1	46.0
F	51.8	58.5	52.5	60.9
K	55.9	54.4	57.5	49.4
HS	50.7	55.3	52.8	54.7
D	53.3	65.2	51.6	64.0
HY	56.5	61.1	58.0	61.9
PD	55.0	63.1	57.4	65.8
MF	56.0	65.0	45.0	45.2
PA	53.8	59.4	55.0	67.3
PT	53.4	64.2	56.0	64.0
SC	54.1	65.5	56.4	66.4
MA	57.0	60.0	58.2	63.4
SI	47.9	54.4	48.5	55.1
ES*	52.0	49.0	46.8	42.0
MT*	10.7	18.1	11.7	23.5

*Raw score.

When a sufficient amount of usable tape-recorded information had been obtained from the expert, the material was translated into computer language (GATE-20); this became the set of programmed decision rules (*2*). The rules were ordered in what was thought to be an optimal arrangement; the hit rates with these initial rules were 63 and 88 percent, respectively, for the valid positive and valid negative categories. The discrepancies between the expert's and the computer's hit percentages, which favored the expert in the valid positive case and favored the computer in the valid negative category, can be explained on the basis of my error. The error may possibly have been introduced through a nonoptimal ordering of the decision rules. However that may be, the task from that point on was clearly to sharpen the decision rules until they were superior to those of the original decision maker. This improvement was achieved by (i) a trial and error ordering and reordering of the rules; (ii) addition and deletion of portions of the rules; (iii) statistical searching and checking; and (iv) continual testing of the revised rules against the criterion sample. The completed set of decision rules included the original expert interpreter's information, a number of rank difference scores, and various slope characteristics of the profile pattern that the expert had failed to observe (*3*). With the revised set of decision rules, the hit rates for the 126 profiles were 91 and 84 percent, respectively, for the valid positive and valid negative categories. This was a considerable improvement over the expert's Q-sort.

Table 2. Portions of tape-recorded protocol and corresponding MMPI rules.

Protocol	MMPI rule
1) "... Now I'm going to divide these into two piles ... on the left [least adjusted] I'm throwing MMPI's with at least four scales primed. . . ."	1) If four or more clinical scales are equal to, or greater than, a T-score of 70, call maladjusted.
2) ". . . If the elevations are lopsided to the right with the left side of the profile fairly low, I'm throwing the MMPI's to the left [least adjusted]..."	2) If Pa or Sc and Pa, Pt, or Sc are equal to or greater than Hs, D, or Hy, then call maladjusted.
3) ". . . Here are a couple of nice, normal looking MMPI's . . . all scales are hugging around a T-score of 50, and Es is nice and high ... over to the right side [most adjusted]. . . ."	3) Call adjusted if at least 5 clinical scales are between T-scores 40 and 60 and if the Es scale's raw score is 45 or higher.

The efficacy of any statistical formula, model, or set of rules is frequently judged on the basis of its ability to hold up with new samples. Moderately satisfactory results have thus far been achieved when the new rules have been tested against a sample of well-adjusted and maladjusted counseling clients from the University of Nebraska ($N = 116$) and from Brigham Young University ($N = 100$). The hit rates for the Nebraska sample were 72 and 94 percent, respectively, for the valid negative and valid positive categories; for the Brigham Young sample the percentages were 80 and 64. To some extent the size of the cross-validation shrinkage can be accounted for on two bases: (i) the differences in the ratio of "maladjusted" to "adjusted" profiles between the criterion and the new samples, and (ii) the fact that both of the cross-validation samples, in terms of the way they were selected, resembled only a subgroup of the criterion sample (for example, the counseling group of 65 students). The percentage ratios of "maladjusted" to "adjusted" profiles for the criterion group, the Nebraska sample, and the Brigham Young sample were 36:64, 31:69, and 50:50, respectively. Perhaps if greater care had been exercised in selecting comparable cross-validation samples the size of the shrinkage might have been smaller.

The results obtained in this study have a threefold significance. (i) The decision rules should aid a counseling center in detecting emotional maladjustment in an entering freshman class. (ii) The use of the computer in an entirely new area of intelligent problem-solving has been demonstrated. (iii) This study could pave the way for rigorous investigation of clinical decision-making, which is more subtle than personality test interpretation.

BENJAMIN KLEINMUNTZ
Department of Psychology, Carnegie Institute of Technology, Pittsburgh, Pennsylvania

References and Notes

1. A. Newell, J. C. Shaw, H. A. Simon, *Psychol. Rev.* **65**, 151 (1958); A. Newell and H. A. Simon, *Science* **134**, 2011 (1961).
2. Grateful acknowledgment for assistance in computational and computer programming is made to R. Dale Shipp, graduate mathematics student at Carnegie Institute of Technology.
3. The revised set of MMPI rules will be furnished to interested readers upon request.
4. This research was supported in part by the National Institute of Mental Health, grant No. M-5701.

20 November 1962

Antibodies to Genetic Types of Gamma Globulin after Multiple Transfusions

Abstract. *Seventeen of 24 sera from children who had received multiple transfusions contained agglutinating antibodies against a Gm factor absent in the individual's serum gamma globulin. Each of these agglutinators was highly specific for a single Gm factor, and all proved useful as reagents for genetic typing. The accumulated evidence indicates that they resulted from genetically foreign gamma globulin introduced by transfusion.*

Agglutinating antibodies, used for delineation of the various genetic types of human globulin, occur rarely in normal sera and occasionally in sera from patients with rheumatoid arthritis (*1*). Through the use of such antibodies a number of genetic factors in human globulin have been delineated. These factors occur as alleles at two genetic loci (the Gm and Inv loci) (*2*). The implications for the administration of foreign genetic types of gamma globulin as transfusions or as pooled gamma globulin have not been ascertained. As an approach to this question, a search for antibodies against Gm and Inv factors has been carried out in children who have received multiple transfusions for chronic anemia. Blumberg *et al.* (*3*) found precipitating antibodies against beta lipoproteins in a small number of individuals in a similar group. In the study reported here, 17 of 24 children were found to have agglutinating antibodies against one of the Gm factors.

Twenty of the 24 children had thalassemia major, three had congenital nonspherocytic hemolytic anemia, and one had elliptocytosis. All had received at least three transfusions and some had received more than 40; the last transfusion had been given approximately 1 month prior to testing. With techniques previously reported from this laboratory (*4*), Gm-specific antibodies were detected by their ability to agglutinate group O, Rh_0 erythrocytes coated with appropriate anti-D sera. The specificity of the agglutinators was determined by inhibition of agglutination with normal sera of known Gm type, and Gm typing of all the sera was carried out with inhibition techniques in which test reagents of known specificity were used. Gm typing of sera containing anti-Gm agglutinators was accomplished by a variety of procedures in which the effects of these agglutinators in the typing system were nullified. Inhibition by 2-mercaptoethanol proved particularly useful.

The data for the patients studied are summarized in Table 1. Antibodies against factors Gm(a), Gm(b), or Gm(x) were found in sera from 15 of the 20 children with thalassemia and in sera from two of the three children with congenital hemolytic anemia. In each case agglutination was inhibited only by individual sera of appropriate Gm type; all such sera proved useful as highly specific typing reagents. No patient had serum containing antibodies against more than one Gm factor, and in each instance the patient's serum lacked the factor for which the agglutinator was specific. Anti-Gm(x) agglutinators were found only in those individuals whose sera contained the Gm(a) factor. The one patient whose serum lacked the Gm(b) factor had a low but definite anti-Gm(b) titer. Anti-Inv agglutinators were carefully sought but were not found, although at least seven of the sera lacked the Inv(a) factor. Precipitating antibodies to specific types of gamma globulin have not been demonstrated in any of these sera.

Table 1. Anti-Gm agglutinators in the 24 individual patients grouped according to Gm phenotype.

Agglutinator	
Type	**Titer**
a + b + x +	
	0
	0
	0
a + b + x −	
Anti(x)	1:80
Anti(x)	1:40
Anti(x)	1:8
	0
	0
	0
a − b + x −	
Anti(a)	1:1280
Anti(a)	1:320
Anti(a)	1:320
Anti(a)	1:160
Anti(a)	1:160
Anti(a)	1:80
Anti(a)	1:80
Anti(a)	1:80
Anti(a)	1:40
Anti(a)	1:40
Anti(a)	1:16
Anti(a)	1:16
Anti(a)	1:8
	0
a + b − x −	
Anti(b)	1:10

Sera from 27 pediatric patients who had not received transfusions were screened for anti-Gm agglutinators. No such agglutinators were found in this control group. Serum from 24 of these patients lacked at least one of the Gm factors.

The studies reported here revealed that 71 percent of the sera from a group of children who had received multiple transfusions contained anti-Gm agglutinators, and the facts which have been presented strongly support the conclusion that these antibodies resulted from the introduction of foreign gamma globulin through multiple transfusions. This incidence of specific anti-Gm agglutinators contrasts markedly with the low incidence in normal individuals previously reported (*1*). The possibility that normal individuals with sera containing such agglutinators may have received foreign gamma globulin should be carefully reconsidered. The clinical significance, if any, of the presence of these anti-Gm antibodies in patients that have received multiple transfusions remains to be explored (*5*).

JAMES C. ALLEN
HENRY G. KUNKEL
Rockefeller Institute, New York